Reducing the Asymptotic Bias of Weak Instruments Estimation Using Independently Repeated Cross-sectional Information

Zongwu Cai\textsuperscript{a,b} and Ying Fang\textsuperscript{b}

\textsuperscript{a} Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA.
\textsuperscript{b} Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen, Fujian Province, 361005, China.

Abstract: We show that independently repeated cross-sectional data can reduce the asymptotic bias when instruments are weakly correlated to the endogenous variables. When both $N$ and $T$ go to infinite, we can obtain consistent estimators even if instruments are weak.

Keywords: Bias reduction; Weak instruments; Panel data.

JEL classification: C13 C33

Correspondence author: Zongwu Cai Tel. (704) 687-2650, Fax (704) 687-6415, E-mail zcai@uncc.edu.

This version: Oct. 22, 2008

\textsuperscript{*} Cai acknowledges funds provided by the University of North Carolina at Charlotte.
Reducing the Asymptotic Bias of Weak Instruments Estimation
Using Independently Repeated Cross-sectional Information

1. Introduction

It is well known that the two-stage least square (TSLS) estimator and the limited information maximum likelihood (LIML) estimator provide a poor approximation in a cross-sectional model when instruments are weakly correlated to the endogenous explanatory variables (e.g., Nelson and Startz, 1990a, 1990b; Bound, etc., 1995; Staiger and Stock, 1997; Stock, Wright and Yogo, 2002; among others). Chao and Swanson (2005) obtain consistent estimation when available instruments are weak and the number of instruments goes to infinity with the sample size. However, they find that the TSLS estimator achieves consistency under more stringent condition than that of the LIML estimator. In this paper, we consider to reduce the asymptotic bias in an independently repeated cross-sectional dataset when instruments are weakly correlated to the endogenous variables. Indeed, we show that the bias term has the order of $O(T^{-1})$. When both $N$ and $T$ tend to infinity, the consistent estimation of the TSLS and LIML estimators can be achieved. Finally, we conduct a simple Monte Carlo simulation to illustrate the finite sample performance.

2. The Model and the Estimation Method

Without loss of generality, we consider the following simple simultaneous equations model in an independently repeated cross-sectional data set:

\[ y_{it} = \alpha_i + \beta^T Y_{it} + u_{it} \]  
\[ Y_{it} = \Pi T Z_{it} + V_{it} \]

where $1 \leq i \leq N,  1 \leq t \leq T$, $y_{it}$ is a scalar dependent variable, $Y_{it}$ is a $p \times 1$ vector of endogenous variables, $Z_{it}$ is a $q \times 1$ ($q \geq p$) vector of excluded instruments, and $\{ \alpha_i \}$ is independent across individuals $i$. We assume that $\{Z_{it}, u_{it}\}$ and $\{Z_{it}, V_{it}\}$ are independent across both $N$ and $T$. Following the local-to-zero asymptotics (Staiger and Stock, 1997), we assume that

\[ \Pi = C / \sqrt{N} \]  

It is well known that the weak instruments problem doesn’t affect the consistent estimation of the coefficients of included exogenous variables. To save notations, we focus on the simple model without any included exogenous variables. A general model with included exogenous variables can be simplified to the above model by projecting out them.
where $C$ is a $q \times p$ matrix of constants that contain in a compact set. To remove the individual effect $\{\alpha_i\}$, both equations are multiplied by the forward orthogonal deviations operator $A$ (Arellano, 2003), where $A^T A = I_T - e e^T / T$, $A A^T = I_{T-1}$, $I_T$ is an identity matrix with dimension $T$ and $e$ is a vector of ones. The transformed model can be represented as

$$y^*_i = Y^*_i \beta + u^*_i$$

(4)

$$Y^*_i = Z^*_i \Pi + V^*_i$$

(5)

where $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})^T$, $y^*_i = Ay_i$, $Y^*_i = AY_i$, and $Z^*_i = AZ_i$ respectively. Thus, $\text{var}(u^*_i) = \sigma_u^2 I_{T-1}$ if $\text{var}(u_i) = \sigma_u^2 I_T$ in the original model. The k-class estimator is given by

$$\hat{\beta}_{k\text{-class}} = [Y^{*T} (I - kM_{Z^*}) Y^*]^{-1} [Y^{*T} (I - kM_{Z^*}) y^*]$$

(6)

where $M_X = I - X (X^T X)^{-1} X^T$. When $k_{TSL} = 1$, the k-class estimator is just a TSLS estimator. When $k_{LIML}$ is the smallest root of the determinantal equation

$$[\bar{Y}^{*T} \bar{Y}^* - k \bar{Y}^{*T} M_{Z^*} \bar{Y}^*]$$

where $\bar{Y}^* = (y^*, Y^*)$, the k-class estimator is the LIML estimator. Note that

$$y^* = (y_1^{*T}, y_2^{*T}, \ldots, y_N^{*T})^T$$

$Y^* = (Y_1^{*T}, Y_2^{*T}, \ldots, Y_N^{*T})^T$ and $Z^* = (Z_1^{*T}, Z_2^{*T}, \ldots, Z_N^{*T})^T$.

The above model has a lot of applications in empirical studies. For example, Andreoni and Payne (2003) examine whether government grants crowd out private charities by employing panel data from arts and social science organizations. They apply the TSLS estimators by using several sets of instruments, and all F-test on instruments in the first stage are relatively small, which means it possibly suffers from the weak instruments problem. Other examples using TSLS estimators in panel data include Fishback, etc. (2002), Gruber and Hungerman (2007), and Andreoni and Payne (2007).

3. Large Sample Theory

In this section, we show that the asymptotic bias of the $k$–class estimators can be reduced when the number of independently repeated cross sections $T$ increases. As $T$ goes to infinity, we can achieve consistent estimator. To derive asymptotic results, we make the following assumptions.
Assumption 1: \( \Pi = C / \sqrt{N} \) where \( C \) is a fixed \( q \times p \) matrix.

Assumption 2: \( (u^T u / NT, V^T u / NT, V^T V / NT) \to^p (\sigma_u^2, \Sigma_{vv}, \Sigma_{vv}) \).

Assumption 3: \( Z^T Z / NT \to^p \Sigma_{zz} = E(Z_u Z_u^T) - E(Z_u)E(Z_u^T) \) is a finite, positive definite \( q \times q \) matrix.

Assumption 4: \( (Z^T u^*, \sqrt{NT}, Z^T V^* / \sqrt{NT}) \to^d (\Psi_{zu}, \Psi_{zv}) \), where \( \Psi_{zu} \), \( \text{vec}(\Psi_{zv}) \) is distributed \( N(0, \Sigma \otimes \Sigma_{zz}) \), \( \Sigma = \begin{pmatrix} \sigma_u^2 & \Sigma_{vu}^T \\ \Sigma_{vu} & \Sigma_{vv} \end{pmatrix} \), where \( \text{vec}(\Psi) \) is a vector formed by stacking the column of \( \Psi \) under each, and \( \otimes \) denotes the Kronecker product.

Convergences in Assumptions 2-4 are not primitive assumptions but hold under weak primitive conditions. Assumptions 2 and 3 follow from the weak law of large numbers. Assumption 4 follows from triangular arrays central limit theorems.

**Theorem 1** Suppose that Assumptions 1-4 hold for the model defined in (4) and (5). Also suppose that \( N(k - 1) \to^p \kappa_T \) as \( N \) tends to infinity, then

\[
\hat{\beta}_{k-\text{class}} - \beta_0 \to^d \frac{1}{\sqrt{T - 1}} B,
\]

where

\[
B = [D^T \Sigma_{zz}^{-1} D - \kappa_T \Sigma_{vv}]^{-1}[D^T \Sigma_{zz}^{-1} \Psi_{zu} - \kappa_T \sqrt{T - 1} \Sigma_{vu}^T],
\]

\[
D = \Sigma_{zz} C + \Psi_{zv} / \sqrt{T - 1}.
\]

**Proof.** See the Appendix.

It is obvious that \( B \) is a mixture of finite random variables. For the TSLS estimator \( (k=1) \), the result is simplified to \( B = [D^T \Sigma_{zz}^{-1} D]^{-1} D^T \Sigma_{zz}^{-1} \Psi_{zu} \). Theorem 1 shows that the asymptotic bias shrinks as \( T \) becomes large. To understand this result, we remind you the so called concentration parameter, a measure of strength of instruments in the literature. The concentration parameter is defined as

\[
\Sigma_{vv}^{-1/2} \Pi T Z^T Z \Pi \Sigma_{vv}^{-1/2} \to^p (T - 1)\Sigma_{vv}^{-1/2} C^T \Sigma_{zz} C \Sigma_{vv}^{-1/2}
\]

which grows as \( T \) increases. When \( T \) goes to infinity, the concentration parameter also increases to infinity. On the other hand, to see how fast the asymptotic bias shrinks to zero, we can show that when \( p = 1 \), the asymptotic bias for the TSLS \( (k = 1) \) is given by

\[
E[B] / \sqrt{T - 1} = q(T - 1)^{-1} \Sigma_{vv} [C^T \Sigma_{zz} C]^{-1} + O(T^{-2})
\]

which has the order \( O(T^{-1}) \). Note that the proof of (8) is similar to that for Theorem
2.2 in Li (2006) and omitted here. Therefore, when $T \to \infty$, we can achieve the consistent estimation. The consistent result and asymptotic normality of the TSLS estimator are summarized in the following corollary.

**Corollary 1** Suppose that Assumptions 1-4 hold, then as both $N$ and $T$ tend to infinity, (a) $\hat{\beta}_{TSLS} \to^{p} \beta_0$; and (b) $\sqrt{T} (\hat{\beta}_{TSLS} - \beta_0) \to^{d} N[0, \sigma_u^2 (C' \Sigma_{ZZ} C)^{-1}]$.

Note that the asymptotic distribution depends on $C$ which is never identified under Assumption 1. Therefore, we cannot make use of the above asymptotic normality to test the coefficient $\beta$. To make inference under weak instruments, we refer to Anderson and Rubin (1949), Kleigerben (2002), Moreira (2003), and among others.

When both $N$ and $T$ tend to infinity, the LIML estimator is asymptotically equivalent to the TSLS estimator. It follows from the following theorem.

**Theorem 2** Suppose that Assumptions 1-4 hold, then as both $N$ and $T$ tend to infinity, $NT(k_{LIML} - 1) \to^{p} 0$.

**Proof.** See the Appendix.

4. Monte Carlo Simulations

In this section, we consider the following model for Monte Carlo simulations:

\[ y_t = \alpha_i + 9Y_t + u_t \]  
(6)

\[ Y_t = (0.7 / \sqrt{N}) Z_t + v_t \]  
(7)

where $Z_t$ is generated from a uniform distribution $U(2, 10)$, $\alpha_i$ is generated from a standard normal distribution. $u_t \overset{d}{\sim} N(0,1)$ and $v_t \overset{d}{\sim} N(0,1)$ are generated jointly from a bivariate normal distribution with the correlation coefficient $\rho = 0.7$. Clearly, $\{Z_t\}$ is independent of $u_t$ and $v_t$. We consider three cases: (a) $T$ is fixed ($T=50$), and $N$ takes values of 50, 150, 250, 350, and 450 respectively; (b) $N$ is fixed ($N=50$), and $T$ takes values of 50, 150, 250, 350, and 450 respectively; and (c) $N=2T$, and $T$ takes values of 20, 40, 60, 80, and 100 respectively. We compute the average absolute bias of the TSLS and LIML estimators respectively, and the median of absolute bias as well. 1000 replications are performed for each pair of $N$ and $T$. All simulation results are provided in Tables 1-3.

When $T$ is fixed, as Table 1 shows, the increase of $N$ reduce the bias of neither TSLS or LIML estimators. When $N$ is fixed but $T$ grows, Table 2 shows that the average absolute bias is reduced from 0.0714 (when $T$ is 50) to 0.0235 (when $T$ is 450) for TSLS estimator, and from 0.0685 to 0.0237 for LIML estimator. The median of the
absolute bias also decreases significantly when $T$ grows large. Table 3 shows that the bias can be reduced when $N$ and $T$ grow proportionally. All these simulation results are consistent to our theory.

Table 1 Average bias and median bias when $T$ is fixed

<table>
<thead>
<tr>
<th>$T=50$</th>
<th>Average Absolute Bias</th>
<th>Median of Absolute Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSLS</td>
<td>LIML</td>
</tr>
<tr>
<td>$N=50$</td>
<td>0.0714</td>
<td>0.0697</td>
</tr>
<tr>
<td>$N=150$</td>
<td>0.0724</td>
<td>0.0705</td>
</tr>
<tr>
<td>$N=250$</td>
<td>0.0732</td>
<td>0.0691</td>
</tr>
<tr>
<td>$N=350$</td>
<td>0.0715</td>
<td>0.0726</td>
</tr>
<tr>
<td>$N=450$</td>
<td>0.0703</td>
<td>0.0726</td>
</tr>
</tbody>
</table>

Table 2 Average bias and median bias when $N$ is fixed

<table>
<thead>
<tr>
<th>$N=50$</th>
<th>Average Absolute Bias</th>
<th>Median of Absolute Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSLS</td>
<td>LIML</td>
</tr>
<tr>
<td>$T=50$</td>
<td>0.0714</td>
<td>0.0685</td>
</tr>
<tr>
<td>$T=150$</td>
<td>0.0402</td>
<td>0.0401</td>
</tr>
<tr>
<td>$T=250$</td>
<td>0.0310</td>
<td>0.0317</td>
</tr>
<tr>
<td>$T=350$</td>
<td>0.0264</td>
<td>0.0271</td>
</tr>
<tr>
<td>$T=450$</td>
<td>0.0235</td>
<td>0.0237</td>
</tr>
</tbody>
</table>

Table 3 Average bias and median bias when $N=2T$

<table>
<thead>
<tr>
<th>$N=2T$</th>
<th>Average Absolute Bias</th>
<th>Median of Absolute Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSLS</td>
<td>LIML</td>
</tr>
<tr>
<td>$T=20$</td>
<td>0.1133</td>
<td>0.1229</td>
</tr>
<tr>
<td>$T=40$</td>
<td>0.0805</td>
<td>0.0810</td>
</tr>
<tr>
<td>$T=60$</td>
<td>0.0649</td>
<td>0.0654</td>
</tr>
<tr>
<td>$T=80$</td>
<td>0.0536</td>
<td>0.0565</td>
</tr>
<tr>
<td>$T=100$</td>
<td>0.0490</td>
<td>0.0518</td>
</tr>
</tbody>
</table>

5. Conclusions

This paper shows that the asymptotic bias arising from weak instruments shrinks when independently repeated cross-sectional data are available. As the number of independently repeated cross sections $T$ goes to infinity, we can achieve the consistent estimation. In future research we consider the weak instruments problem in a dynamic panel data model.

Appendix

Proof of Theorem 1 The bias of the $k$-class estimator is given by
\[ \hat{\beta}_{k\text{-class}} - \beta_0 = [Y^T (I - kM_{\tilde{z}}) Y ]^{-1} [Y^T (I - kM_{\tilde{z}}) u^*] \rightarrow \\
[Y^T P_{\tilde{z}} Y - (\kappa / N) Y^T M_{\tilde{z}} Y ]^{-1} [Y^T P_{\tilde{z}} u^* - (\kappa / N) Y^T M_{\tilde{z}} u^*] \]

We have the following limits holding jointly
(i) \[ Y^T P_{\tilde{z}} Y = (Y^T Z^* / \sqrt{N}) (Z^T Z^* / N)^{-1} (Z^T Y^* / \sqrt{N}) \]
\[ \rightarrow (T - 1) [\Sigma_{zz} C + \frac{1}{\sqrt{T - 1}} \Psi_{zu}]^T \Sigma_{zz}^{-1} [\Sigma_{zz} C + \frac{1}{\sqrt{T - 1}} \Psi_{zu}]. \]

(ii) \[ Y^T P_{\tilde{z}} u^* = (Y^T Z^* / \sqrt{N}) (Z^T Z^* / N)^{-1} (Z^T u^* / \sqrt{N}) \]
\[ \rightarrow \sqrt{T - 1} [\Sigma_{zz} C + \frac{1}{\sqrt{T - 1}} \Psi_{zu}]^T \Sigma_{zz}^{-1} \Psi_{zu}. \]

Note that \( Y^T P_{\tilde{z}} Y^* / N \rightarrow 0 \) and \( Y^T P_{\tilde{z}} u^* / N \rightarrow 0 \). The result of the theorem follows from (i), (ii), and that fact that \( V^T V^* / N \rightarrow (T - 1) \Sigma_{vv} \) and \( V^T u^* / N \rightarrow \Sigma_{vu} \) Q.E.D.

**Proof of Theorem 2** \( k_{LML} \) is the smallest root of the determinantal equation
\[ |Y^* Y^* - kY^* M_{\tilde{z}} Y^*| \]. Let \( J = \begin{pmatrix} 1 & 0 \\ -\beta & 1 \end{pmatrix} \) and note that \( Y^* J = \begin{pmatrix} u^* & Y^* \end{pmatrix} \). Since \( J \) is a non-singular matrix, the roots of the modified determinantal equation
\[ |J^T Y^* Y^* / NT - kJ^T Y^* M_{\tilde{z}} Y^* / NT| = 0 \] has the same roots of the original determinantal equation.
\[ J^T Y^* Y^* / NT - J^T Y^* M_{\tilde{z}} Y^* / NT \]
\[ = \begin{pmatrix} \frac{u^T u^*}{NT} & \frac{u^T Y^*}{NT} \\ \frac{Y^T u^*}{NT} & \frac{Y^T Y^*}{NT} \end{pmatrix} - k \begin{pmatrix} \frac{u^T M_{\tilde{z}} u^*}{NT} & \frac{u^T M_{\tilde{z}} Y^*}{NT} \\ \frac{Y^T M_{\tilde{z}} u^*}{NT} & \frac{Y^T M_{\tilde{z}} Y^*}{NT} \end{pmatrix} \]
\[ \rightarrow \begin{pmatrix} \sigma_u^2 \Sigma_v^T \\ \Sigma_v \end{pmatrix} - \begin{pmatrix} \sigma_u^2 \Sigma_v^T \\ \Sigma_v \end{pmatrix}. \]

It follows that \( NT (k_{LML} - 1) \rightarrow 0 \) Q.E.D.

**References**

Chao, J. and N. Swanson, 2005, Consistent estimation with a large number of weak instruments, Econometrica, 73, 1673-1692.
Nelson, C. and R. Startz, 1990a, Some further results on the exact small sample properties of the instrumental variable estimator, Econometrica, 58, 967-976.
Nelson, C. and R. Startz, 1990b, the distribution of instrumental variable estimator and its t-ratio when the instrument is a poor one, Journal of Business, 63, 5125-5140