Efficient Instrumental Variable Estimation of Semiparametric Dynamic Panel Data Models *

Badi H. Baltagi
Department of Economics, Texas A&M University
College Station, TX 77843-4228
Email: Badi@econ.tamu.edu

Qi Li
Department of Economics Texas A&M University
College Station, TX 77843-4228
Email: Qi@econ.tamu.edu

Abstract

This paper proposes some new semiparametric instrumental variable estimators for estimating a dynamic panel data model. Monte Carlo experiments show that the new estimators perform much better than the estimators suggested by Li and Stengos (1996).

Key words: Dynamic panel; semiparametric estimation; instrumental variable; efficient estimation; Monte Carlo simulation.

*Both authors would like to thank the Bush program in Economics of Public Policy, and the Private Enterprise Research Center, Texas A&M University, for supporting their research.

1 Introduction and the Model

There is a rich literature on semiparametric estimation of panel data models, see Horowitz and Markatou (1996), Ullah and Roy (1998), Ullah and Mundra (1999), Li and Hsiao (1998), to mention only a few. However, not much attention have been paid to the problem of instrumental variable estimation of semiparametric panel data models. Exceptions are Li and Stengos (1996), and Li and Ullah (1998). In this paper we consider the problem of efficient instrumental variable estimation of semiparametric models. Our Monte Carlo results show that the proposed estimators have substantial efficiency gain over the one suggested by Li and Stengos (1996).

We consider the following semiparametric dynamic panel data model

$$y_{it} = x_{it}'\beta + \theta(z_{it}) + u_{it}, \qquad (i = 1, ..., N;\ t = 1, ..., T) \tag{2.1}$$

where x_{it} is of dimension $p \times 1$, β is a $p \times 1$ vector of unknown parameters, z_{it} is of dimension d, $\theta(\cdot)$ is an unspecified smooth function. We allow x_{it} to contain some lagged values of y_{it} (say, the first element of x_{it} is $y_{i,t-1}$) so that model (2.1) is a semiparametric dynamic panel data model. We assume z_{it} is weakly exogenous in the sense that $E(u_{it}|z_{is}) = 0$ for $s \leq t$. Also, we assume the data is independent across the i index, so we have $Var(u) = I_N \otimes \Omega$, where $\Omega = var(u_i)$, u is of dimension $NT \times 1$, and u_i is of dimension $T \times 1$. We consider the case of large N and fixed T. Therefore, all asymptotics in this paper are for $N \to \infty$.

We allow the possibility that the error u_{it} is serially correlated. For example the error u_{it} could have a one-way error component specification, $u_{it} = \mu_i + \nu_{it}$, where μ_i is a random individual effect, which renders the errors serially correlated.

Li and Stengos (1996) propose a semiparametric instrumental variable (IV) estimator for β. However, not only is the semiparametric IV estimator of Li and Stengos (1996) not efficient, but it may also be inconsistent under certain situations. In this paper we propose some new semiparametric estimators that avoid the problems of the estimator suggested by Li and Stengos (1996). To simplify the notation and exposition, we will first discuss some infeasible estimators.

2 Semiparametric Instrumental Variable Estimators

2.1 Some Infeasible Estimators

Equation (2.1) contains an unknown function $\theta(.)$. Following Robinson (1998), we first eliminate $\theta(.)$. Taking the expectation of (2.1) conditional on z_{it}, then subtracting it from (2.1) we get

$$y_{it} - E(y_{it}|z_{it}) = (x_{it} - E(x_{it}|z_{it}))'\beta + u_{it} \equiv v_{it}'\beta + u_{it}, \tag{2.2}$$

where we have used $E(u_{it}|z_{it}) = 0$, and $v_{it} \stackrel{def}{=} x_{it} - E(x_{it}|z_{it})$. Note that v_{it} and u_{it} are correlated because v_{it} contains lagged values of y_{it}, and u_{it} is serially correlated. Suppose there exists a $q \times 1$ ($q \geq p$) instrumental variable w_{it} that is correlated with x_{it} and uncorrelated with u_{it}, see equation (2.4) below for details. Then we can estimate β using IV-OLS as follows.

$$\tilde{\beta}_{IVO} = (v'ww'v)^{-1}v'ww'(y - E(y|z)) = \beta + (v'ww'v)^{-1}v'ww'u, \tag{2.3}$$

where w and v are of dimension $n \times q$, $n \times p$, with typical rows given by w_{it}' and v_{it}'. y, $E(y|z)$ and u are all $n \times 1$ vectors with typical row elements given by y_{it}, $E(y_{it}|z_{it})$ and u_{it}, respectively. Denote w_i and u_i the $T \times q$, and $T \times 1$ matrices, with typical rows given by w_{it}' and u_{it}. The following conditions are needed to derive the asymptotic distribution of $\tilde{\beta}_{IVO}$.

$$\begin{aligned} &w'v/n \xrightarrow{p} E[w_{it}v_{it}'] \equiv A, \\ &w'uu'w \xrightarrow{p} E[w_i'u_iu_i'w_i]/T \equiv B, \\ &w'u/n \xrightarrow{p} E[w_{it}u_{it}] = 0, \end{aligned} \tag{2.4}$$

where $n = NT$. Using (2.4), and standard central limit theorem arguments, it is easy to show that

$$\sqrt{n}(\tilde{\beta}_{IVO} - \beta) \to N(0, V_O) \text{ in distribution}, \tag{2.5}$$

where $V_O = Q^{-1}A'BAQ^{-1}$, $Q = A'A$ with A and B defined in (2.4).

Next we consider the case where $Var(u) = \Sigma \equiv I_N \otimes \Omega$ is known. Also we assume that the error is conditionally homoskedastic, i.e., $Var(u|w) = \Sigma$. In this case, we can estimate β using the following IV-GLS estimator:

$$\begin{aligned} \tilde{\beta}_{IVG} &= (v'\Sigma^{-1}w(w'\Sigma^{-1}w)^{-1}w'\Sigma^{-1}v)^{-1}v'\Sigma^{-1}w(w'\Sigma^{-1}w)^{-1}w'\Sigma^{-1}(y - E(y|z)) \\ &= \beta + (v'\Sigma^{-1}w(w'\Sigma^{-1}w)^{-1}w'\Sigma^{-1}v)^{-1}v'\Sigma^{-1}w(w'\Sigma^{-1}w)^{-1}w'\Sigma^{-1}u. \end{aligned} \tag{2.6}$$

To derive the asymptotic distribution of $\tilde{\beta}_{IVG}$, we make the following assumptions:

$$\begin{aligned}
&w'\Sigma^{-1}v/n \xrightarrow{p} E[w_i'\Omega v_i]/T \equiv \bar{A}, \\
&w'\Sigma^{-1}w/n \xrightarrow{p} E[w_i'u_iu_i'w_i]/T = \bar{B}, \\
&w'\Sigma^{-1}u/n \xrightarrow{p} E[w_i\Omega u_i]/T = 0, \\
&Var(u|w) = Var(u) = I_N \otimes \Omega. \qquad (2.7)
\end{aligned}$$

Note that the condition that $E[w_i\Omega u_i] = 0$ requires that w_{it} is strictly exogenous, i.e., $E(u_{it}|w_{is}) = 0$ for all t and s. Therefore, conditions in (2.7) are stronger than those in (2.4).

Using (2.7), it is straightforward to show that

$$\sqrt{n}(\tilde{\beta}_{IVG} - \beta) \to N(0, V_G) \text{ in distribution,} \qquad (2.8)$$

where $V_G = (\bar{A}\bar{B}\bar{A}')^{-1}$ with $\bar{A}$ and $\bar{B}$ defined in (2.7).

It can be shown that $V_G - V_O$ is negative semidefinite (e.g., White (1984)). Hence, $\tilde{\beta}_{IVG}$ is asymptotically more efficient than $\tilde{\beta}_{IVO}$. When the conditional homoskedastic error assumption is violated, $\tilde{\beta}_{IVG}$ is still a $\sqrt{n}$-consistent estimator for β, but it will have a different asymptotic variance, and it may not be more efficient than $\hat{\beta}_{IVO}$.

Finally we propose an IV estimator that is based on the within transformation of the data. The within transformation has the advantage of computationally simple, requiring only a least squares regression on within transformed variables. Define $\xi_{it} = E(y_{it}|z_{it})$, and denote the within transformed variables by: $\tilde{y}_{it} = y_{it} - \bar{y}_{i.}$, $\tilde{v}_{it} = v_{it} - \bar{v}_{i.}$, $\tilde{w}_{it} = w_{it} - \bar{w}_{i.}$, and $\tilde{\xi}_{it} = \xi_{it} - \bar{\xi}_{i.}$, where $\bar{y}_{i.} = T^{-1}\sum_{j=1}^{N} y_{jt}$, $\bar{v}_{i.}$, $\bar{w}_{i.}$ and $\bar{\xi}_{i.}$ are similarly defined. Applying the within transformation to (2.2), we get

$$\tilde{y} - \tilde{\xi} = \tilde{v}\beta + \tilde{u}. \qquad (2.9)$$

Therefore, the IV within estimator of β is given by

$$\tilde{\beta}_{IVW} = (\tilde{v}'\tilde{w}\tilde{w}'\tilde{v})^{-1}\tilde{v}'\tilde{w}\tilde{w}'(\tilde{y} - \tilde{\xi}) = \beta + (\tilde{v}'\tilde{w}\tilde{w}'\tilde{v})^{-1}\tilde{v}'\tilde{w}\tilde{w}'\tilde{u}. \qquad (2.10)$$

We need the following conditions to derive the asymptotic distribution of $\tilde{\beta}_{IVW}$.

$$\begin{aligned}
&\tilde{w}'\tilde{v}/n \xrightarrow{p} E[\tilde{w}_i\tilde{v}_i'] \equiv \tilde{A}, \\
&\tilde{w}'\tilde{u}\tilde{u}'\tilde{w}/n \xrightarrow{p} E[w_i'\tilde{u}_i\tilde{u}_i'w_i] = \tilde{B}, \\
&\tilde{w}'\tilde{u}/n \xrightarrow{p} E[\tilde{w}_i'\tilde{u}_i] = 0. \qquad (2.11)
\end{aligned}$$

Similar to the case of $\tilde{\beta}_{IVG}$, the condition $E[\tilde{w}_i'\tilde{u}_i] = 0$ requires that w_{it} is strictly exogenous. Thus, we can see some trade-off in regularity conditions among $\tilde{\beta}_{IVO}$, β_{IVG} and $\tilde{\beta}_{IVW}$. $\tilde{\beta}_{IVO}$ requires z_{it} to be a weakly exogenous variable, while $\tilde{\beta}_{IVG}$ and $\tilde{\beta}_{IVW}$ require the stronger condition that z_{it} is strictly exogenous and the error is conditional homoskedastic.

Under conditions (2.11), it is easy to show that

$$\sqrt{n}(\tilde{\beta}_{IVW} - \beta) \to N(0, V_W) \text{ in distribution,} \tag{2.12}$$

where $V_W = \tilde{Q}^{-1}\tilde{A}\tilde{B}\tilde{A}'$, $\tilde{Q} = \tilde{A}'\tilde{A}$ with $\tilde{A}$ and $\tilde{B}$ defined in (2.11).

Li and Stengos (1996) considered IV-OLS estimation, while Li and Ullah (1998) considered IV-GLS and IV-Within estimation of a dynamic semiparametric model. Both papers assume the existence of some legitimate instrumental variables, but they do not discuss how to obtain optimal instruments. In the next subsection we discuss the issue of obtaining optimal instruments for a dynamic panel data model. The simulation results reported in section 3 show that the new IV estimators proposed in this paper perform much better than those considered in Li and Stengos (1996), and Li and Ullah (1998).

2.2 The Choice of Instrumental Variables

We now turn to the important problem of how to choose the instrumental variable w_{it}. Consider the simple case that both x_{it} and z_{it} are scalars, $x_{it} = y_{i,t-1}$ and z_{it} is exogenous. In this case, Li and Stengos (1996) suggest using $w_{LS,it} = z_{i,t-1}$ as an IV for $v_{it} = y_{i,t-1} - E(y_{i,t-1}|z_{it})$ because $z_{i,t-1}$ is uncorrelated with u_{it} and is *possibly* correlated with v_{it}.[1] However, given that the functional form of $\theta(.)$ is unknown, $w_{LS,it} = z_{i,t-1}$ may be weakly correlated, or even uncorrelated, with v_{it}. By the exogenous assumption of z_{it}, we know that $E(u_{it}|z_{i,t-1}) = 0$, so that $w_{LS,it} = z_{i,t-1}$ is uncorrelated with the error term u_{it}. But we could also have $E[v_{it}z_{i,t-1}] = 0$, so that $w_{LS,it} = z_{i,t-1}$ is uncorrelated with v_{it}. Hence, $w_{LS,it} = z_{i,t-1}$ may not be a legitimate instrumental variable. To see this is indeed possible, consider the case that z_{it} is an i.i.d. process satisfying

$$E(z_{it}) = 0 \text{ and } E(z_{it}^3) = 0. \tag{2.13}$$

[1] Li and Stengos (1996) suggest using $z_{i,t-1} - E(z_{i,t-1}|z_{it})$ as an IV for v_{it}. However, one can write $z_{i,t-1} = [z_{i,t-1} - E(z_{i,t-1}|z_{it})] + E(z_{i,t-1}|z_{it})$, and since v_{it} is orthogonal to $E(z_{i,t-1}|z_{it})$, we have $E[v_{it}z_{i,t-1}] = E\{v_{i,t-1}[z_{i,t-1} - E(z_{i,t-1}|z_{it})]\}$. Hence, the instrumental variable $z_{i,t-1} - E(z_{i,t-1}|z_{it})$ suggested in Li and Stengos (1996) is equivalent to $z_{i,t-1}$.

Also assume that $\theta(z_{it}) = z_{it}^2$. In this case, we have

$$y_{it} = \beta y_{i,t-1} + z_{it}^2 + u_{it}. \tag{2.14}$$

From model (2.14) it is easy to see that $y_{i,t-1}$ and z_{it} are independent with each other. Hence, $E(y_{i,t-1}|z_{it}) = E(y_{i,t-1}) \overset{def}{=} \mu_y$. Using (2.13), (2.14) and the fact that $E(u_{i,t-1}|z_{i,t-1}) = 0$, we have

$$\begin{aligned} & E[v_{it}w_{LS,it}] = E[(y_{i,t-1} - \mu_y)z_{i,t-1}] = E[y_{i,t-1}z_{i,t-1}] + 0 \\ = \;& \beta E(y_{i,t-2})E(z_{i,t-1}) + E(z_{i,t-1}^3) + E(u_{i,t-1}z_{i,t-1}) = 0. \end{aligned} \tag{2.15}$$

This shows that $w_{LS,it} = z_{i,t-1}$ is not correlated with the endogenous regressor v_{it}, and hence it is not a legitimate IV. Of course we have made some strong assumptions above in order to show that $w_{LS,it}$ is not a legitimate IV. In practice, it is unlikely that $w_{LS,it}$ is uncorrelated with v_{it}. But it is possible that the correlation between $w_{LS,it}$ and v_{it} is weak so that the IV estimation suggested by Li and Stengos (1996) can lead to imprecise estimation in such situations.

Given the structure of model (2.2) with $x_{it} = y_{i,t-1}$, we know that $v_{it} = y_{i,t-1} - E(y_{i,t-1}|z_{it})$ must be correlated with *some function* of $z_{i,t-1}$, although it may not be correlated with the particular function $w_{LS,it} = z_{i,t-1}$. Newey (1990) discussed optimal IV estimation in general nonlinear regression (parametric) models with independent data. If we consider the unrealistic case that $E(y_{it}|z_{it})$ and $E(x_{it}|z_{it})$ are known. Then, $v_{it} = x_{it} - E(x_{it}|z_{it})$ is known. In this case, (2.2) is a parametric linear model, and we can use the result of Newey (1990) to obtain an optimal IV for v_{it}. Here we use the terminology of an optimal IV in the sense that the IV variable w_{it} has maximum correlation with the endogenous variable v_{it}. If we restrict ourselves to choose an IV as a function of $z_{i,t-1}$, then the optimal IV is simply the optimal projection of v_{it} on $z_{i,t-1}$, or the conditional mean function of v_{it} given $z_{i,t-1}$. Therefore, the optimal IV for v_{it} is

$$\begin{aligned} w_{it} &= E(v_{it}|z_{i,t-1}) = E[y_{i,t-1} - E(y_{i,t-1}|z_{it})|z_{i,t-1}] \\ &= E(y_{i,t-1}|z_{i,t-1}) - E[E(y_{i,t-1}|z_{it})|z_{i,t-1}]. \end{aligned} \tag{2.16}$$

In practice we do not know the conditional expectations, and we need to use some non-parametric methods to compute these conditional expectations. (2.16) is computationally very demanding because it involves a double conditional expectation $E[E(.|z_{it})|z_{i,t-1}]$. However, note that $y_{i,t-1} = [y_{i,t-1} - E(y_{i,t-1}|z_{it})] + E(y_{i,t-1}|z_{it}) \equiv v_{it} + E(y_{i,t-1}|z_{it})$, and that $E(y_{i,t-1}|z_{it})$ is orthogonal to v_{it}. Therefore,

$$\begin{aligned} E[v_{it}E(y_{i,t-1}|z_{i,t-1})] &= E\{v_{it}E[v_{it} + E(y_{i,t-1}|z_{it})|z_{i,t-1}]\} \\ &= E\{v_{it}[E(v_{it}|z_{i,t-1}) + \mu_y]\} = E\{v_{it}E[v_{it}|z_{i,t-1}]\} + \mu_y E[v_{it}] \\ &= E\{v_{it}E[v_{it}|z_{i,t-1}]\}, \end{aligned} \tag{2.17}$$

where we used the assumption that z_{it} is an i.i.d process with $E(y_{i,t-1}|z_{it}) = E(y_{i,t-1}) = \mu_y$ and the fact that $E(v_{it}) = 0$. Equation (2.17) tells us that $E(y_{i,t-1}|z_{i,t-1})$ serves as an equivalent IV as $E(v_{it}|z_{i,t-1})$. However, $E(y_{i,t-1}|z_{i,t-1})$ is based upon a single conditional expectation and hence is much easier to compute than $E(v_{it}|z_{i,t-1})$ which involves a double conditional expectation (see (2.17)). Therefore, instead of using (2.17), we suggest using

$$w_{it} = E(y_{i,t-1}|z_{i,t-1}), \tag{2.18}$$

as the instrumental variable for v_{it}.

Even when z_{it} is not an i.i.d process, say (y_{it}, z_{it}) is a stationary process in t, then it is easy to show that $E(y_{i,t-1}|z_{i,t-1})$ and v_{it} are positively correlated and therefore is a legitimate IV.

In the general case we can have $x_{it} = (y_{i,t-1}, x'_{2,it})'$, where $x_{2,it}$ is exogenous. Then our IV will be $w_{it} = (w_{1,it}, w'_{2,it})'$, where $w_{1,it} = E(y_{i,t-1}|z_{i,t-1})$ and $w_{2,it} = x_{2,it} - E(x_{2,it}|z_{it})$ (which is equivalent to $x_{2,it}$).

2.3 Feasible Estimators

The estimators $\tilde{\beta}_{IVO}$, $\tilde{\beta}_{IVG}$, and $\tilde{\beta}_{IVW}$ in section 2.1 are not feasible because the conditional mean functions $E(y_{it}|z_{it})$, $E(x_{it}|z_{it})$ and $E(w_{it}|z_{it})$ are unknown. The feasible estimators can be obtained by replacing the unknown conditional mean functions by some nonparametric estimators, say the nonparametric kernel estimators.

Following Robinson (1988), we use the kernel estimation method to estimate the unknown conditional expectations. Specifically we estimate $f(z_{it})$, $\xi_{it} \equiv E(y_{it}|z_{it})$, $E(x_{it}|z_{it})$, and $w_{it} = E(x_{it}|z_{i,t-1})$ by $\hat{f}_{it}$, $\hat{\xi}_{it}$, $\hat{x}_{it}$ and $\hat{w}_{it}$, where

$$\hat{f}_{it} = \frac{1}{NTh^d} \sum_{j=1}^{N} \sum_{s=1}^{T} K_{it,js}, \tag{2.19}$$

$$\hat{\xi}_{it} \equiv \hat{E}(y_{it}|z_{it}) = \frac{1}{NTh^d} \sum_{j=1}^{N} \sum_{s=1}^{T} y_{js} K_{it,js}/\hat{f}_{it}, \tag{2.20}$$

$$\hat{x}_{it} \equiv \hat{E}(x_{it}|z_{it}) = \frac{1}{NTh^d}\sum_{j=1}^{N}\sum_{s=1}^{T} x_{js}K_{it,js}/\hat{f}_{it}, \tag{2.21}$$

and

$$\hat{w}_{it} = (\hat{w}_{1,it}, \hat{w}'_{2,it})', \tag{2.22}$$

with

$$\hat{w}_{1,it} \equiv \hat{E}(y_{i,t-1}|z_{i,t-1}) = \frac{1}{NTh^d}\sum_{j=1}^{N}\sum_{s=1}^{T} y_{j,s-1}K_{i,t-1,js}/\hat{f}_{i,t-1}, \tag{2.23}$$

$$\hat{w}_{2,it} \equiv x_{2,it} - \hat{E}(x_{2,it}|z_{i,t-1}) = x_{2,it} - \frac{1}{NTh^d}\sum_{j=1}^{N}\sum_{s=1}^{T} x_{2,js}K_{i,t-1,js}/\hat{f}_{i,t-1}, \tag{2.24}$$

where $K_{it,js} = K((z_{it} - z_{js})/h)$ is the kernel function and h is the smoothing parameter.

We estimate $v_{it} = x_{it} - E(x_{it}|z_{it})$ by $\hat{v}_{it} = x_{it} - \hat{x}_{it}$. Then in vector-matrix notation, the feasible version of $\tilde{\beta}_{IVO}$ is obtained from (2.3) by replacing $\xi_{it} \equiv E(y_{it}|z_{it})$, $v_{it} = x_{it} - E(x_{it}|z_{it})$, and w_{it} by their kernel estimators $\hat{\xi}_{it}$, $\hat{v}_{it} = x_{it} - \hat{x}_{it}$ and $\hat{w}_{it}$, respectively.

$$\hat{\beta}_{IVO} = [\hat{v}'\hat{w}\hat{w}'\hat{v}]^{-1}\hat{v}'\hat{w}\hat{w}'(y - \hat{\xi}). \tag{2.25}$$

For a feasible version of $\tilde{\beta}_{IVG}$, we also need a consistent estimator of $\Sigma = I_N \otimes \Omega$. If $u_{it} = \mu_i + \nu_{it}$, follows an one-way error component model, where μ_i is i.i.d. $(0, \sigma_\mu^2)$, ν_{it} is i.i.d. $(0, \sigma_\nu^2)$, μ_i and ν_{it} are independent with each other. Then we have $\Omega = \sigma_1^2\bar{J}_T + \sigma_\nu^2 E_T$, and $\Omega^{-1} = (1/\sigma_1^2)\bar{J}_T + (1/\sigma_\nu^2)E_T$, where $\bar{J}_T = J_T/T$, J_T is a $T \times T$ matrix with all elements equal to one, and $E_T = I_T - \bar{J}_T$. In this case, a consistent estimator of Σ^{-1} is $\hat{\Sigma}^{-1} = I_N \otimes \hat{\Omega}^{-1}$ with $\hat{\Omega}^{-1} = (1/\hat{\sigma}_1^2)\bar{J}_T + (1/\hat{\sigma}_\nu^2)E_T$, $\hat{\sigma}_1^2 = \hat{u}'[I_N \otimes \bar{J}_T]\hat{u}/T$, $\hat{\sigma}_\nu^2 = \hat{u}'[I_N \otimes E_T]\hat{u}/[N(T-1)]$, and $\hat{u} = y - x\hat{\beta}_{IVO}$.

Hence, the feasible IV-GLS estimator is given by

$$\hat{\beta}_{IVG} = (\hat{v}'\hat{\Sigma}^{-1}\hat{w}(\hat{w}'\hat{\Sigma}^{-1}\hat{w})^{-1}\hat{w}'\hat{\Sigma}^{-1}\hat{v})^{-1}\hat{v}'\hat{\Sigma}^{-1}\hat{w}(\hat{w}'\hat{\Sigma}^{-1}\hat{w})^{-1}\hat{w}'\hat{\Sigma}^{-1}(y - \hat{\xi}), \tag{2.26}$$

For a feasible version of $\tilde{\beta}_{IVW}$, we need to replace $\tilde{v}_{it} = v_{it} - T^{-1}\sum_{s=1}^{T} v_{is}$, $\tilde{w}_{it} = w_{it} - T^{-1}\sum_{s=1}^{T} w_{is}$, $\tilde{\xi}_{it} = E(y_{it}|z_{it}) - T^{-1}\sum_{s=1}^{T} E(y_{it}|z_{it})$, by their corresponding estimators: $\hat{v}_{it} - T^{-1}\sum_{s=1}^{T}\hat{v}_{is}$, $\hat{w}_{it} - T^{-1}\sum_{s=1}^{T}\hat{w}_{is}$, and $\hat{y}_{it} - T^{-1}\sum_{s=1}^{T}\hat{y}_{it}$, respectively. We will use the notation $\tilde{v}_f$, $\tilde{w}_f$ and $\tilde{\xi}_f$ to denote these feasible estimators. A feasible version of $\tilde{\beta}_{IVW}$ is given by

$$\hat{\beta}_{IVW} = (\tilde{v}'_f\tilde{w}_f\tilde{w}'_f\tilde{v}_f)^{-1}\tilde{v}'_f\tilde{w}_f\tilde{w}'_f(\tilde{y} - \tilde{\xi}_f). \tag{2.27}$$

The asymptotic distributions of $\sqrt{n}(\hat{\beta}_{IVO}-\beta)$, $\sqrt{n}(\hat{\beta}_{IVG}-\beta)$, and $\sqrt{n}(\hat{\beta}_{IVW}-\beta)$ are the same as those given by (2.5), (2.8) and (2.12), respectively. The proofs of these results are similar to the proof of Theorem 1 in Li and Stengos (1996), and are therefore omitted here.

After obtaining a $\sqrt{n}$ consistent estimator of β, say $\hat{\beta}$ ($\hat{\beta}$ can be $\hat{\beta}_{IVO}$, $\hat{\beta}_{IVG}$ or $\hat{\beta}_{IVW}$), one can estimate $\theta(z)$ based upon

$$y_{it} - x_{it}'\hat{\beta} = \theta(z_{it}) + u_{it} + x_{it}'(\beta - \hat{\beta}) \equiv \theta(z_{it}) + \epsilon_{it}, \tag{2.28}$$

where $\epsilon_{it} = u_{it} + x_{it}'(\beta - \hat{\beta})$. A nonparametric kernel estimator of $\theta(z)$ is given by

$$\hat{\theta}(z) = \frac{\sum_i \sum_t (y_{it} - x_{it}'\hat{\beta}) K(\frac{z_{it}-z}{h})}{\sum_i \sum_t K(\frac{z_{it}-z}{h})}. \tag{2.29}$$

Because $\hat{\beta} - \beta = O_p(n^{-1/2})$, which is faster than the usual nonparametric convergence rate, it can be easily shown that $\hat{\theta}(z)$ has the same asymptotic distribution as the case where β is known. We will not present the asymptotic distribution of $\hat{\theta}(z)$ here since it is a standard nonparametric estimation result.

3 Monte Carlo Results

We consider the following data generating process (DGP):

$$\begin{aligned} y_{it} &= \beta y_{i,t-1} + \gamma_1 z_{it} + \gamma_2 z_{it}^2 + \mu_i + \nu_{it} \\ &= \beta y_{i,t-1} + \theta(z_{it}) + \mu_i + \nu_{it}. \end{aligned} \tag{2.30}$$

where z_{it} is independent and uniformly distributed in the interval of $[-\sqrt{3}, \sqrt{3}]$, μ_i is i.i.d. $N(0, \sigma_\mu^2)$, ν_i is i.i.d. $N(0, \sigma_\nu^2)$. We fix the total error variance $\sigma^2 \stackrel{def}{=} \sigma_\mu^2 + \sigma_\nu^2 = 10$. Define $\rho = \sigma_\mu^2/\sigma^2$. We choose $\rho = 0.2$, 0.5 and 0.8. β is fixed at $\beta = 0.5$, and we consider three sets of values of $(\gamma_1, \gamma_2) = (1, 0)$, $(1, 1)$ and $(0, 1)$.

For comparison, we use two sets of IV's in the simulations. The first set is $w_{it} = E(y_{i,t-1}|z_{i,t-1})$ as suggested in this paper, and the resulting estimators are denoted as $\hat{\beta}_{IVO1}$, $\hat{\beta}_{IVG1}$, and $\hat{\beta}_{IVW1}$, respectively. The second set is the one suggested in Li and Stengos (1996) using $w_{it} = z_{i,t-1} - E(z_{i,t-1}|z_{it})$, and the resulting estimators are denoted as $\hat{\beta}_{IVO2}$, $\hat{\beta}_{IVG2}$, and $\hat{\beta}_{IVW,2}$, respectively. We report the estimated bias, standard deviation (Std), and root mean square errors (Rmse) for all the estimators. These are computed via $Bias(\hat{\beta}) = M^{-1}\sum_{j=1}^{M}(\hat{\beta}_j - \beta)$, $Std(\hat{\beta}) = \{M^{-1}\sum_{j=1}^{M}[\hat{\beta}_j - Mean(\hat{\beta})]^2\}^{1/2}$, and $Rmse(\hat{\beta}) = \{M^{-1}\sum_{j=1}^{M}(\hat{\beta}_j - \beta)^2\}^{1/2}$, where M

is the number of replications. We use $M = 2000$ in all experiments. We choose $T = 6$, $N = 100$ and 200.

When $(\gamma_1, \gamma_2) = (0, 1)$, only z_{it}^2 is present in $\theta(z_{it})$, and the instrument suggested by Li and Stengos (1996), i.e., $w_{LS,it} = z_{i,t-1} - E(z_{i,t-1}|z_{it})$ is uncorrelated with the endogenous variable $v_{it} = y_{i,t-1} - E(y_{i,t-1}|z_{it})$. In this case, $\hat{\beta}_{IVO2}$, $\hat{\beta}_{IVG2}$, and $\hat{\beta}_{IVW2}$ are inconsistent estimators for β. Indeed as expected in this case, $\hat{\beta}_{IVO2}$, $\hat{\beta}_{IVG2}$, and $\hat{\beta}_{IVW2}$ all have very large variances and hence large MSEs. In contrast, our new IV estimators $\hat{\beta}_{IVO1}$, $\hat{\beta}_{IVG1}$, and $\hat{\beta}_{IVW1}$ all perform quite well.

When $(\gamma_1, \gamma_2) = (1, 1)$, both z_{it} and z_{it}^2 are present in $\theta(z_{it})$, and the instrument suggested by Li and Stengos (1996) is a legitimate IV, but it is not an optimal IV because it does not capture the z_{it}^2 component of $\theta(z_{it})$. Our new IV estimator $\hat{\beta}_{IVO1}$ outperforms $\hat{\beta}_{IVO2}$. Also, $\hat{\beta}_{IVG1}$ outperforms $\hat{\beta}_{IVG2}$, and $\hat{\beta}_{IVW1}$ outperforms $\hat{\beta}_{IVW2}$.

When $(\gamma_1, \gamma_2) = (1, 0)$, only z_{it} is present in $\theta(z_{it})$, and the instrument suggested by Li and Stengos (1996) is in fact optimal for the DGP considered here (since $\theta(z_{it}) = z_{it}$). As expected all the estimators perform well in this case.

Summarizing the simulation results, we see that our optimal IV estimators dominate the IV estimators suggested by Li and Stengos (1996). Next, we compare the relative MSE performance of our IV-OLS, IV-GLS and Within estimators ($\hat{\beta}_{IVO1}$, $\hat{\beta}_{IVG1}$ and $\hat{\beta}_{IVW1}$).

When ρ is small ($\rho = 0.2$), $\hat{\beta}_{IVO1}$ has a slightly smaller MSE than those of $\hat{\beta}_{IVG}$ and $\hat{\beta}_{IVW}$. For a large value of ρ ($\rho = 0.8$), $\hat{\beta}_{IVG1}$ $\hat{\beta}_{IVW1}$ enjoy much smaller MSEs than that of $\hat{\beta}_{IVO1}$. This is as expected since $\hat{\beta}_{IVG1}$ takes into account the variance structure, and the within transformation wipes out the individual effects whether ρ is small or large. When ρ is large, Σ is more accurately estimated, and it also implies a smaller within transformed errors. Hence, it leads to smaller MSEs for $\hat{\beta}_{IVG1}$ and $\hat{\beta}_{IVW1}$. On the other hand, $\hat{\beta}_{IVO1}$ is an OLS-type IV estimator, which ignores the variance structure of the error term. Hence when ρ gets larger, the MSE of $\hat{\beta}_{IVO1}$ also gets larger.

It is interesting to observe that the performance of $\hat{\beta}_{IVG1}$ and $\hat{\beta}_{IVW1}$ are almost identical. Given that $\hat{\beta}_{IVW1}$ is computational simpler than $\hat{\beta}_{IVG1}$, we recommend the use of $\hat{\beta}_{IVW1}$ in practice.

References

Horowitz, J.L. and M. Markatou (1996). "Semiparametric estimation of regression models for panel data," *The Review of Economic Studies* 63, 145-168.

Li, Q. and T. Stengos (1996). "Semiparametric estimation of partially linear panel data models," *Journal of Econometrics* 71, 389-397.

Li, Q. and C. Hsiao (1998). "Testing serial correlation in semiparametric panel data models," *Journal of Econometrics* 87, 207-237.

Li, Q. and A. Ullah (1998). "Estimating partially linear models with one-way error components," *Econometric Reviews* 17, 145-166.

Newey, W. (1990). "Efficient instrumental variable estimation of nonlinear models," *Econometrica* 58, 809-837.

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

Ullah, A. and N. Roy (1998). Nonparametric and semiparametric econometrics of panel data. In: A. Ullah and D.E.A. Giles (Eds). *Handbook on Applied Economic Statistics* (pp.579-604), Ch. 17. Marcel Dekker.

Ullah, A. and K. Mundra (1999). *Semiparametric panel data estimation: An application to immigrates homelink effect on U.S. producer trade flows* Working paper 15, Department of Economics, University of California at Riverside.

White, H. (1984). *Asymptotic Theory for Econometricians* (Academic Press, New York).

Table 1: $(\gamma_1, \gamma_2) = (0, 1)$

	$N = 100$								
	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
	Bias	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*
$\hat{\beta}_{IVO1}$	.025	.141	.143	.064	.127	.143	.104	.105	.148
$\hat{\beta}_{IVG1}$	−.030	.165	.167	−.023	.139	.141	−.016	.096	.097
$\hat{\beta}_{IVW1}$	−.038	.168	.172	−.030	.140	.143	−.020	.095	.097
$\hat{\beta}_{IVO2}$	−.530	46.6	46.6	.539	50.8	50.8	.518	13.9	13.9
$\hat{\beta}_{IVG2}$	.015	29.3	29.3	−1.129	31.0	31.0	.462	57.7	57.7
$\hat{\beta}_{IVW2}$	.132	18.3	18.3	.204	30.8	30.8	2.90	100.9	100.9

Table 2: $(\gamma_1, \gamma_2) = (1, 1)$

	$N = 100$								
	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
	Bias	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*
$\hat{\beta}_{IVO1}$	.010	.098	.099	.025	.096	.099	.041	.090	.099
$\hat{\beta}_{IVG1}$	−.018	.109	.110	−.016	.090	.091	−.014	.060	.062
$\hat{\beta}_{IVW1}$	−.021	.111	.113	−.018	.090	.092	−.015	.060	.062
$\hat{\beta}_{IVO2}$	−.009	.139	.140	−.023	.157	.158	−.042	.218	.222
$\hat{\beta}_{IVG2}$	−.018	.109	.110	−.016	.090	.091	−.010	.077	.078
$\hat{\beta}_{IVW2}$	−.007	.151	.151	−.008	.119	.119	−.010	.077	.078

Table 3: $(\gamma_1, \gamma_2) = (1, 0)$

	$N = 100$								
	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
	Bias	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*
$\hat{\beta}_{IVO1}$	.018	.125	.126	.044	.118	.126	.071	.105	.126
$\hat{\beta}_{IVG1}$	−.010	.143	.143	−.005	.120	.120	−.001	.082	.082
$\hat{\beta}_{IVW1}$	−.014	.147	.148	−.008	.121	.121	−.003	.082	.082
$\hat{\beta}_{IVO2}$	−.008	.139	.139	−.023	.158	.159	−.049	.528	.530
$\hat{\beta}_{IVG2}$	−.006	.146	.146	−.007	.117	.117	−.009	.077	.077
$\hat{\beta}_{IVW2}$	−.006	.150	.151	−.008	.118	.118	−.010	.076	.077

Table 4: $(\gamma_1, \gamma_2) = (0, 1)$

	$N = 200$								
	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
	Bias	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*
$\hat{\beta}_{IVO1}$	.018	.096	.098	.041	.91	.100	.063	.082	.103
$\hat{\beta}_{IVG1}$	−.021	.113	.115	−.018	.093	.095	−.015	.063	.065
$\hat{\beta}_{IVW1}$	−.026	.115	.118	−.021	.094	.096	−.017	.063	.065
$\hat{\beta}_{IVO2}$	−.709	24.3	24.3	.747	21.7	21.7	.504	12.6	12.6
$\hat{\beta}_{IVG2}$	.463	47.9	47.9	2.59	58.6	58.6	−.282	10.2	10.2
$\hat{\beta}_{IVW2}$	−1.62	52.7	52.7	−3.06	114.9	114.9	1.89	72.6	72.6

Table 5: $(\gamma_1, \gamma_2) = (1, 1)$

	$N = 200$								
	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
	Bias	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*
$\hat{\beta}_{IVO1}$	.007	.068	.069	.016	.067	.069	.026	.064	.069
$\hat{\beta}_{IVG1}$	−.010	.143	.143	−.012	.063	.064	−.001	.042	.043
$\hat{\beta}_{IVW1}$	−.016	.078	.079	−.014	.063	.064	−.012	.042	.043
$\hat{\beta}_{IVO2}$	−.004	.097	.097	−.010	.101	.102	−.016	.107	.108
$\hat{\beta}_{IVG2}$	−.006	.146	.146	−.007	.117	.117	−.009	.077	.077
$\hat{\beta}_{IVW2}$	−.005	.078	.079	−.006	.084	.084	−.007	.055	.055

Table 6: $(\gamma_1, \gamma_2) = (1, 0)$

	$N = 200$								
	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
	Bias	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*	*Bias*	*Std*	*Rmse*
$\hat{\beta}_{IVO1}$	.010	.092	.092	.026	.088	.092	.043	.081	.092
$\hat{\beta}_{IVG1}$	−.008	.102	.103	−.005	.085	.085	−.002	.057	.057
$\hat{\beta}_{IVW1}$	−.010	.104	.105	−.007	.085	.085	−.003	.056	.057
$\hat{\beta}_{IVO2}$	−.004	.097	.097	−.101	.101	.101	−.016	.106	.107
$\hat{\beta}_{IVG2}$	−.006	.146	.146	−.007	.117	.117	−.009	.077	.077
$\hat{\beta}_{IVW2}$	−.005	.105	.105	−.006	.084	.084	−.007	.054	.055