Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

Inferences for a Partially Varying Coefficient Model With Endogenous Regressors

Zongwu Cai
Department of Economics, University of Kansas, Lawrence, KS 66045; Wang Yanan Institute for Studies in Economics, Ministry of Education Key Laboratory of Econometrics and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China (caiz@ku.edu)

Ying Fang, Ming Lin, and Jia Su
Wang Yanan Institute for Studies in Economics, Department of Statistics, School of Economics, Ministry of Education Key Laboratory of Econometrics and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China (yifst1@xmu.edu.cn; linming50@xmu.edu.cn; jia43zhu@163.com)

In this article, we propose a new class of semiparametric instrumental variable models with partially varying coefficients, in which the structural function has a partially linear form and the impact of endogenous structural variables can vary over different levels of some exogenous variables. We propose a three-step estimation procedure to estimate both functional and constant coefficients. The consistency and asymptotic normality of these proposed estimators are established. Moreover, a generalized F-test is developed to test whether the functional coefficients are of particular parametric forms with some underlying economic intuitions, and furthermore, the limiting distribution of the proposed generalized F-test statistic under the null hypothesis is established. Finally, we illustrate the finite sample performance of our approach with simulations and two real data examples in economics.

KEY WORDS: Endogeneity; Functional coefficients; Generalized F-test; Instrumental variables models; Nonparametric test; Profile least squares.

1. INTRODUCTION

Instrumental variables (IV) models have been widely used in empirical studies to correct potential endogeneity between regressors and structural errors and identify causal relations among several economic variables. By assuming the coefficients of all economic variables to be constant, linear IV models are commonly used but they might be too restrictive for some real economic applications. For example, in the literature of labor economics, one main research topic is to estimate the relationship between income and education. A linear IV model is typically adopted because of the endogeneity of years-of-schooling, the measurement of education. However, Schultz (1997) realized that marginal returns to education vary with the level of working experience and Card (2001) found that returns to education could be underestimated if a linear IV model was adopted by assuming the additive separation of education and working experience. In other words, a classical linear IV model might not be suitable for this typical application in the labor economics literature.

Our motivation comes mainly from studying the above example and other real applications in economics. Specifically, inspired by the above example, we propose a new class of semiparametric functional-coefficient IV models by allowing the effect of some structural variables to be functions of some exogenous variables. In such a way, the new model adopts a partially linear form, in which the coefficients of some variables are restricted to be constant but the coefficients of other variables are assumed to be functional, depending on some exogenous continuous economic variables selected based on some economic theories. Moreover, we allow for the existence of endogeneity in the structural equation. In the above example, the coefficient of education can be an unknown function of working experience. Apart from introducing model flexibility, the semiparametric framework and the functional-coefficient setup are helpful to alleviate the curse of dimensionality in a multivariate regression framework and to avoid the so-called ill-posed inverse problem in general nonparametric instrumental variables models; see Newey and Powell (2003).

The partially linear structure of the functional-coefficient instrumental variables model leads itself naturally to a three-stage estimation procedure as described later in detail. The first stage constitutes a projection of endogenous variables on a set of instrumental variables, the second stage is to estimate the constant coefficients by a profile least-squares approach, and the last stage uses the kernel method to estimate the functional coefficients. Particularly, we propose a new modified estimator of the constant coefficients and show that it may be more efficient than the conventional profile least squares estimator in some empirically relevant cases. We develop the large sample theory for the proposed three-stage estimators. Moreover, we propose two tests on testing both the constant parameters and the functional coefficients, respectively. In particular, a generalized F-test statistic is proposed novelly for testing the functional coefficients in a nonparametric instrumental variables model.

Journal of Business & Economic Statistics
January 2019, Vol. 37, No. 1
DOI: 10.1080/07350015.2017.1294079

This article builds on a vast amount of literature on nonparametric estimation of instrumental variables models. Newey and Powell (2003) pointed out that the identification of a general nonparametric instrumental variables model depends on a Fredholm integral equation of the first kind, which leads to the so-called ill-posed inverse problem. Newey, Powell, and Vella (1999) proposed a two-step nonparametric estimator of triangular simultaneous equation models. Ai and Chen (2003) considered a general model with conditional moment restrictions containing unknown functions, and a sieve minimum distance estimator was proposed. Both papers developed estimators using series approximation under regularization by compactness. Hall and Horowitz (2005) and Darolles et al. (2011) proposed to solve the ill-posed problem by the Tikhonov regularization. However, estimation of such general nonparametric instrumental variables models requires some strong regularizations and sometimes precludes from implementing practicable inferences.

To avoid the ill-posed inverse problem but at the same time to retain some model specification flexibility, it is common to impose some restrictions on a general nonparametric instrumental variables model. Das (2005) considered estimation of a nonparametric model whose endogenous variables are restricted to be a univariate discrete variable with finite support. Cai et al. (2006) proposed a two-step estimation of a nonparametric instrumental variables model under a fully functional-coefficient representation for structural equations. Our model extends it to a semiparametric (partially linear) framework and covers the aforementioned two models as special cases.

A functional-coefficient representation allows a regression model to be linear in some components with their coefficients given by unknown functions of other variables. Hastie and Tibshirani (1993) first introduced the functional-coefficient model into the literature, whereas Chen and Tsay (1993), Cai, Fan, and Li (2000a) and Cai, Fan, and Yao (2000b) explored the functional-coefficient model under a time series framework. Compared to a fully nonparametric model, in addition to capturing nonlinearity and heterogeneity, see Cai (2010) for details, a functional-coefficient model can accommodate structural information by choosing the smoothing variable and the functional form as well. Hence, testing on functional coefficients becomes a vehicle to test structural information and the underlying economic theory. To this end, Cai, Fan, and Yao (2000b) and Fan, Zhang, and Zhang (2001) proposed a generalized likelihood ratio test on functional coefficients model without endogeneity and showed that the well-known Wilks phenomenon holds for this case. Moreover, Fan and Huang (2005) extended the generalized likelihood ratio test to a varying-coefficient partially linear model without endogenous regressors. Su, Murtazashvili, and Ullah (2013) considered a nonparametric Wald-type statistic to test the constancy of the functional coefficients in an instrumental variables model. The generalized F-test proposed in this article allows regressors to be endogenous and can be used to test whether the functional coefficients are of any particular parametric forms.

Zhou and Liang (2009) is another article related to our research. Motivated by measurement errors, Zhou and Liang (2009) considered a semiparametric varying-coefficient partially linear model with endogeneity only in the linear part. They proposed a semiparametric profile least-squares based procedure to estimate the constant and varying coefficients, and applied a generalized likelihood ratio test to test the functional forms of the varying coefficients. By contrast, this article considers a general varying-coefficient instrumental variables model which allows for endogeneity existing in all structural variables except the smoothing variable, and we develop a new modified estimator of the constant coefficients and demonstrate its efficiency gains over the profile least-squares estimator in some empirically important cases. Moreover, we propose a novel generalized F-test for testing the varying-coefficient part and the proposed generalized F-test statistic can have an interpretation as an extended Wald-type statistic similar to that in Su, Murtazashvili, and Ullah (2013) with being always nonnegative in finite sample case, which is different in some way from the generalized likelihood ratio test statistic proposed in Zhou and Liang (2009).

The rest of the article is organized as follows. Section 2 introduces the model and three related motivated empirical examples are provided. Section 3 provides a three-stage estimation procedure and the asymptotic properties of the proposed estimators are established. Section 4 develops tests on both the constant coefficients and the functional coefficients. In particular, a generalized F-test statistic on the functional coefficients is proposed and its limiting distribution under the null hypothesis is derived. A Monte Carlo simulation study is conducted in Section 5 to demonstrate the finite sample performance of the proposed estimators and test statistics. Section 6 applies our methods to estimate the return to education and the growth effect of foreign direct investment (FDI), respectively. Finally, Section 7 concludes the article and all mathematical proofs are relegated to the Appendix in the supplementary material.

2. MODEL SETUP AND ILLUSTRATIVE EXAMPLES

A semiparametric functional-coefficient instrumental variables model assumes the following form:

$$Y_k = A_1^T(u_k)X_{k,1} + A_2^T(u_k)X_{k,2} + \beta_1^T W_{k,1} + \beta_2^T W_{k,2} + \varepsilon_k, \quad (1)$$

where Y_k is an observed scalar random variable, $X_{k,1}$ and $W_{k,1}$ are endogenous random variables, $X_{k,2}$ and $W_{k,2}$ are exogenous random variables, ε_k is a random error, A^T denotes the transpose of A, the coefficients $A(u_k) = (A_1^T(u_k), A_2^T(u_k))^T$ are unknown functions of exogenous continuous random variable u_k, and $\beta = (\beta_1^T, \beta_2^T)^T$ are constant coefficients. Without losing the generality, we assume the smoothing variable u_k to be a scalar to save notations.

The above model is general enough to include many popular models in the literature. For example, if the parametric part is excluded from model (1), it becomes a nonparametric functional-coefficient model proposed by Cai et al. (2006). When $X_{k,1}$ is a univariate discrete endogenous variable, model (1) is reduced to the model in Das (2005). If $A(u_k)$ takes the form of a threshold function, then model (1) includes the threshold instrumental variables model in Caner and Hansen (2004) as a special case. Moreover, if terms $A_1^T(u_k)X_{k,1}$ and $\beta_2^T W_{k,2}$ are excluded, model (1) reduces to the model with measurement in errors studied in Zhou and Liang (2009). Finally, if we restrict all variables to be exogenous, our model is reduced to a semiparametric functional-coefficient model studied by many authors in the statistics and econometrics literature, including but not limited to Hastie and Tibshirani (1993), Chen and Tsay (1993), Cai, Fan, and Li (2000a), Cai, Fan, and Yao (2000b), Li et al. (2002), Zhang, Lee, and Song (2002), Fan and Huang (2005), and Ahmad, Leelahanon, and Li (2005), among others.

Before discussing the model inference procedures, we illustrate our motivation on why we need to propose this new class of semiparametric functional-coefficient instrumental variables models by the following three real examples in the economics literature.

Example 2.1: Return to Education

Of interest in labor economics is to estimate the returns to education. Since education in a wage model is regarded as an endogenous variable due to unobservable heterogeneity in ability and schooling choices, a traditional empirical strategy is to employ a linear instrumental variables regression model. However, Schultz (1997) argued that marginal returns to education vary with different levels of working experience. As found in Card (2001), the returns to education would be underestimated if one would ignore the nonlinearity and the interacted impact between education and working experience. The above empirical findings suggest that it should be appropriate to use a semiparametric functional-coefficient instrumental variables model in which the impact of education allows to vary with the level of individual working experience.

Example 2.2: FDI and Economic Growth

It is an important issue in macroeconomics and international economics literature to examine the role of foreign direct investment (FDI) in the economic growth. Kottaridi and Stengos (2010) and Cai, Chen, and Fang (2016) found that a beneficial effect of FDI on economic growth exists only for countries at high level of income. In other words, the coefficient of FDI inflows in a growth model should be functional over the initial income level of each country. The nonlinear and heterogeneous impacts of FDI inflows motivate the adoption of a partially functional-coefficient model. Interestingly, as argued by Li and Liu (2004), the ratio of FDI to GDP is an endogenous variable. Therefore, to model the impact of FDI on the economic growth, we need to take care of both nonlinearity and endogeneity.

Example 2.3: Curse of Resources and Quality of Institutions

One of the core issues in development economics literature is to identify the impact of natural resources on economic development. The curse of resources summarizes the finding that resource abundant regions tend to grow slower than resource poor regions (Sachs and Warner 1997, 2001). However, we cannot neglect the fact that countries rich in natural resources constitute both growth losers and winners. Mehlum, Moene, and Torvik (2006) argued that quality of institutions is decisive for the resource curse. The variance of growth performance among natural resources abundant regions depends on how resource rents are distributed via the institutional arrangement, and whether it is grabber friendly or producer friendly. Therefore, this suggests that the coefficient of resource abundance should depend on the quality of institutions. Moreover, the resource abundance is often measured by the share of primary exports in GNP. Stijns (2005) argued that this measure is an endogenous variable and then we need to employ an instrumental variables regression.

In summary, model (1) is suitable for the aforementioned three real examples in economics and it provides an alternative to the existing literature to capture both nonlinearity and endogeneity as well as heterogeneity. The advantages of using the new model given by (1) to analyze the first two illustrative examples will be reported in details in Section 6.

3. ESTIMATION PROCEDURES

3.1 Notation

In this article, we consider an instrumental variables model with partially functional coefficients given in (1), reexpressed as follows:

$$Y_k = A^T(u_k)X_k + \beta^T W_k + \varepsilon_k, \quad k = 1, \ldots, n, \tag{2}$$

where X_k is a $p \times 1$ random vector, including endogenous random variables $X_{k,1}$ and exogenous random variables $X_{k,2}$, and W_k is a $d \times 1$ random vector, including endogenous random variables $W_{k,1}$ and exogenous random variables $W_{k,2}$.

Since $X_{k,1}$ and $W_{k,1}$ are endogenous, we need to find instrumental variables to estimate the structural equation. Let Z_k, including the constant term, the exogenous variables $X_{k,2}$, $W_{k,2}$ and other instrumental variables, be a $q \times 1$ vector, where $q \geq p + d$, which is the identification condition that the number of instruments is larger than the number of endogenous variables. Then, $E(\varepsilon_k \mid u_k, Z_k) = 0$. By taking a linear projection of Y_k on Z_k conditionally on u_k, we obtain that

$$\begin{aligned} Y_k = \mathrm{Proj}_{Z_k}(Y_k \mid u_k) + \varepsilon_k^* &= A(u_k)^T \mathrm{Proj}_{Z_k}(X_k \mid u_k) \\ &\quad + \beta^T \mathrm{Proj}_{Z_k}(W_k \mid u_k) + \varepsilon_k^*, \end{aligned}$$

where $\mathrm{Proj}_{Z_k}(V_k \mid u_k) = E(V_k Z_k^T \mid u_k)[E(Z_k Z_k^T \mid u_k)]^{-1} Z_k$ with $V_k = (Y_k, X_k^T, W_k^T)^T$ and $\varepsilon_k^* = Y_k - \mathrm{Proj}_{Z_k}(Y_k \mid u_k)$. Obviously,

$$E\left[\mathrm{Proj}_{Z_k}(X_k \mid u_k)\, \varepsilon_k^* \mid u_k\right] = 0$$
$$\text{and} \quad E\left[\mathrm{Proj}_{Z_k}(W_k \mid u_k)\, \varepsilon_k^* \mid u_k\right] = 0.$$

Therefore, the above structure suggests a three-stage estimation procedure. The first stage is to estimate $\mathrm{Proj}_{Z_k}(X_k \mid u_k)$ and $\mathrm{Proj}_{Z_k}(W_k \mid u_k)$ by a regression of (X_k, W_k) on Z_k conditional on u_k, and the second stage is to use a profile least-squares approach to estimate the constant coefficients β, and finally, the functional coefficients $A(u)$ is estimated by a nonparametric method such as the local linear fitting method.

We now define some notations which will be used throughout the article. To this end, we let $\Pi_X(u) = \mathbf{\Gamma}(u)^{-1} E(Z_k X_k^T \mid u_k = u)$ and $\Pi_W(u) = \mathbf{\Gamma}(u)^{-1} E(Z_k W_k^T \mid u_k = u)$ with $\mathbf{\Gamma}(u) = E(Z_k Z_k^T \mid u_k = u)$. Then, $\mathrm{Proj}_{Z_k}(X_k \mid u_k) = \Pi_X^T(u_k) Z_k$ and $\mathrm{Proj}_{Z_k}(W_k \mid u_k) = \Pi_W^T(u_k) Z_k$. We further define $v_{X,k} = X_k - \Pi_X^T(u_k) Z_k$ and $v_{W,k} = W_k - \Pi_W^T(u_k) Z_k$. By definition, the reduced form equations become

$$X_k^T = Z_k^T \Pi_X(u_k) + v_{X,k}^T \quad \text{and} \quad W_k^T = Z_k^T \Pi_W(u_k) + v_{W,k}^T,$$

where $E(Z_k v_{X,k}^T \mid u_k) = \mathbf{0_{q \times p}}$ and $E(Z_k v_{W,k}^T \mid u_k) = \mathbf{0_{q \times d}}$. Here, $\mathbf{0}$ stands for a matrix of zeros. We also use $K(u)$ to denote the kernel function and let $K_h(u) = K(u/h)/h$.

3.2 Three-Stage Estimation

There are some methods available in the literature to estimate a semiparametric functional-coefficient model and a semiparametric functional-coefficient instrumental variables model; see, for example, Fan and Huang (2005), Zhou and Liang (2009), and Cai and Xiong (2012). In this article, we adopt a three-stage estimation procedure. The first stage is to estimate the reduced form equation. Zhou and Liang (2009) and Cai and Xiong (2012) considered nonparametric reduced form equation. However, when the number of instruments is large, a fully nonparametric regression might become problematic for a moderate sample size due to the so-called curse of dimensionality. To overcome this problem, we propose using linear projection conditional on the smoothing variable, so that the reduced form equation becomes a functional-coefficient model as well. Then, we apply a profile least-squares method to estimate the constant coefficients. Particularly, we develop a novel modified approach to estimate the constant coefficients that can remove some effects of residues from the reduced form equation. When the variation in the reduced form equation is large or the structural equation and the reduced form equation have positive correlated variations, the modified estimator can be more efficient than the conventional profile least squares estimator, which can be evidenced from our simulation study (see Section 5). The last stage is to estimate the functional coefficients by a local linear regression.

Suppose that $\{Y_k, X_k, W_k, Z_k, u_k, k = 1, \ldots, n\}$ is a random sample observed from model (2). Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, $\mathbf{X} = (X_1, \ldots, X_n)^T$, $\mathbf{W} = (W_1, \ldots, W_n)^T$, $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\mathbf{M} = (A^T(u_1)X_1, \ldots, A^T(u_n)X_n)^T$,

$$\mathbf{D}_Z(u) = \begin{pmatrix} Z_1^T & \frac{u_1-u}{h} Z_1^T \\ \vdots & \vdots \\ Z_n^T & \frac{u_n-u}{h} Z_n^T \end{pmatrix} \quad \text{and} \quad \mathbf{D}_X(u) = \begin{pmatrix} X_1^T & \frac{u_1-u}{h} X_1^T \\ \vdots & \vdots \\ X_n^T & \frac{u_n-u}{h} X_n^T \end{pmatrix}.$$

Then, model (2) can be written as a matrix form as follows:

$$\mathbf{Y} = \mathbf{M} + \mathbf{W}\beta + \boldsymbol{\varepsilon}.$$

At the first stage, we estimate the linear projection of (X_k^T, W_k^T) on Z_k conditional on u_k by $(\widehat{X}_k^T, \widehat{W}_k^T)$, that is, let

$$\begin{aligned}\widehat{X}_k^T &= Z_k^T \widehat{\Pi}_X(u_k) = \left(Z_k^T,\ \mathbf{0}_{1\times q}\right) \\ &\quad \times \left[\mathbf{D}_Z(u_k)^T \mathbf{H}(u_k) \mathbf{D}_Z(u_k)\right]^{-1} \mathbf{D}_Z(u_k)^T \mathbf{H}(u_k) \mathbf{X},\end{aligned}$$

and

$$\begin{aligned}\widehat{W}_k^T &= Z_k^T \widehat{\Pi}_W(u_k) = \left(Z_k^T,\ \mathbf{0}_{1\times q}\right) \\ &\quad \times \left[\mathbf{D}_Z(u_k)^T \mathbf{H}(u_k) \mathbf{D}_Z(u_k)\right]^{-1} \mathbf{D}_Z(u_k)^T \mathbf{H}(u_k) \mathbf{W},\end{aligned}$$

where $\mathbf{H}(u) = \mathrm{diag}\left(K_h(u_1 - u), \ldots, K_h(u_n - u)\right)$. Furthermore, we define

$$\widehat{\mathbf{D}}_X(u) = \mathbf{D}_Z(u)\left[\mathbf{D}_Z(u)^T \mathbf{H}(u) \mathbf{D}_Z(u)\right]^{-1} \mathbf{D}_Z(u)^T \mathbf{H}(u) \mathbf{D}_X(u),$$

which is the first stage estimator.

The second stage is to estimate β by minimizing

$$\sum_{k=1}^n \left[Y_k - A(u_k)^T \widehat{X}_k - \beta^T \widehat{W}_k\right]^2. \tag{3}$$

For any given value of β, based on the local polynomial fitting theory, a local linear estimate of $A(u_k)$ and its derivative is given by

$$\begin{pmatrix} \widehat{A}(u_k) \\ h\widehat{A}'(u_k) \end{pmatrix} = \left[\widehat{\mathbf{D}}_X^T(u_k) \mathbf{H}(u_k) \widehat{\mathbf{D}}_X(u_k)\right]^{-1} \widehat{\mathbf{D}}_X^T(u_k) \mathbf{H}(u_k) \left(\mathbf{Y} - \widehat{\mathbf{W}}\beta\right). \tag{4}$$

where $\widehat{\mathbf{W}} = (\widehat{W}_1, \ldots, \widehat{W}_n)^T$. Substituting $\widehat{A}(u_k)$ for $A(u_k)$ in (3), it becomes

$$\begin{aligned}&\sum_{k=1}^n \left[Y_k - \widehat{A}(u_k)^T \widehat{X}_k - \beta^T \widehat{W}_k\right]^2 \qquad (5)\\ &= \left[\mathbf{Y} - \widehat{\mathbf{S}}\left(\mathbf{Y} - \widehat{\mathbf{W}}\beta\right) - \widehat{\mathbf{W}}\beta\right]^T \left[\mathbf{Y} - \widehat{\mathbf{S}}\left(\mathbf{Y} - \widehat{\mathbf{W}}\beta\right) - \widehat{\mathbf{W}}\beta\right] \\ &= \left(\mathbf{Y} - \widehat{\mathbf{W}}\beta\right)^T \left(\mathbf{I}_n - \widehat{\mathbf{S}}\right)^T \left(\mathbf{I}_n - \widehat{\mathbf{S}}\right) \left(\mathbf{Y} - \widehat{\mathbf{W}}\beta\right),\end{aligned}$$

where

$$\widehat{\mathbf{S}} = \begin{pmatrix} \left(\widehat{X}_1^T,\ \mathbf{0}_{1\times p}\right)\left[\widehat{\mathbf{D}}_X(u_1)^T \mathbf{H}(u_1) \widehat{\mathbf{D}}_X(u_1)\right]^{-1} \widehat{\mathbf{D}}_X(u_1)^T \mathbf{H}(u_1) \\ \vdots \\ \left(\widehat{X}_n^T,\ \mathbf{0}_{1\times p}\right)\left[\widehat{\mathbf{D}}_X(u_n)^T \mathbf{H}(u_n) \widehat{\mathbf{D}}_X(u_n)\right]^{-1} \widehat{\mathbf{D}}_X(u_n)^T \mathbf{H}(u_n) \end{pmatrix}.$$

and $\mathbf{I}_n$ is the $n \times n$ identity matrix. By minimizing (5), we obtain the conventional profile least-squares estimator of β given by

$$\widehat{\beta} = [\widehat{\mathbf{W}}^T (\mathbf{I}_n - \widehat{\mathbf{S}})^T (\mathbf{I}_n - \widehat{\mathbf{S}}) \widehat{\mathbf{W}}]^{-1} \widehat{\mathbf{W}}^T (\mathbf{I}_n - \widehat{\mathbf{S}})^T (\mathbf{I}_n - \widehat{\mathbf{S}}) \mathbf{Y}. \tag{6}$$

Here, we propose a modified estimator as follows:

$$\widetilde{\beta} = [\widehat{\mathbf{W}}^T (\mathbf{I}_n - \widehat{\mathbf{S}})^T (\mathbf{I}_n - \widetilde{\mathbf{S}}) \mathbf{W}]^{-1} \widehat{\mathbf{W}}^T (\mathbf{I}_n - \widehat{\mathbf{S}})^T (\mathbf{I}_n - \widetilde{\mathbf{S}}) \mathbf{Y},$$

where

$$\widetilde{\mathbf{S}} = \begin{pmatrix} \left(X_1^T,\ \mathbf{0}_{1\times p}\right)\left[\widehat{\mathbf{D}}_X(u_1)^T \mathbf{H}(u_1) \widehat{\mathbf{D}}_X(u_1)\right]^{-1} \widehat{\mathbf{D}}_X(u_1)^T \mathbf{H}(u_1) \\ \vdots \\ \left(X_n^T,\ \mathbf{0}_{1\times p}\right)\left[\widehat{\mathbf{D}}_X(u_n)^T \mathbf{H}(u_n) \widehat{\mathbf{D}}_X(u_n)\right]^{-1} \widehat{\mathbf{D}}_X(u_n)^T \mathbf{H}(u_n) \end{pmatrix}.$$

The main reason of doing so is to remove the effects of residues from the reduced form equation. In practice, to avoid the overwhelming influence on the estimation of β from the tail behavior of the smoothing variable u_k, we can follow the suggestion from Cai and Masry (2000) and Fan and Huang (2005) to trim some observations of u_k in its sparse regions.

Finally, after obtaining $\widetilde{\beta}$, the functional coefficients at any given value u can be estimated by

$$\begin{pmatrix} \widetilde{A}(u) \\ h\widetilde{A}'(u) \end{pmatrix} = \left[\widehat{\mathbf{D}}_X^T(u) \mathbf{H}(u) \widehat{\mathbf{D}}_X(u)\right]^{-1} \widehat{\mathbf{D}}_X^T(u) \mathbf{H}(u) \left(\mathbf{Y} - \mathbf{W}\widetilde{\beta}\right)$$

rather than the estimate given in (4).

3.3 Asymptotic Properties

Before studying the asymptotic properties of the proposed estimators, some technical assumptions are listed below for deriving the large sample theories for the three-stage estimators and the generalized F-test statistic proposed in the next section. To save notations, we define $\boldsymbol{\Phi}(u) = f(u)\, \Pi_X^T(u) \boldsymbol{\Gamma}(u) \Pi_X(u)$, $\boldsymbol{\Psi}(u) = f(u)\, \Pi_X^T(u) \boldsymbol{\Gamma}(u) \Pi_W(u)$ and $\boldsymbol{\Lambda}(u) = f(u)\, \Pi_X^T(u) E(Z_k \varepsilon_k^2 Z_k^T \mid u_k = u) \Pi_X(u)$, where $f(u)$ is the density function of u_k. Note that all the assumptions listed below are not necessary to be the weakest possibly.

Assumptions:

A1: Observations $(Y_k, X_k, W_k, Z_k, u_k)$, $k = 1, \dots, n$, are independent and identically distributed, where X_k is a $p \times 1$ random vector, W_k is a $d \times 1$ random vector, the instrument Z_k is a $q \times 1$ random vector with $q \geq p + d$, and u_k is an exogenous random variable such that $E(\varepsilon_k \mid Z_k, u_k) = 0$.

A2: The random variable u_k has a bounded compact support Ω, and its density function is Lipschitz continuous and bounded away from zero.

A3: $A(u)$, $\Pi_X(u)$ and $\Pi_W(u)$ have continuous second derivatives in Ω.

A4: The matrices $\mathbf{\Gamma}(u)$, $\mathbf{\Phi}(u)$, $\mathbf{\Psi}(u)$ and $\mathbf{\Lambda}(u)$ are nonsingular for any $u \in \Omega$.

A5: The kernel function $K(\cdot)$ is a symmetric density function with compact support and furthermore satisfies the Lipschitz condition.

A6: There exists an $s > 2$ such that $\sup_{u\in\Omega}\{E[\|Z_k\|^{2s} \mid u_k = u]\} < \infty$, $\sup_{u\in\Omega} \{E[\|X_k\|^{2s} \mid u_k = u]\} < \infty$, $\sup_{u\in\Omega}\{E[\|W_k\|^{2s} \mid u_k = u]\} < \infty$ and $n^{2\epsilon-1}h \to \infty$ for some $\epsilon < 1 - s^{-1}$.

A7: $n \to \infty$, $h \to 0$ and $nh^{2+\kappa} \to \infty$ for some $\kappa > 0$.

All the above assumptions are fairly standard in the literature of instrumental variables models and functional-coefficient models, such as in Fan and Huang (2005), Cai et al. (2006), and Zhou and Liang (2009). Assumption 1 requires that $q \geq p + d$ and Assumption 4 requires that the conditional covariance matrices are nonsingular. These conditions are sufficient for the model identification. Assumptions 2, 3, and 5 are quite common in the literature of local linear estimation. Assumptions 2, 5, and 6 suffice to employ a lemma in Mack and Silverman (1982) (see Lemma 1 in the Supplementary Appendix) to obtain the uniform convergence of a kernel estimator.

We now present the consistency and asymptotic normality of the estimators proposed in Section 3.2. The following theorem states the asymptotic distribution of $\sqrt{n}(\widetilde{\beta} - \beta)$ with the detailed proof given in the supplementary Appendix.

Theorem 1. Suppose that Assumptions A1–A7 hold. When $nh^4 \to 0$, we have

$$\sqrt{n}\left(\widetilde{\beta} - \beta\right) \xrightarrow{d} N\left(0, \Sigma_1^{-1}\Sigma_1^*\Sigma_1^{-1}\right),$$

where $\Sigma_1 = E(\Upsilon_k\Upsilon_k^T)$ and $\Sigma_1^* = E(\Upsilon_k\varepsilon_k^2\Upsilon_k^T)$ with

$$\Upsilon_k = \Pi_W^T(u_k)Z_k - \mathbf{\Psi}^T(u_k)\mathbf{\Phi}^{-1}(u_k)\Pi_X^T(u_k)Z_k.$$

Furthermore, if ε_k is conditionally homoscedastic; that is, $E(\varepsilon_k^2 \mid u_k, Z_k) = \sigma_\varepsilon^2$, then, we have

$$\sqrt{n}\left(\widetilde{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma_\varepsilon^2\Sigma_1^{-1}\right)$$

since $\Sigma_1^* = \Sigma_1\sigma_\varepsilon^2$.

The above theorem includes the conclusion in Fan and Huang (2005) as a special case. When both X_k and W_k are exogenous, they are included in Z_k. Therefore, $X_k = \Pi_X^T(u_k)Z_k$ and $W_k = \Pi_W^T(u_k)Z_k$. Furthermore, if ε_k is conditionally homoscedastic, then Σ_1 becomes $E(W_kW_k^T) - E[E(W_kX_k^T \mid u_k)E(X_kX_k^T \mid u_k)E(X_kW_k^T \mid u_k)]$, which is exactly the result of Theorem 4.1 in Fan and Huang (2005).

To present the asymptotic property of the local linear estimator of the functional coefficients, we define

$$\begin{pmatrix} \widetilde{A}(u) \\ h\widetilde{A}'(u) \end{pmatrix} = \left[\widehat{\mathbf{D}}_X^T(u)\mathbf{H}(u)\widehat{\mathbf{D}}_X(u)\right]^{-1}\widehat{\mathbf{D}}_X^T(u)\mathbf{H}(u)\left(\mathbf{Y} - \mathbf{W}\overline{\beta}\right), \tag{7}$$

where $\overline{\beta}$ is an estimate of β. Now, we have the next theorem.

Theorem 2. Suppose that $\overline{\beta}$ is a $\sqrt{n}$-consistent estimate of β and under Assumptions A1–A7, we have

$$\sqrt{nh}\left[\widetilde{A}(u) - A(u) - \frac{1}{2}h^2\mu_2 A''(u)\big(1 + o_p(1)\big)\right] \xrightarrow{d} N\left(0, \Sigma_2\right),$$

where $\mu_2 = \int t^2K(t)dt$ and $\Sigma_2 = \nu_0\mathbf{\Phi}^{-1}(u)\mathbf{\Lambda}(u)\mathbf{\Phi}^{-1}(u)$ with $\nu_0 = \int K^2(t)dt$. If ε_k is conditionally homoscedastic, then, the asymptotic variance reduces to $\Sigma_2 = \nu_0\sigma_\varepsilon^2\mathbf{\Phi}^{-1}(u)$ since $\mathbf{\Lambda}(u) = \sigma_\varepsilon^2\mathbf{\Phi}(u)$. Furthermore, if $nh^5 \to 0$, we have

$$\sqrt{nh}\left[\widetilde{A}(u) - A(u)\right] \xrightarrow{d} N\left(0, \Sigma_2\right). \tag{8}$$

Clearly, the bias term in the above theorem is of order h^2 and is the same as the bias in a semiparametric functional-coefficient model (Fan and Huang 2005) and in a nonparametric functional-coefficient instrumental variables model (Cai et al. 2006). The estimator of the functional coefficients is oracle in the sense that the estimator performs as well as the constant coefficients β and the reduced form coefficients $\Pi_X(u)$, $\Pi_W(u)$ would be known.

Bandwidth selection is a challenging issue in semiparametric models because of the nature of the multi-stage estimation. There are two bandwidths involved in the proposed three-stage estimation procedure. We use the same bandwidth in the first stage and the second stage to remove some effects of residues in the reduced form equation on the estimation of β. The optimal bandwidth $h = Cn^{-1/5}$ for a fully nonparametric functional coefficient model cannot be applied since we require that nh^4 tends to 0 (under-smoothing). Following Cai et al. (2006), we suggest employing a small bandwidth in the first two stages to eliminate the bias as much as possible. Specifically, we use cross-validation to select a bandwidth h_1 for the first-stage fitting, then the bandwidth used in the first two stages is $h = n^{-a}h_1$ with $a > \frac{1}{4}$. When β is estimated, the estimation of $A(u)$ in the third stage becomes a conventional functional-coefficient model and some bandwidth selection methods available in literature are applicable; see, for example, Li and Racine (2007).

4. INFERENCES

4.1 Inferences on Constant Coefficients

It is clear from Theorem 1 that to establish a confidence interval of β, we need to construct consistent estimators of Σ_1 and Σ_1^*, described as follows. First, Lemma 4 in the supplementary Appendix shows that

$$\widehat{\Sigma}_1 = \frac{1}{n}\widehat{\mathbf{W}}^T\left(\mathbf{I}_n - \widehat{\mathbf{S}}\right)^T\left(\mathbf{I}_n - \widehat{\mathbf{S}}\right)\widehat{\mathbf{W}} \xrightarrow{p} \Sigma_1.$$

Second, let $\widehat{\varepsilon}_k = Y_k - \widehat{A}^T(u_k)X_k - W_k\widetilde{\beta}$. By the similar argument in proving Lemma 4, it can be shown that

$$\widehat{\Sigma}_1^* = \frac{1}{n}\widehat{\mathbf{W}}^T\left(\mathbf{I}_n - \widehat{\mathbf{S}}\right)^T\begin{pmatrix}\widehat{\varepsilon}_1^2 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \widehat{\varepsilon}_n^2\end{pmatrix}\left(\mathbf{I}_n - \widehat{\mathbf{S}}\right)\widehat{\mathbf{W}} \xrightarrow{p} \Sigma_1^*.$$

Now, we consider the following hypothesis testing problem for β :

$$H_0 : F\beta = C \quad \text{versus} \quad H_1 : F\beta \neq C,$$

where F is a $b \times d$ full rank matrix with $b \leq d$ and C is a $b \times 1$ vector. Clearly, a Wald-type test statistic based on Theorem 1 is given by

$$W_n(F, C) = \left(F\widetilde{\beta} - C\right)^T\left[F\widehat{\Sigma}_1^{-1}\widehat{\Sigma}_1^*\widehat{\Sigma}_1^{-1}F^T\right]^{-1}\left(F\widetilde{\beta} - C\right).$$

The following corollary presents the limiting distribution of the proposed Wald-type test statistic.

Corollary 1. Suppose that Assumptions A1–A7 hold. When $nh^4 \to 0$, under the null hypothesis, we have

$$W_n(F, C) \xrightarrow{d} \chi_b^2$$

where χ_b^2 denotes a chi-square distribution with degrees of freedom b.

The consequence of Corollary 1 is to perform test at the significance level α as follows. We reject $H_0 : F\beta = C$ if $W_n(F, C) > \chi_b^2(\alpha)$, where $\chi_b^2(\alpha)$ is the α critical value of χ_b^2.

4.2 Inferences on Functional Coefficients

First, based on Theorem 2, we can construct a pointwise confidence interval for $A(u)$ for each given point u with higher-order bias ignored. To do so, we need to develop a consistent estimator of Σ_2. To this end, we define $\widehat{\mathbf{X}} = (\widehat{X}_1, \ldots, \widehat{X}_n)^T$ and

$$\widehat{\Sigma}_2 = \nu_0\left[\frac{1}{n}\widehat{\mathbf{X}}^T\mathbf{H}(u)\widehat{\mathbf{X}}\right]^{-1} \times \left[\frac{1}{n}\widehat{\mathbf{X}}^T\begin{pmatrix}\widehat{\varepsilon}_1^2K_h(u_1 - u) & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \widehat{\varepsilon}_n^2K_h(u_n - u)\end{pmatrix}\widehat{\mathbf{X}}\right] \times \left[\frac{1}{n}\widehat{\mathbf{X}}^T\mathbf{H}(u)\widehat{\mathbf{X}}\right]^{-1}.$$

Similar to the proof of Lemma 3 in the Supplementary Appendix, we can show that $\widehat{\Sigma}_2 \xrightarrow{p} \Sigma_2$. Therefore, one can easily construct a pointwise confidence interval of $A(u)$ by using the formulation given in (8).

Now, we consider constructing a test statistic on the functional coefficients $A(u)$. Testing on $A(u)$ is of particular interest in our model because we may want to impose some structural restrictions on the functional coefficients, which actually reflects some economic intuitions. Hence, testing on functional coefficients sometimes is equivalent to testing on underlying economic theories. Consider the following general hypothesis testing problem:

$$H_0 : A(u) = A_0(u; \theta) \quad \text{versus} \quad H_1 : A(u) \neq A_0(u; \theta),$$

where $A_0(u, \theta)$ is a known parametric function of u. The above hypothesis testing problem is a nonparametric test against a parametric form. It is general enough to include many interesting cases. For example, when $A_0(u, \theta) = \theta$, it becomes testing linearity. When θ is zero, it reduces to a significance test. If $A_0(u, \theta)$ is taken to be a given parametric function of u with some economic stories, then our interest is to test whether or not some economic theories hold.

Following the idea in Cai and Tiwari (2000), we construct a generalized F-test statistic based on the ratio of residual sum of squares (RSS) of the partially varying coefficient model to that for the parametric model. Note that we allow the parameter θ in the functional coefficients to be unknown, but to be consistently estimated with the rate of square-root-n under the null hypothesis. Let $\overline{\theta}$ and $\overline{\beta}$ be $\sqrt{n}$-consistent estimates of θ and β under the null hypothesis, respectively. The generalized F-test statistic is

$$\lambda_n(A_0) = \frac{\text{RSS}_0 - \text{RSS}_1}{\text{RSS}_1/n},$$

where $\text{RSS}_0 = \frac{1}{n}\sum_{k=1}^n S_0(u_k)$ is the RSS under the null hypothesis and $\text{RSS}_1 = \frac{1}{n}\sum_{k=1}^n S_1(u_k)$ is the RSS under the alternative hypothesis with

$$S_0(u) = \left[\mathbf{Y} - \widehat{\mathbf{D}}_X(u)\begin{pmatrix}A_0(u; \overline{\theta})\\ hA_0'(u; \overline{\theta})\end{pmatrix} - \mathbf{W}\overline{\beta}\right]^T\mathbf{H}(u) \times \left[\mathbf{Y} - \widehat{\mathbf{D}}_X(u)\begin{pmatrix}A_0(u; \overline{\theta})\\ hA_0'(u; \overline{\theta})\end{pmatrix} - \mathbf{W}\overline{\beta}\right],$$

and

$$S_1(u) = \left[\mathbf{Y} - \widehat{\mathbf{D}}_X(u)\begin{pmatrix}\widetilde{A}(u)\\ h\widetilde{A}'(u)\end{pmatrix} - \mathbf{W}\overline{\beta}\right]^T\mathbf{H}(u) \times \left[\mathbf{Y} - \widehat{\mathbf{D}}_X(u)\begin{pmatrix}\widetilde{A}(u)\\ h\widetilde{A}'(u)\end{pmatrix} - \mathbf{W}\overline{\beta}\right].$$

Here, $(\widetilde{A}(u_k), h\widetilde{A}'(u_k))$ are the nonparametric estimates of the functional coefficients and the first derivatives obtained by (7). From the proof of Lemma 6 in the Supplementary Appendix, it follows that

$$\text{RSS}_0 - \text{RSS}_1 = \frac{1}{n}\sum_{k=1}^n\left[\begin{pmatrix}\widetilde{A}(u_k)\\ h\widetilde{A}'(u_k)\end{pmatrix} - \begin{pmatrix}A_0(u_k; \overline{\theta})\\ hA_0'(u_k; \overline{\theta})\end{pmatrix}\right]^T \widehat{\mathbf{D}}_X^T(u_k)\mathbf{H}(u_k)\widehat{\mathbf{D}}_X(u_k)\left[\begin{pmatrix}\widetilde{A}(u_k)\\ h\widetilde{A}'(u_k)\end{pmatrix} - \begin{pmatrix}A_0(u_k; \overline{\theta})\\ hA_0'(u_k; \overline{\theta})\end{pmatrix}\right].$$

Therefore, the generalized F-test statistic can have an interpretation as an extended Wald-type statistic similar to that in Su, Murtazashvili, and Ullah (2013) with being always nonnegative in finite sample case, which is different in some way from the generalized likelihood ratio test statistic proposed by Zhou and Liang (2009). The following theorem presents the limiting distribution of the generalized F-test statistic, λ_n, under the null hypothesis.

Theorem 3. Suppose that Assumptions A1–A7 hold, when $nh^4 \to 0$, under the null hypothesis, we have

$$\sigma_n^{-1}\{\lambda_n - \mu_n\} \xrightarrow{d} N(0,1),$$

where

$$\mu_n = \frac{(\nu_0 + \mu_2^{-1}\nu_2)\, E\Big\{\text{Trace}\,\{\boldsymbol{\Phi}^{-1}(u_k)\boldsymbol{\Lambda}(u_k)\}\Big\}}{h\, E\left\{\left[\varepsilon_k + \upsilon_{X,k}^T A_0(u_k;\theta)\right]^2 f(u_k)\right\}}$$

and

$$\sigma_n^2 = \frac{2\int g^2(t)\,dt\, E\Big\{\text{Trace}\,\{\boldsymbol{\Phi}^{-1}(u_k)\boldsymbol{\Lambda}(u_k)\boldsymbol{\Phi}^{-1}(u_k)\boldsymbol{\Lambda}(u_k)f(u_k)\}\Big\}}{h\, E^2\left\{\left[\varepsilon_k + \upsilon_{X,k}^T A_0(u_k;\theta)\right]^2 f(u_k)\right\}}$$

with $\nu_0 = \int K^2(t)\,dt$, $\nu_2 = \int t^2K^2(t)\,dt$, $\mu_2 = \int t^2 K(t)\,dt$ and $g(t) = \int K(s)K(t+s)\,ds + \mu_2^{-1}\int s(t+s)K(s)K(t+s)\,ds$. Furthermore, if ε_k is conditionally homoscedastic, then, $\boldsymbol{\Phi}^{-1}(u_k)\boldsymbol{\Lambda}(u_k) = \sigma_\varepsilon^2\mathbf{I}_p$ so that

$$\mu_n = \frac{(\nu_0 + \mu_2^{-1}\nu_2)\, p\,\sigma_\varepsilon^2}{h\, E\left\{\left[\varepsilon_k + \upsilon_{X,k}^T A_0(u_k;\theta)\right]^2 f(u_k)\right\}}$$

and

$$\sigma_n^2 = \frac{2\int g^2(t)\,dt\, p\,\sigma_\varepsilon^4\, E\,[f(u_k)]}{h\, E^2\left\{\left[\varepsilon_k + \upsilon_{X,k}^T A_0(u_k;\theta)\right]^2 f(u_k)\right\}}.$$

Although we obtain the asymptotic distribution of λ_n, the test might be sensitive to the bandwidth h in finite sample case. To gain a better performance of the proposed test for the small-sample case in practice, we suggest using a bootstrap method to calculate the p-value for λ_n. We adopt the wild bootstrap method proposed by Davidson and MacKinnon (2010) for heteroscedastic noises. Specifically, let $\widehat{\varepsilon}_k = Y_k - \widehat{A}^T(u_k)X_k - W_k\overline{\beta}$, $\widehat{\upsilon}_{X,k} = X_k - \widehat{\Pi}_X^T(u_k)Z_k$ and $\widehat{\upsilon}_{W,k} = W_k - \widehat{\Pi}_W^T(u_k)Z_k$. Then, the bootstrap sample is generated as follows:

$$Y_k^* = A_0^T(u_k, \overline{\theta})X_k^* + \overline{\beta}^T W_k^* + \varepsilon_k^*,$$
$$X_k^{*T} = Z_k^T\widehat{\Pi}_{X,k}(u_k) + \upsilon_{X,k}^{*T},$$

and

$$W_k^{*T} = Z_k^T\widehat{\Pi}_{W,k}(u_k) + \upsilon_{W,k}^{*T}$$

for $k = 1, \ldots, n$, where $(\varepsilon_k^*, \upsilon_{X,k}^{*T}, \upsilon_{W,k}^{*T}) = (\widehat{\varepsilon}_k, \widehat{\upsilon}_{X,k}^T, \widehat{\upsilon}_{W,k}^T)e_k^*$, and

$$e_k^* = \begin{cases} -\frac{\sqrt{5}-1}{2}, & \text{with probability } \frac{\sqrt{5}+1}{2\sqrt{5}};\\ \frac{\sqrt{5}+1}{2}, & \text{with probability } \frac{\sqrt{5}-1}{2\sqrt{5}}.\end{cases}$$

We generate B datasets using the above bootstrap sampling scheme and let $\lambda_{n,1}^*, \ldots, \lambda_{n,B}^*$ be the corresponding bootstrap test statistics. The bootstrap p-value is $\sum_{i=1}^B I(\lambda_{n,i}^* \geq \lambda_n)/B$, where $I(\cdot)$ is the indicator function.

5. A MONTE CARLO SIMULATION STUDY

In this section, we present a simulated example to examine the finite sample performance of the proposed three-stage estimators and the generalized F-test on the functional coefficients as well as the proposed Wald-type test on the constant coefficients. For this purpose, we consider the following semiparametric functional-coefficient instrumental variables model:

$$Y_k = A(u_k)X_k + \beta_1 W_{k,1} + \beta_2 W_{k,2} + \varepsilon_k,$$

where the coefficients $A(u) = (1.6 + 0.6u)\exp\{-0.4(u-3)^2\}$, $\beta_1 = -1$ and $\beta_2 = 1$. The smoothing variable u_k follows a uniform[2, 6] distribution, $W_{k,2}$ is exogenous following a $N(0,1)$ distribution, and X_k and $W_{k,1}$ are endogenous following the reduced form equations:

$$X_k = \left[0.5 + \sin^2(u_k)\right]Z_{k,1} + \upsilon_{k,1},$$

and

$$W_{k,1} = \left[0.5 + \cos^2(u_k)\right]Z_{k,2} + \upsilon_{k,2},$$

where instrumental variable $Z_{k,1}$ and $Z_{k,2}$ are independently generated from a uniform[0, 4] distribution and the noises

$$\begin{pmatrix}\varepsilon_k\\ \upsilon_{k,1}\\ \upsilon_{k,2}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\sigma_\upsilon & \rho\sigma_\upsilon\\ \rho\sigma_\upsilon & \sigma_\upsilon^2 & 0\\ \rho\sigma_\upsilon & 0 & \sigma_\upsilon^2\end{pmatrix}\right),$$

where σ_υ controls the variation of residues in the reduced form equation, and ρ controls the correlation between the residues in the structural equation and in the reduced form equation.

First, we examine the effect of bandwidth selection on the performance of the proposed estimator $\widetilde{\beta}$. We set $\sigma_\upsilon^2 = 1$, $\rho = 0.7$ and conduct simulations by considering three sample sizes as $n = 250$, 500, and 1000. For each sample size, we replicate the experiment 1000 times. The Epanechnikov kernel function $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ is used. The cross-validation criterion suggests using $h_1 = 2.5n^{-1/5}$ for the first-stage fitting. To meet the requirement that nh^4 tends to 0, we fix the bandwidth for estimating β at three values $h = 1.25n^{-1/3}$, $2.5n^{-1/3}$ and $5n^{-1/3}$. The means and standard deviations of the estimated 1000 values for $\widetilde{\beta}$ under different settings are reported in Table 1. It shows that the performance of $\widetilde{\beta}$ is not sensitive to the choice of bandwidth. The estimator $\widetilde{\beta}$ is consistent and its simulated standard deviations are close to $\text{SD}_a(\beta)$, the asymptotic standard deviations provided by Theorem 1.

Table 1. Means and standard deviations of the Estimator $\widetilde{\beta}$

		$\widetilde{\beta}_1$		$\widetilde{\beta}_2$	
Sample size	Bandwidth	Mean	Std. dev.	Mean	Std. dev.
$n = 250$	$1.25\,n^{-1/3}$	−0.9944	0.0368	0.9979	0.0643
	$2.5\,n^{-1/3}$	−0.9965	0.0378	0.9976	0.0639
	$5.0\,n^{-1/3}$	−0.9957	0.0388	0.9979	0.0634
	SD_a	—	0.0388	—	0.0632
$n = 500$	$1.25\,n^{-1/3}$	−0.9978	0.0265	0.9988	0.0450
	$2.5\,n^{-1/3}$	−0.9989	0.0270	0.9986	0.0445
	$5.0\,n^{-1/3}$	−0.9987	0.0272	0.9984	0.0440
	SD_a	—	0.0275	—	0.0447
$n = 1000$	$1.25\,n^{-1/3}$	−0.9997	0.0187	0.9998	0.0321
	$2.5\,n^{-1/3}$	−1.0003	0.0190	0.9999	0.0321
	$5.0\,n^{-1/3}$	−1.0006	0.0191	0.9999	0.0320
	SD_a	—	0.0194	—	0.0316

NOTE: SD_a is the asymptotic standard deviation provided by Theorem 1.

Table 2. RMSE's of different estimators for β

Sample Size	σ_v	$\widetilde{\beta}_1$	$\widehat{\beta}_1$	β_1^C	$\widetilde{\beta}_2$	$\widehat{\beta}_2$	β_2^C
$n = 250$	$\sqrt{1/2}$	0.0384	0.0564	0.0551	0.0631	0.0723	0.0687
	1	0.0378	0.0632	0.0565	0.0639	0.0777	0.0735
	$\sqrt{2}$	0.0376	0.0780	0.0630	0.0647	0.0875	0.0784
	2	0.0367	0.1002	0.0623	0.0639	0.1022	0.0825
$n = 500$	$\sqrt{1/2}$	0.0279	0.0368	0.0365	0.0461	0.0503	0.0486
	1	0.0270	0.0413	0.0378	0.0445	0.0509	0.0492
	$\sqrt{2}$	0.0264	0.0493	0.0419	0.0445	0.0549	0.0500
	2	0.0267	0.0622	0.0464	0.0459	0.0660	0.0568
$n = 1000$	$\sqrt{1/2}$	0.0188	0.0235	0.0250	0.0315	0.0332	0.0327
	1	0.0190	0.0269	0.0267	0.0321	0.0355	0.0342
	$\sqrt{2}$	0.0192	0.0300	0.0272	0.0324	0.0378	0.0358
	2	0.0186	0.0400	0.0306	0.0317	0.0405	0.0381

NOTE: ρ is fixed at 0.7.

Then, we compare the performance of different estimators for β. We simulate the data with $\rho = 0.7$ and different σ_v's. The bandwidth is fixed at $h = 2.5n^{-1/3}$. Table 2 reports the root mean squared errors (RMSE) of different estimators for β based on 1000 independent simulations, where $\widehat{\beta}$ is the profile least squares estimator given in (6), and β^C is the estimator proposed in Cai et al. (2006), that is, after obtaining $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{W}}$, a local constant method is used to compute $\widehat{\beta}(u_k)$ at each u_k, $k = 1, \ldots, n$, then β^C is obtained by taking average of all $\widehat{\beta}(u_k)$'s. It can be seen clearly from Table 2 that $\widetilde{\beta}$ is more efficient than $\widehat{\beta}$ and β^C in this example and its RMSE is not sensitive to σ_v. Meanwhile, the RMSE's of $\widehat{\beta}$ and β^C increase as σ_v increases.

Now, we estimate the functional coefficient by $\widetilde{A}(\cdot)$ in (7), where $\widetilde{\beta}$ is used as a $\sqrt{n}$-consistent estimator of β and the bandwidth in this stage is determined by the cross-validation criterion. Figure 1 displays the estimated $A(\cdot)$ (dashed line) based on a typical sample with sample size $n = 500$, together with the true function (solid line). The estimated 90% pointwise confidence intervals (dash-dotted lines) computed based on Theorem 2 are also provided. It shows evidently that $\widetilde{A}(\cdot)$ estimates the true coefficient function very well. We use the root mean squared error (RMSE) as the performance measure of estimating $A(\cdot)$, which is defined as

$$\mathrm{RMSE} = \left\{ \frac{1}{m} \sum_{i=1}^{m} \left[\widetilde{A}(u^i) - A(u^i) \right]^2 \right\}^{1/2}$$

where u^i $(i = 1, \ldots, m)$ are the equally spaced grid points on the support of u_k. Here, we take m as 100. Figure 2 reports the boxplots of 1000 RMSE values of $\widetilde{A}(\cdot)$ for three sample sizes. One can see clearly that as the sample size increases, the RMSE shrinks toward zero.

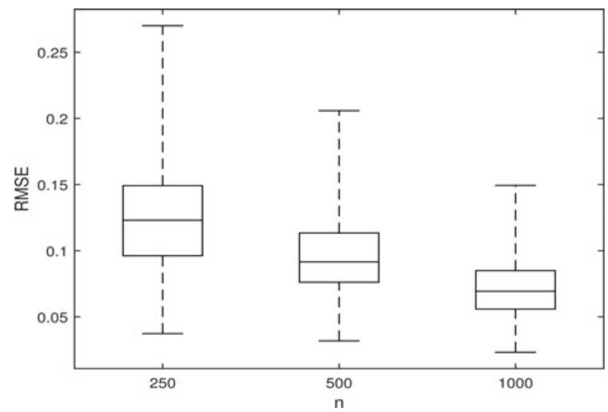

Figure 2. The boxplots of the RMSE values of $\widetilde{A}(\cdot)$ in 1000 independent simulations for three sample sizes $n = 250$, 500, and 1000.

Next, we study the size and power performance of testing β_1 with the null hypothesis as $H_0 : \beta_1 = -1$, $\beta_2 = 1$ against the alternative hypothesis $H_1 : \beta_1 = -1 + \gamma_1$, $\beta_2 = 1$. The power is indexed by γ_1. We use the Wald statistic proposed in Corollary 1 for sample size $n = 500$ and the bandwidth used here is $2.5n^{-1/3}$. We conduct 1000 simulations to obtain test sizes and powers of the proposed test. The power curves for three significance levels are plotted in Figure 3. When $\gamma_1 = 0$, the power collapses to the test size. The simulated sizes of the proposed test are 1.0%, 4.7%, and 10.8% corresponding to three significance levels 1% (dotted line), 5% (dashed line), and 10% (solid line), respectively. This implies that the simulated sizes are close to the nominal sizes and it concludes our test can deliver a correct test size. When γ_1 deviates from 0, the power curves tend to 1

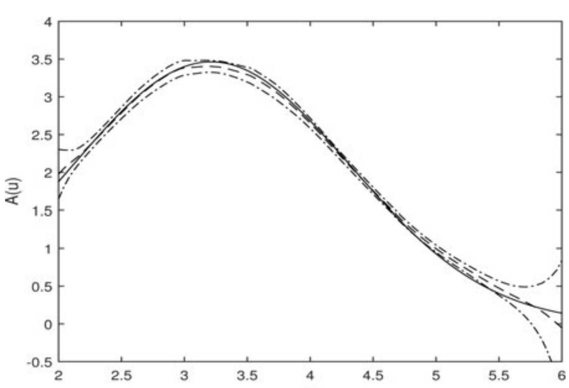

Figure 1. The estimated $A(u)$ when the sample size $n = 500$. The solid line represents the true $A(u)$, and the dashed line denotes the estimated values. The two dash-dotted lines are the 90% pointewise confidence intervals with bias ignored.

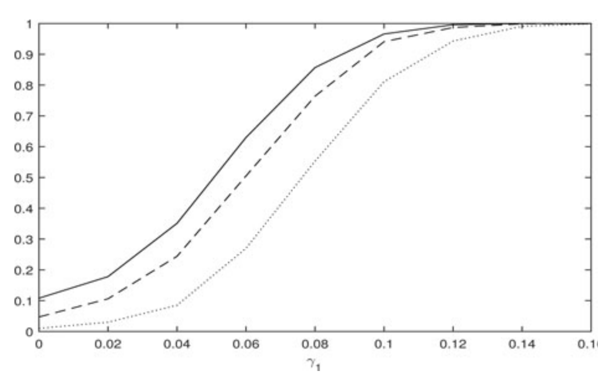

Figure 3. The power curves of testing $H_0 : \beta_1 = -1$, $\beta_2 = 1$ against $H_1 : \beta_1 = -1 + \gamma_1$, $\beta_2 = 1$ for sample size $n = 500$. The dotted line is the power curve for 1% significance level, the dashed line and the solid line are for 5% and 10% significance levels, respectively.

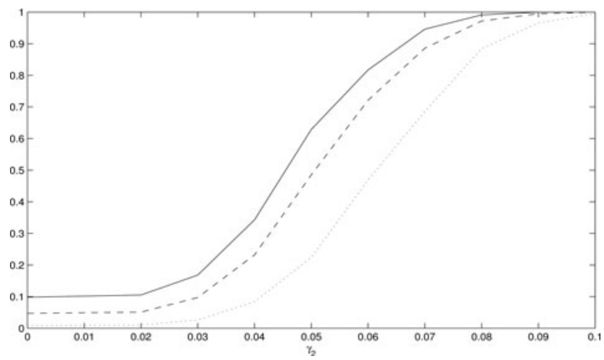

Figure 4. The power curves of the generalized F-test on the functional coefficient $A(u)$ for sample size $n = 500$. The dotted line is the power curve under 1% significance level, the dashed line and the solid line are for 5% and 10% significance levels, respectively.

quickly. This means that our test is reasonably powerful. Indeed, one can observe clearly from Figure 3 that the power is over 90% for both significance levels 10% and 5% when $\gamma_1 \geq 0.1$.

Finally, we apply the proposed generalized F-test to test the functional coefficient $A(\cdot)$. To examine the power performance, we consider the following hypothesis testing problem:

$$H_0 : A(u) = A_0(u) \quad \text{versus} \quad H_1 : A(u) = (1 + \gamma_2)A_0(u),$$

where $A_0(u) = (1.6 + 0.6u)\exp\{-0.4(u - 3)^2\}$. The power function is indexed by γ_2 and when $\gamma_2 = 0$, the alternative hypothesis becomes the null one. We use the unrestricted residual bootstrap procedure provided in Section 4.2 to determine the critical value. The number of bootstrap replications is set as $B = 399$. Figure 4 plots the power curves of the test obtained from 1000 simulations. The estimated test sizes (when $\gamma_2 = 0$) of the generalized F-test are 0.9%, 4.7%, and 9.8% corresponding to three significance levels 1% (dotted line), 5% (dashed line), and 10% (solid), respectively. It shows that the test can provide the correct test sizes and the power curves increase to 1 as γ_2 deviates from 0. It can be seen clearly from Figure 4 that the powers for both significance levels 10% and 5% are above 90% when $\gamma_0 \geq 0.08$. This means that the proposed generalized F-test is practically useful.

6. REAL EXAMPLES

6.1 Return to Education

We first consider an empirical example of return to education using the data of female youth aged between 16 and 25 years from the 1985 wave of the Australian longitudinal survey. We begin with a linear model

$$Y_k = \beta_0 + \beta_1 W_{k,1} + \beta_2 W_{k,2} + \beta_3 W_{k,3} + \beta_4 W_{k,4} + \beta_5 W_{k,5} + \alpha X_k + \varepsilon_k, \tag{9}$$

where Y_k is the natural logarithm of hourly wages, X_k is years of schooling, $W_{k,1}$ to $W_{k,5}$ denotes the marital status, government employment, union status, a dummy of whether is Australian born, and working experience, respectively. Since X_k is endogenous due to unobservable heterogeneity in ability and schooling choices, we follow Das, Newey, and Vella (2003) and Cai et al. (2006) to employ an index of labor market attitudes as the instrument. The index of labor market attitudes ranges from 0.6 to 3.0, which are individual respondence to survey questions related to work, social roles, and school attitudes. The higher scores indicate positive work attitudes. Table 3 reports the results of the two-stage least-squares (TSLS) estimators. The impact of education on wages is 0.154, which is strongly significant with a p-value less than 0.01. The coefficient of working experience is also positive and significant, with an estimated value of 0.093 and a p-value less than 0.01. All other coefficients are also significant. Union status and government employment have positive effects on wages, while marital status and whether Australian born have negative effects.

However, Schultz (1997) argued that return to education depends on different levels of working experience. Furthermore, Card (2001) found that ignoring the nonlinearity between years of schooling and working experience would underestimate the impact of education on wages. These features lead to a semiparametric model allowing the coefficient of years of schooling to depend on working experience. We use the working experience as the smoothing variable in our model; that is, $u_k = W_{k,5}$. Assuming the coefficient of X_k is a function of u_k, we apply the proposed generalized F-test to test whether or not other coefficients vary with the level of working experience. Since X_k is nearly constant conditional on some values of u_k in the dataset, we assume β_0 to a constant to avoid this local multicollinearity. Table 4 reports the bootstrap constancy test results

Table 4. Constancy test for other coefficients in Model (9)

	Marital status β_1	Government employment β_2	Union status β_3	Australian born β_4
Generalized F-statistic	0.0982	0.0832	2.0188	0.5451
p-value	0.7043	0.6466	<0.01***	0.0100**

NOTE: Significance codes: 0.01 '***', 0.05 '**', 0.10 '*'.

Table 3. Estimation results of linear instrument variables model in (9)

Variables	Constant	Marital status	Government employment	Union status	Australian born	Working experience	Education
Coefficient	−0.7045	−0.0599	0.0263	0.0479	−0.0558	0.0930	0.1542
p-value	0.0283**	0.0282**	0.0819*	0.0892*	0.0190**	<0.01***	<0.01***

NOTE: Significance codes: 0.01 '***', 0.05 '**', 0.10 '*'.

Table 5. Estimations of constant coefficients in Model (10)

	Constant β_0	Marital status β_1	Government employment β_2
Estimates	0.1554	−0.0462	0.0236
Wald Statistic	0.6151	3.7411	3.3314
p-value	0.4329	0.0531*	0.0680*

NOTE: Significance codes: 0.01 '***', 0.05 '**', 0.10 '*'.

of other coefficients. It shows that the coefficients of marital status and government employment can be treated as constant coefficients. Therefore, we consider the semiparametric functional-coefficient instrumental variable model as follows:

$$Y_k = \beta_0 + \beta_1 W_{k,1} + \beta_2 W_{k,2} + \beta_3(u_k) W_{k,3} + \beta_4(u_k) W_{k,4} + \alpha(u_k) X_k + \varepsilon_k, \quad (10)$$

where $\beta_3(u)$, $\beta_4(u)$, and $\alpha(u)$ are the functional coefficients.

We perform our three-stage estimation to model (10). Table 5 reports the estimation results of all the constant coefficients and the associated p-values of the Wald statistics for significance test. The coefficients of marital status and government employment are significant. The signs of these coefficients are the same as those in the linear instrument variables model (9).

Figure 5 plots the local linear estimate of the functional coefficient $\alpha(u)$ of years of schooling together with its 90% pointwise confidence interval with higher order bias ignored. We find that the values of $A(u)$ are significantly positive and increase with working experience. However, the estimated line is almost a straight line with positive slope which confirms the argument proposed by Schultz (1997) and Card (2001). Figures 6 and 7 plot the estimated functional coefficients of union status and Australian born, respectively. For individuals with long working experience, the effect of union status becomes less important and less significant. The estimated coefficient of the dummy variable of whether is Australian born varies with the level of working experience but is insignificant in almost the whole data range.

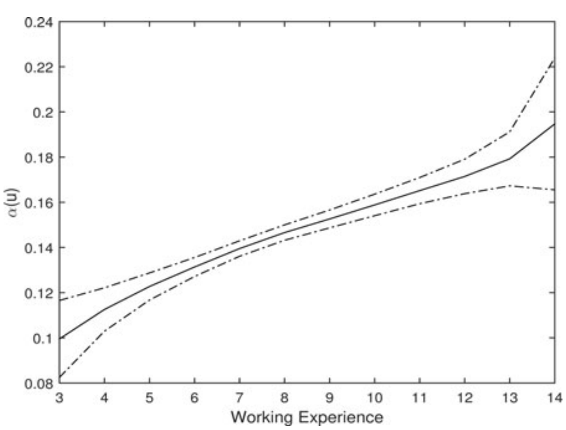

Figure 5. The estimated functional coefficient of years of schooling and its 90% pointwise confidence interval with bias ignored.

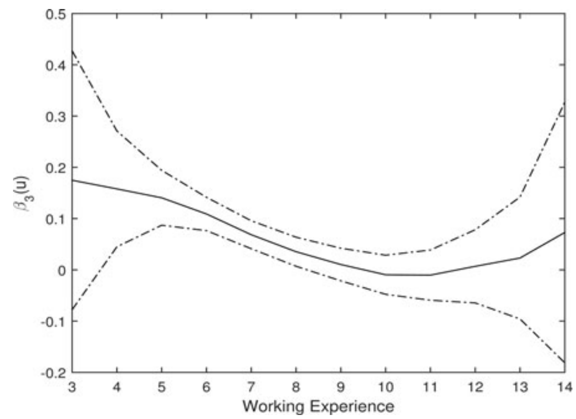

Figure 6. The estimated functional coefficient of union status and its 90% pointwise confidence interval with bias ignored.

6.2 FDI and Economic Growth

Now, we consider an empirical example of investigating how FDI has impact on economic growth in FDI host countries. To this end, we begin with a common linear model in the growth literature

$$y_{it} = \beta_0 + \beta_1 I_{it}^d + \beta_2 n_{it} + \beta_3 h_{it} + \beta_4 I_{it}^f + \beta_5 u_{it} + \varepsilon_{it}, \quad (11)$$

where y_{it} is the growth rate of income per capita in country or region i during period t, I_{it}^d is the domestic investment rate, n_{it} is the population growth rate, h_{it} is the human capital, I_{it}^f is the ratio of net FDI to GDP, and u_{it} is the logarithm of initial GDP per capita. The main interest here is to test whether the coefficient β_4 is significantly positive.

However, since there may exist unobserved factors or omitted variables correlated with FDI and affecting economic growth not only through the channel of FDI, the FDI rate is usually regarded as an endogenous variable in the above empirical growth model. A natural choice of the instrumental variable is the lagged variable of the FDI rate. Recent studies include Alfaro et al. (2004), Durham (2004), and Kottaridi and Stengos (2010).

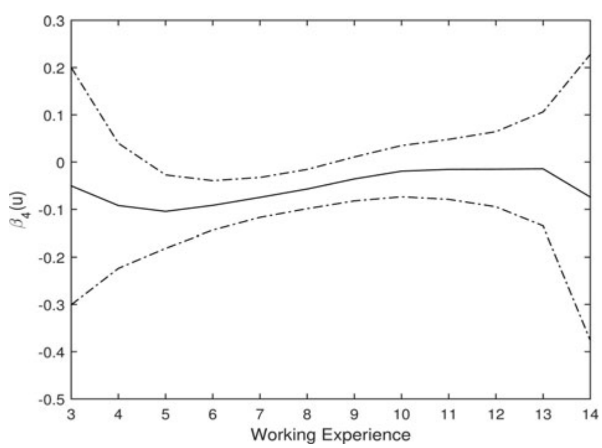

Figure 7. The estimated functional coefficient of Australian born and its 90% pointwise confidence interval with bias ignored.

Table 6. Summary statistics of economic variables

Variables	y_{it}	I_{it}^{d}	n_{it}	h_{it}	I_{it}^{f}	u_{it}
Minimum	−0.0606	0.0623	−0.0072	0.3660	−0.0538	4.9407
Maximum	0.0928	0.5812	0.0784	12.2470	0.5115	10.5001
Mean Value	0.0167	0.2176	0.0179	5.4112	0.0301	7.7056
Median	0.0170	0.2141	0.0193	5.1585	0.0092	7.6179

The sample used in this section includes 88 countries or regions from 1970 to 1999. We follow the literature; see, for example, Maasoumi, Racine, and Stengos (2007), Durlurf, Kourtellos, and Tan (2008), and Kottaridi and Stengos (2010), taking 5-year averages to smooth yearly fluctuations in macroeconomic variables. More specifically, we use the average data in 1975–1979, 1985–1989, and 1995–1999 as cross-sectional data to estimate the impact of FDI on economic growth. The average FDI rates in 1970–1974, 1980–1984, and 1990–1994 are used as instrument variables for the three periods, respectively. The domestic investment is computed by the average of the domestic gross fixed capital formation measured by the U.S. dollars in 2000 constant values. The population growth is computed by the average annual growth rate in each period, and the human capital is measured by mean years of schooling in each period. All these data are available to be downloaded from World Development Indicators. Finally, the data of FDI flows are available from United Nations Conference on Trade and Development, which are computed based on U.S. dollars in 2000 constant values. Table 6 summarizes the basic statistics of these economic variables.

Table 7 reports the estimation results of the linear instrumental variables models in (11). Models 1–3 in Columns 2–4 consider subsamples based on the level of the initial GDP per capita, which represent low, median and high income groups, respectively. Each group contains 88 observations. The marginal effect of domestic investment on economic growth is all positive but decreases from the low income to high income groups. The effect of FDI looks very different in three groups. In the low income group, the effect of FDI is negative with a p-value of 0.801. However, in the median and high-income groups, the effect turns to be positive with the estimated values 0.089 and 0.035, respectively. The median group estimate is not statistically significant with a p-value of 0.562, but the high group estimate is significant with a p-value of 0.057. The estimated coefficients of population growth and human capitals are neither significant. Models 4–6 move to the complete sample with different model specifications. In all these specifications, the effect of domestic investment is significantly positive and the estimates looks quite stable. The effects of FDI are neither significant in Models 4 and 5, yet the coefficient of FDI and its interacted terms with the initial condition and the squared term of the initial condition are all strongly significant in Model 6. The estimated coefficient of FDI is −6.21 with a p-value of 0.035, and the estimated coefficients of interacted terms are 1.39 with a p-value of 0.026 and −0.076 with a p-value of 0.021, respectively.

The nonlinear relationship between economic growth and FDI demonstrated in Table 7 is consistent to some arguments in the growth literature on the FDI effect on economic growth. The nonlinearity in FDI effect is mainly coming from the absorptive capacity in FDI host countries, which means that host countries need some minimum conditions to absorb the positive spillover effects from FDI. Nunnenkamp (2004) emphasized on the importance of the initial condition for host countries to absorb

Table 7. Estimation results of linear instrument variables models

Models	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	−0.8075	−0.0044	0.0771	−0.0154	−0.0147	0.0014
	(0.7977)	(0.9342)	(0.0673)	(0.1867)	(0.2167)	(0.9260)
I_{it}^{d}	0.6565	0.1705	0.1635	0.1427	0.1442	0.1554
	(0.7491)	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
n_{it}	3.5230	−0.5139	−0.3492	−0.1635	−0.1509	−0.1245
	(0.7951)	(0.2246)	(0.1648)	(0.3020)	(0.3680)	(0.4787)
h_{it}	−0.0165	0.0013	0.0007	0.0008	0.0009	0.0012
	(0.8124)	(0.5269)	(0.5612)	(0.3546)	(0.3512)	(0.2338)
I_{it}^{f}	−10.1055	0.0899	0.0350	0.0158	−0.1317	−6.2156
	(0.8012)	(0.5617)	(0.0570)	(0.4693)	(0.8156)	(0.0353)
u_{it}	0.1296	−0.0022	−0.0101	−0.0002	−0.0003	−0.0027
	(0.8034)	(0.7305)	(0.0446)	(0.9266)	(0.8774)	(0.2122)
$u_{it} \times I_{it}^{f}$					0.0148	1.3899
					(0.7921)	(0.0262)
$u_{it}^{2} \times I_{it}^{f}$						−0.0765
						(0.0211)
Sample	88	88	88	264	264	264

NOTE: The values in the parentheses are p-values.

Table 8. Constancy test for other coefficients in Model (12)

Coefficients	β_0	β_1	β_2	β_3
Generalized F-Statistic	5.8122	1.4380	10.4241	1.5842
p-Value	0.3960	0.9398	0.1930	0.8321

NOTE: Significance codes: 0.01 '***', 0.05 '**', 0.10 '*'.

Table 9. Estimation results of constant coefficients in Model (12)

	β_0	β_1	β_2	β_3
Estimates	−0.0159	0.1463	−0.1224	0.0005
Wald Statistic	4.1191	30.6234	0.4194	0.5962
p-Value	0.0424**	<0.01***	0.5172	0.4400

NOTE: Significance codes: 0.01 '***', 0.05 '**', 0.10 '*'.

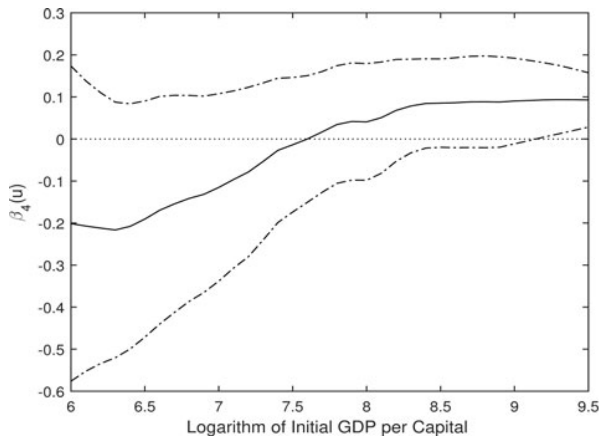

Figure 8. The estimated functional coefficient of FDI rate and its 90% pointwise confidence interval with bias ignored.

the positive effects from the adoption of FDI. To deal with this nonlinear relation, we propose a partially linear instrumental variables model. We allow the coefficient of FDI to depend on the initial condition in the host country. The initial condition is measured by the logarithm of initial GDP per capita in each period for the host country. Hence, we obtain the following empirical equation:

$$y_{it} = \beta_0 + \beta_1 I_{it}^d + \beta_2 n_{it} + \beta_3 h_{it} + \beta_4(u_{it}) I_{it}^f + \varepsilon_{it}, \quad (12)$$

where u_{it} is the logarithm of the initial GDP per capita in the period t in country i.

Before estimating the above partially functional coefficient instrumental variable model in (12), we first examine whether the other coefficients also depend on the initial condition u_{it} as well. Table 8 reports the testing results based on the generalized F-test using the bootstrapping method. The p-values are 0.378, 0.944, 0.185, and 0.787 for the coefficients from β_0 to β_3, respectively. In other words, the generalized F-test cannot reject the null hypothesis of constancy for the intercept term, the coefficients of domestic investment, population growth and human capita. The testing results support the use of a partially functional coefficient model given in (12).

Table 9 reports the estimation results of the constant coefficients in the partially functional coefficient instrumental variables model in (12). The intercept term and the coefficient of domestic investment are significant. The estimate of the coefficient of domestic investment is 0.148 with a p-value close to zero. The coefficients of population growth and human capita are not statistically significant. All these results have a similar pattern and magnitude compared to estimation results of these four coefficients in the linear model. Figure 8 presents the plot of the estimated functional coefficient of FDI rate. The estimated line has a clear pattern increasing with the level of initial GDP per capita. The effect of FDI is negative (not significant) for low initial GDP per capita, and then increases to significantly positive. The threshold value occurs around 7.5 of the logarithm of the GDP per capita. This empirical finding is in line with the hypothesis of absorptive capacity. The host countries with sufficiently high level of economic development can benefit from the positive spill-over effect of FDI on economic growth, but poor countries failed to do so.

7. CONCLUSION

In this article, we study a new class of semiparametric functional-coefficient instrumental variables models. We propose a three-stage estimator of the constant and functional coefficients. A novel generalized F-test is developed to allow for endogeneity in structural regressors in a partially functional coefficient framework. Moreover, we illustrate our estimation and testing approach with a simulated example. Our method works reasonably well in small samples. We then apply our method to the estimation of return to education and the growth effect of FDI, respectively. In the example of return to education, we find a strong evidence to reconfirm the argument as in Schultz (1997). In the example of FDI and economic growth, our results strongly support the hypothesis of the economic theory of the so-called absorptive capacity.

SUPPLEMENTARY MATERIALS

The supplementary PDF files contains additional proofs.

ACKNOWLEDGMENTS

The authors thank the Editor, the Associate Editor, and two anonymous referees for their helpful and constructive comments and suggestions that improved significantly the article. In particular, the authors are indebted to one of the reviewers who initiated the idea of using the conditional linear projection in Section 3.1. Finally, the authors also acknowledge the financial supports from NSFC with grants #71271179, #11101341, #71131008 (Key Project), #71631004 (Key Project), and the National Science Fund for Distinguished Young Scholars (#71625001).

[Received February 2015. Revised February 2017.]

REFERENCES

Ahmad, I., Leelahanon, S., and Li, Q. (2005), "Efficient Estimation of a Semiparametric Partially Linear Varying Coefficient model," *Annals of Statistics*, 33, 258–283. [3]

Ai, C., and Chen, X. (2003), "Efficient Estimation of Models with Conditional Moment Restrictions Containing unknown Functions," *Econometrica*, 71, 1795–1843. [2]

Alfaro, L., Chanda, S., Kalemli-Ozcan, S., and Sayek, S. (2004), "FDI and Economic Growth: The Role of Local Financial Markets," *Journal of International Economics*, 64, 89–112. [10]

Cai, Z. (2010), "Functional Coefficient Models for Economic and Financial Data," in: *Handbook of Functional Data Analysis*, eds. F. Ferraty and Y. Romain, Oxford: Oxford University Press, pp. 166–186. [2]

Cai, Z., Chen, L., and Fang, Y. (2016), "Quantile Panel Data Models with Partially Varying Coefficients with an Application to the Growth Effect of FDI," Working Paper, Department of Economics, University of Kansas. [3]

Cai, Z., Das, M., Xiong, H., and Wu, X. (2006), "Functional Coefficient Instrumental Variables Models," *Journal of Econometrics*, 133, 207–241. [2,5,8,9]

Cai, Z., Fan, J., and Li, R. (2000a), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902. [2]

Cai, Z., Fan, J., and Yao, Q. (2000b), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–956. [2]

Cai, Z., and Masry, E. (2000), "Nonparametric Estimation of Additive Nonlinear ARX time series: Local Linear Fitting and Projection," *Econometric Theory*, 16, 465–501. [4]

Cai, Z., and Tiwari, R. C. (2000), "Application of a Local Linear Autoregressive Model to BOD Time Series," *Environmetrics*, 11, 341–350. [6]

Cai, Z., and Xiong, H. (2012), "Partially Varying Coefficient Instrumental Variables Models," *Statistica Neerlandica*, 66, 85–110. [4]

Caner, M., and Hansen, B. E. (2004), "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843. [2]

Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160. [1,3,9,10]

Chen, R., and Tsay, R. S. (1993), "Functional-Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308. [2]

Darolles, S., Fan, Y., Florens, J. P., and Renault, E. (2011), "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565. [2]

Das, M. (2005), "Instrumental Variables Estimators for Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335–361. [2]

Das, M., Newey, W. K., and Vella, F. (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58. [9]

Davidson, R., and MacKinnon, J. M. (2010), "Wild Bootstrap Tests for IV Regressions," *Journal of Business & Economic Statistics*, 28, 128–144. [7]

Durham, J. B. (2004), "Absorptive Capacity and the Effects of Foreign Direct Investment and Equity Foreign Portfolio Investment on Economic Growth," *European Economic Review*, 48, 285–306. [10]

Durlurf, S. N., Kourtellos, A., and Tan, C. (2008), "Are any Growth Theories Robust?" *Economic Journal*, 118, 329–346. [11]

Fan, J., and Huang, T. (2005), "Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models," *Bernoulli*, 11, 1031–1057. [2,4,5]

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *Annals of Statistics*, 29, 153–193. [2]

Hall, P., and Horowitz, J. (2005), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904–2929. [2]

Hastie, T. J., and Tibshirani, R. J. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Series B, 55, 757–796. [2]

Kottaridi, C., and Stengos, T. (2010), "Foreign Direct Investment, Human Capital and Non-Linearities in Economic Growth," *Journal of Macroeconomics*, 32, 858–871. [3,10]

Li, Q., Huang, C. J., Li, D., and Fu, T. (2002), "Semiparametric Smooth Coefficient Model," *Journal of Business & Economic Statistics*, 20, 412–422. [3]

Li, Q., and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, New York: Princeton University Press. [5]

Li, X., and Liu, X. (2004), "Foreign Direct Investment and Economic Growth: An Increasingly Endogenous Relationship," *World Development*, 33, 393–407. [3]

Maasoumi, E. J., Racine, J. S., and Stengos, T. (2007), "Growth and Convergence: A Profile of Distribution Dynamics and Mobility," *Journal of Econometrics*, 136, 483–508. [11]

Mack, Y. P., and Silverman, B. W. (1982), "Weak and Strong Uniform Consistency of Kernel Regression Estimates," *Probability Theory and Related Fields*, 61, 405–415. [5]

Mehlum, H., Moene, K., and Torvik, R. (2006), "Institutions and the Resource Curse," *Economic Journal*, 116, 1–20. [3]

Newey, W. K., and Powell, J. L. (2003), "Nonparametric Instrumental Variables Estimation," *Econometrica*, 71, 1565–1578. [1]

Newey, W. K., Powell, J. L., and Vella, F. (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603. [2]

Nunnenkamp, P. (2004), *To What Extent Can Foreign Direct Investment Help Achieve International Development Goals?* New York: Oxford University Press. [11]

Sachs, J. D., and Warner, A. M. (1997), "Sources of Slow Growth in African Economies," *Journal of African Economies*, 6, 335–376. [3]

——— (2001), "The Curse of Natural Resources," *European Economic Review*, 45, 827–838. [3]

Schultz, T. P. (1997), "Human Capital, Schooling and Health," IUSSP, XXIII. General Population Conference, Yale University. [1,3,9,10,12]

Stijns, J. P. C. (2005), "Natural Resource Abundance and Economic Growth Revisited," *Resources Policy*, 30, 107–130. [3]

Su, L., Murtazashvili, I., and Ullah, A. (2013), "Local Linear GMM Estimation of Functional Coefficient IV Models with an Application to Estimating the Rate of Return to Schooling," *Journal of Business & Economic Statistics*, 31, 184–207. [2,6]

Zhang, W., Lee, S. Y., and Song, X. (2002), "Local Polynomial Fitting in Semivarying Coefficient Model," *Journal of Multivariate Analysis*, 82, 166–188. [3]

Zhou, Y., and Liang, H. (2009), "Statistical Inference for Semiparametric Varying-Coefficient Partially Linear Models with Generated Regressors," *Annals of Statistics*, 37, 427–458. [2,4,5,6]