Econometric Theory, **21**, 2005, 865–869. Printed in the United States of America.
DOI: 10.1017/S0266466605050437

NOTES AND PROBLEMS

INSTRUMENTAL VARIABLES ESTIMATION WITH PANEL DATA

JEFFREY M. WOOLDRIDGE
Michigan State University

The system two stage least squares estimator for the linear panel data model is shown to have different characterizations depending on the choice of instrument matrix. The more general estimator, where, in effect, separate reduced form linear projections are estimated for each time period, also has the advantage of being applicable when the number of instruments changes across time periods. The issue of efficient estimation is also treated.

1. MOTIVATION AND RESULTS

This problem explores how different choices of instrument matrices lead to different two stage least squares (2SLS) estimators in estimating models with panel data. In addition to providing a simple characterization of the different estimators—with one system 2SLS estimator allowing a different number of instruments for each time period—this problem provides a simple way to compute an instrumental variables (IV) estimator when the dimension of the instrument set changes with time.

The model is a standard linear panel data model,

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \qquad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{x}_{it}$ is $1 \times K$ for all t. For each t, we have a vector of instruments, $\mathbf{z}_{it}$, which is $1 \times L_t$, $L_t \geq K$. We assume that $\mathrm{E}(\mathbf{z}_{it}' u_{it}) = \mathbf{0}, t = 1, \ldots, T$, so that the instruments are contemporaneously exogenous but not necessarily strictly exogenous. We allow the dimension of $\mathbf{z}_{it}$ to change with t to encompass models where instruments used in some time periods are not necessarily exogenous in other time periods. For example, equation (1) could be in first-differenced form, as in Arellano and Bond (1991). Then, y_{it} is the first difference of a response variable and $\mathbf{x}_{it}$ contains a lagged first difference, and the instruments for time

Address correspondence to Jeffrey M. Wooldridge, Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA; e-mail: wooldri1@msu.edu.

© 2005 Cambridge University Press 0266-4666/05 $12.00

https://doi.org/10.1017/S0266466605050437 Published online by Cambridge University Press

t can include the observed history of the response variable up through time $t-2$. For the asymptotic analysis, we assume fixed T with $N \to \infty$.

Define the $T \times K$ matrix of explanatory variables $\mathbf{X}_i$ as

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{iT} \end{pmatrix} \tag{2}$$

and initially the matrix of instruments for observation i as

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{i2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{z}_{iT} \end{pmatrix}. \tag{3}$$

It is convenient to break the problem into several parts.

(i) Assuming that it exists, the *system* 2SLS estimator of $\boldsymbol{\beta}$ (for any choice of $\mathbf{Z}_i$) is

$$\hat{\boldsymbol{\beta}} = \left[\left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Z}_i \right) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{X}_i \right) \right]^{-1} \times \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Z}_i \right) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{y}_i \right); \tag{4}$$

see Wooldridge (2002, equation (8.29)).

Show that for the definitions of $\mathbf{X}_i$ and $\mathbf{Z}_i$ given previously, $\hat{\boldsymbol{\beta}}$ can be obtained as follows:

(a) For each time period t, run separate reduced form regressions, $\mathbf{x}_{it}$ on $\mathbf{z}_{it}$, $i = 1, \ldots, N$, and obtain the fitted values, $\hat{\mathbf{x}}_{it}, i = 1, \ldots, N$. (Remember, for each i and t, the fitted value is a $1 \times K$ vector regardless of the dimension of $\mathbf{z}_{it}$.)

(b) Obtain $\hat{\boldsymbol{\beta}}$ by applying pooled IV to (1) using instruments $\hat{\mathbf{x}}_{it}$ for $\mathbf{x}_{it}$. In other words,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\mathbf{x}}_{it}' \mathbf{x}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\mathbf{x}}_{it}' y_{it} \right). \tag{5}$$

(ii) How would you estimate the asymptotic variance of $\hat{\boldsymbol{\beta}}$ in (5)?

https://doi.org/10.1017/S0266466605050437 Published online by Cambridge University Press

(iii) If L_t is the same for all t—a common situation in simultaneous equations models with panel data, or measurement error in one or more of the x_{itj}—a different choice of $\mathbf{Z}_i$ is possible:

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} \\ \mathbf{z}_{i2} \\ \vdots \\ \mathbf{z}_{iT} \end{pmatrix}. \tag{6}$$

If we use this choice of $\mathbf{Z}_i$ in equation (4), how can we describe the resulting estimator? In particular, how does it differ from the estimator that uses (3) as the IV matrix?

(iv) How would you obtain an estimator more asymptotically efficient than either of the estimators in (i) or (iii)?

2. PROOFS AND DISCUSSION

The proofs of the results require only straightforward matrix algebra and least squares mechanics. Of perhaps more interest is what the results have to say for how to use instruments in panel data settings.

(i) Straightforward partitioned multiplication shows that

$$\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{Z}_i = \begin{pmatrix} \sum_{i=1}^{N} \mathbf{z}_{i1}'\mathbf{z}_{i1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^{N} \mathbf{z}_{i2}'\mathbf{z}_{i2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sum_{i=1}^{N} \mathbf{z}_{iT}'\mathbf{z}_{iT} \end{pmatrix},$$

$$\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{X}_i = \begin{pmatrix} \sum_{i=1}^{N} \mathbf{z}_{i1}'\mathbf{x}_{i1} \\ \sum_{i=1}^{N} \mathbf{z}_{i2}'\mathbf{x}_{i2} \\ \vdots \\ \sum_{i=1}^{N} \mathbf{z}_{iT}'\mathbf{x}_{iT} \end{pmatrix} \quad \text{and} \quad \sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{y}_i = \begin{pmatrix} \sum_{i=1}^{N} \mathbf{z}_{i1}'y_{i1} \\ \sum_{i=1}^{N} \mathbf{z}_{i2}'y_{i2} \\ \vdots \\ \sum_{i=1}^{N} \mathbf{z}_{iT}'y_{iT} \end{pmatrix}.$$

https://doi.org/10.1017/S0266466605050437 Published online by Cambridge University Press

Now, block-partitioned inversion and multiplication give

$$\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)$$
$$=\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\mathbf{x}_{it}'\mathbf{z}_{it}\right)\left(\sum_{i=1}^{N}\mathbf{z}_{it}'\mathbf{z}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{z}_{it}'\mathbf{x}_{it}\right)$$
$$=\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{\mathbf{x}}_{it}'\mathbf{x}_{it}=\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{x}}_{it}'\mathbf{x}_{it},$$

where $\hat{\mathbf{x}}_{it}=\mathbf{z}_{it}\hat{\boldsymbol{\Pi}}_t$ and $\hat{\boldsymbol{\Pi}}_t=(\sum_{i=1}^{N}\mathbf{z}_{it}'\mathbf{z}_{it})^{-1}(\sum_{i=1}^{N}\mathbf{z}_{it}'\mathbf{x}_{it})$ is the matrix of first-stage regression coefficients (for time period t). Similarly,

$$\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{y}_i\right)=\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{\mathbf{x}}_{it}'y_{it}=\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{x}}_{it}'y_{it}.$$

Putting these results together gives the result. As a practical matter, in virtually any econometrics package it is easy to obtain the fitted values $\hat{\mathbf{x}}_{it}$ for each time period and then obtain $\hat{\boldsymbol{\beta}}$ as the pooled IV estimator, listing $\hat{\mathbf{x}}_{it}$ as instruments for $\mathbf{x}_{it}$.

(ii) As with cross-sectional IV, estimation of the linear projection $L(\mathbf{x}_{it}|\mathbf{z}_{it})=\mathbf{z}_{it}\boldsymbol{\Pi}_t$ by $\hat{\mathbf{x}}_{it}=\mathbf{z}_{it}\hat{\boldsymbol{\Pi}}_t$ does not affect the first-order asymptotic properties of IV. (Technically, $N^{-1/2}\sum_{i=1}^{N}(\mathbf{z}_{it}\hat{\boldsymbol{\Pi}}_t)'u_{it}=N^{-1/2}\sum_{i=1}^{N}(\mathbf{z}_{it}\boldsymbol{\Pi}_t)'u_{it}+o_p(1)$.) Nevertheless, without further assumptions, the errors $\{u_{it}:t=1,2,\ldots,T\}$ can be arbitrarily serially correlated and heteroskedastic. Let $\hat{u}_{it}=y_{it}-\mathbf{x}_{it}\hat{\boldsymbol{\beta}}$ denote the IV residuals where, as usual, these depend on $\mathbf{x}_{it}$, not $\hat{\mathbf{x}}_{it}$. Then

$$\widehat{\text{Avar}\sqrt{N}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}=\left(N^{-1}\sum_{i=1}^{N}\hat{\mathbf{X}}_i'\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\hat{\mathbf{X}}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\hat{\mathbf{X}}_i\right)$$
$$\times\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{X}}_i\right)^{-1}$$
$$=\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{x}}_{it}'\mathbf{x}_{it}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\hat{u}_{it}\hat{u}_{ir}\hat{\mathbf{x}}_{it}'\hat{\mathbf{x}}_{ir}\right)$$
$$\times\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\hat{\mathbf{x}}_{it}\right)^{-1}, \tag{7}$$

where the notation is hopefully clear (e.g., $\hat{\mathbf{u}}_i$ is the $T\times 1$ vector of IV residuals for each i). Therefore, $\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}})$ is the preceding expression but without the divisions by N. This matrix is fully robust to serial correlation and heteroskedasticity in the u_{it}, and it is a special case of formula (8.27) in Wooldridge (2002). Conveniently, this is exactly the matrix that would be computed by using "cluster" options in standard econometrics packages when using pooled IV esti-

https://doi.org/10.1017/S0266466605050437 Published online by Cambridge University Press

mation with instruments $\hat{\mathbf{x}}_{it}$, where each cross-sectional observation acts as its own cluster.

(iii) With $\mathbf{Z}_i$ as in (6), the system 2SLS estimator—say, $\tilde{\boldsymbol{\beta}}$—is easily seen to be

$$\tilde{\boldsymbol{\beta}} = \left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{z}_{it}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}'\mathbf{z}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}'\mathbf{x}_{it}\right)\right]^{-1} \times \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{z}_{it}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}'\mathbf{z}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}'\mathbf{y}_{it}\right). \tag{8}$$

Mechanically, this is the *pooled* 2SLS estimator of (1) using instruments $\mathbf{z}_{it}$ for $\mathbf{x}_{it}$. A simple way to characterize how it differs from the estimator in part (i) is to look at its implicit first-stage regression. The estimator in (8) effectively obtains the first-stage fitted values by pooling across i and t. That is, the first-stage fitted values, say $\tilde{\mathbf{x}}_{it}$, are obtained from the *pooled* regression $\mathbf{x}_{it}$ on $\mathbf{z}_{it}, i = 1,\ldots,N; t = 1,\ldots,T$. As we saw, the estimator that uses (3) as the matrix of instruments implicitly estimates a different first-stage regression for each t. (Of course, when $\dim(\mathbf{z}_{it})$ changes across t, it makes no sense to do a pooled regression in the first stage. When $\dim(\mathbf{z}_{it})$ is constant across t, one has the option of doing a pooled first stage or separate first-stage regressions.)

(iv) The efficient estimator that uses only the moments $\mathrm{E}(\mathbf{z}_{it}'u_{it}) = \mathbf{0}, t = 1,\ldots,T$ is the generalized method of moments (GMM) estimator with an optimal weighting matrix—also called the *minimum chi-square* estimator. (The choice of instrument matrix in (6) means we are only using the moment conditions aggregated across time, $\sum_{t=1}^{T}\mathrm{E}(\mathbf{z}_{it}'u_{it}) = \mathbf{0}$.) The matrix of instruments should be as in (3), because this expresses the full set of moment conditions, and an optimal weighting matrix replaces $(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i)^{-1}$. The system 2SLS estimator can be used in the first step to obtain the optimal weighting matrix. See, for example, Wooldridge (2002, Sect. 8.3.3). GMM with the optimal weight matrix is what Arellano and Bond (1991) propose for the AR(1) model with unobserved effects.

REFERENCES

Arellano, M. & S.R. Bond (1991) Some specification tests for panel data models: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277–298.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

https://doi.org/10.1017/S0266466605050437 Published online by Cambridge University Press