Reducing Asymptotic Bias of Weak Instrumental Estimation Using Independently Repeated Cross-Sectional Information*

Zongwu Cai[a,b] and Ying Fang[b,*] and Jia Su[b]

[a]Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA.

[b]Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics, and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China.

May 9, 2011

In this paper, we consider the instrumental variables estimation (the two-stage least squares estimator and the limited information maximum likelihood estimator) using weak instruments in a repeated measurements or a panel data model. We show that independently repeated cross-sectional data can reduce the asymptotic bias of the instrumental variables estimation when instruments are weakly correlated with endogenous variables. When the number of repeated measurements tends to infinity, we can achieve consistent instrumental variables estimation with weak instruments.

KEY WORDS: Bias reduction; Panel data; Weak instruments.

---

* *Corresponding address*: The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China. *E-mail address*: yifst1@xmu.edu.cn (Y. Fang).

1 Introduction

The consistency of a least squares estimator relies on a crucial assumption of the uncorrelation between regressors and measurement errors. However, in many real applications such as business and economics, this uncorrelated assumption is often violated due to measurement error problems, simultaneous causation, unobserved variables, missing information and other problems. This is well documented as the endogenous problem in the literature. To deal with the problem of endogenous regressors, one method widely used in the literature is to adopt an instrumental variables regression. The validity of using instrumental variables requires two conditions. On one hand, the instrumental variables must be uncorrelated with the measurement errors in the structural equations[1]. On the other hand, the instrumental variables must be highly correlated with endogenous regressors. Recent studies found that in many important applications instrumental variables are weakly correlated with endogenous regressors. When one regresses instrumental variables on endogenous variables, the values of R^2 and the partial F statistics are usually very small when instruments are weak. For example, when estimating returns to education, education is measured by years of schooling and is regarded as an endogenous variable (see, e.g., the review paper by Card (2001)) due to the fact that individual ability can affect both years of schooling and future incomes but unfortunately economists can not observe it. Angrist and Krueger (1991) proposed employing quarter of birth as an instrument of years of schooling to estimate returns to education. However, the R^2 of the first-stage regression is extremely low. This is the so-called weak instruments problem.

The most popular instrumental variables estimation includes the two-stage least squares (TSLS) estimator and the limited information maximum likelihood (LIML) estimator. It is well known in the literature that both two-stage least squares estimator and limited information maximum likelihood estimator provide a poor approximation in a cross-sectional model when instruments are weakly correlated with the endogenous explanatory variables (Nelson and Startz, 1990a, 1990b; Bound, et al., 1995; Staiger and Stock, 1997; Stock, Wright and Yogo, 2002; among others). Indeed, both the two-stage least squares estimation and the limited information maximum estimation lead to inconsistent estimators when weak instrumental variables are used. Chao and Swanson (2005) obtained a consistent instrumental variables estimation when available instruments are weak but the number of instruments goes to infinity with the sample size. However, they found that to achieve consistency, the

[1] The exogeneity of instrumental variables requires a zero correlation between instrumental variables and measurement errors, which is hard to be exactly verified in real applications. Berkowitz, Caner and Fang (2008, 2011) discussed the impact on subsequent estimation and inference when the zero correlation is violated.

two-stage least squares estimator needs more stringent conditions than those for the limited-information maximum likelihood estimator. Based on the best of our knowledge, there are relatively few studies on weak instruments estimation using repeated measurements or panel data.

In this paper, we extend the instrumental variables estimation to a model with panel data or repeated measurements. Let N denote the number of individual units and T denote the number of time periods for each indiviual. When instruments are weakly correlated to the endogenous variables, we consider to reduce the asymptotic bias of an instrumental variables estimation using independently repeated cross-sectional information. We will show that the bias term of an instrumental variables estimation has the order of $O(T^{-1})$ when N goes to infinity. When both N and T go to infinity, the consistent instrumental variables estimation can be achieved. Section 2 introduces the basic statistic model and Section 3 lists the main assumptions and derives the large sample properties of an instrumental variables estimation using weak instruments in a panel data model. In Section 4, we conduct a simple Monte Carlo simulation to illustrate the finite sample performance and Section 5 concludes.

2 The Model and Estimation Method

Without loss of generality[2], we consider the following simple simultaneous equations model with independently repeated cross-sectional data:

$$y_{it} = \alpha_i + \beta^\top Y_{it} + u_{it}, \quad Y_{it} = \gamma_i + \Pi^\top Z_{it} + V_{it}, \qquad 1 \le i \le N,\ 1 \le t \le T, \tag{1}$$

where y_{it} is a scalar dependent variable, Y_{it} is $p \times 1$ vector of endogenous variables, Z_{it} is a $q \times 1$ instrumental variables $(q \ge p)$[3], $B^\top$ denotes the transpose of a matrix or vector B, $\{\alpha_i\}$ and $\{\gamma_i\}$ are independent cross individuals i. In panel data models, $\{\alpha_i\}$ and $\{\gamma_i\}$ are the so called fixed effects which are allowed to be correlated to regressors Y_{it}. We assume that $\{Z_{it}, u_{it}\}$ and $\{Z_{it}, V_{it}\}$ are independent across both N and T. N and T are defined earlier. In panel data applications, T is usually the number of observations across time for a given individual unit.

To model the weak correlation between instrumental variables Z_{it} and endogenous regressors Y_{it}, the value for a significance test of $\Pi = 0$ at the first-stage regression should

[2]It is well known that the weak instruments problem does not affect the consistent estimation of the coefficients of included exogenous variables. To ease notation, we focus on a 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.

[3]For the sake of identification, the number of instrumental variables q must be at least equal to the number of endogenous regressors p. Otherwise, β is not identified.

be small even if N tends to infinity. Therefore, to overcome this problem, we follow the so called local-to-zero asymptotics proposed by Staiger and Stock (1997) and widely accepted in the literature of weak instruments, and assume that

$$\Pi = C/\sqrt{N}, \tag{2}$$

where C is a $q \times p$ matrix of parameters contained in a compact set. It is called local-to-zero since the coefficient Π converges to zero as the sample size goes to infinity.

The above model has a lot of applications in empirical studies. For example, Andreoni and Payne (2003) examined whether or not government grants crowd out private charities by employing panel data from arts and social science organizations. They applied the instrumental variables estimation 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 instrumental variables estimation in panel data include Fishback, et al. (2002), Gruber and Hungerman (2007), and Andreoni and Payne (2007) among others.

To remove the individual effect $\{\alpha_i\}$ and $\{\gamma_i\}$, both equations in (1) are multiplied by the forward orthogonal deviations operator A (Arellano, 2003), where $A^\top A = I_T - ee^\top/T$, $AA^\top = I_{T-1}$, where I_T is an identity matrix with dimension $T \times T$ and e is a vector of ones. The transformed model can be represented as

$$y_i^* = Y_i^* \beta + u_i^*, \qquad Y_i^* = Z_i^* \Pi + V_i^*, \quad 1 \leq i \leq N, \tag{3}$$

where $y_i = (y_{i1}, y_{i2}, \cdots, y_{iT})^\top$, $y_i^* = Ay_i$, and $Z_i^* = A\,Z_i$, respectively. The notations of u_i and u_i^* are defined in the same fashion. Thus, $\mathrm{Var}(u_i^*) = \sigma_u^2\, I_{T-1}$ if $\mathrm{Var}(u_i) = \sigma_u^2 I_T$ in the original model.

We consider two most important instrumental variables estimators, the two-stage least squares estimator and the limited information maximum likelihood estimator. Both estimators can be written as a member of the k-class of estimators (Nagar, 1959; Theil, 1961). The k-class of estimators is defined by

$$\widehat{\beta}_{k-class} = [Y^{*\top}(I - kM_{Z^*})Y^*]^{-1}[Y^{*\top}(I - kM_{Z^*})y^*], \tag{4}$$

where $M_B = I - B(B^\top B)^{-1}B^\top$ for a matrix or vector B and I is an identity matrix. When $k = 1$, the k-class of estimators is reduced to a two-stage least squares estimator. When k is the smallest root of the determinantal equation

$$\mid \overline{Y}^{*\top}\overline{Y}^* - k\overline{Y}^{*\top} M_{Z^*}\overline{Y}^* \mid = 0, \tag{5}$$

where $\overline{Y}^* = \begin{pmatrix} y^* & Y^* \end{pmatrix}$, then the k-class of estimators becomes the limited information maximum likelihood estimator. Note that $y^* = (y_1^{*\top}, \cdots, y_N^{*\top})^\top$, $Y^* = (Y_1^{*\top}, \cdots, Y_N^{*\top})^\top$ and $Z^* = (Z_1^{*\top}, \cdots, Z_N^{*\top})^\top$.

3 Large Sample Theory

In this section, we derive the large sample properties of the two-stage least squares estimator and the limited information maximum likelihood estimator using weak instruments in a panel data model with fixed effects. We show that the asymptotic bias of both estimators can be reduced when the number of independently repeated cross sections T increases. As T goes to infinity, we can achieve the 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: $(u^\top u/NT, V^\top u/NT, V^\top V/NT) \xrightarrow{p} (\sigma_u^2, \Sigma_{Vu}, \Sigma_{VV})$.

Assumption 3: $Z^{*\prime} Z^*/NT \xrightarrow{p} \Sigma_{ZZ} = E(Z_{it}Z_{it}^\top) - E(Z_{it})E(Z_{it})^\top$ is positive definite $q \times q$ matrix.

Assumption 4: $(Z^{*\top}u^*/\sqrt{NT}, Z^{*\top}V^*/\sqrt{NT}) \xrightarrow{d} (\Psi_{Zu}, \Psi_{ZV})$, where $(\Psi_{Zu}^\top, \mathrm{Vec}(\Psi_{ZV}^\top)^\top)$ is distributed as $N(0, \Sigma \otimes \Sigma_{ZZ})$, $\Sigma = \begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}^\top \\ \Sigma_{Vu} & \Sigma_{VV} \end{pmatrix}$, $\mathrm{Vec}(\Psi)$ is a vector formed by stacking the columns of Ψ under each, and $\otimes$ denotes the Kronecker product.

Convergence assumptions 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 Assumptions 1-4 hold for the model defined in (3) and* $NT(k - 1) \xrightarrow{p} \kappa_T$*, then as* N *tends to infinity,*

$$\widehat{\beta}_{k-class} - \beta \xrightarrow{d} B/\sqrt{T},$$

where β *is the true parameter in (3),* $B = \left[D^\top \Sigma_{ZZ}^{-1} D - \kappa_T \Sigma_{ZZ}\right]^{-1} \left[D^\top \Sigma_{ZZ}^{-1} \Psi_{Zu} - \kappa_T \Sigma_{Vu}^\top\right]$ *and* $D = \Sigma_{ZZ} C + \Psi_{ZV}/\sqrt{T}$.

The proof of the above theorem is relegated to the appendix. It is obvious that B is a mixture of finite random variables. For the two-stage least squares estimator ($k = 1$), the above result is simplified to $B = \left[D^\top \Sigma_{ZZ}^{-1} D\right] D^\top \Sigma_{ZZ}^{-1} \Psi_{Zu}$.

When only cross-sectional data are considered ($T = 1$), the asymptotic bias of k-class of estimators is given by

$$\widetilde{B} = \left[\widetilde{D}^\top \Sigma_{ZZ}^{-1} \widetilde{D} - \kappa_T \Sigma_{ZZ}\right]^{-1} \left[\widetilde{D}^\top \Sigma_{ZZ}^{-1} \Psi_{Zu} - \kappa_T \Sigma_{Vu}^\top\right],$$

where $\widetilde{D} = \Sigma_{ZZ} C + \Psi_{ZV}$, which is consistent with the result in Staiger and Stock (1997) when the instrumental variables regression with weak instruments is employed in a cross sectional model. It is interesting to note that the bias $\widetilde{B}$ in the cross sectional model is not in the order of $\sqrt{T}$. Theorem 1 shows that the asymptotic bias shrinks as T becomes large. To understand this result, let us recall the so called concentration parameter, a measurement of strength of instruments in the literature, which is defined as

$$\Sigma_{VV}^{-1/2} \Pi^\top Z^{*\top} Z^* \Pi \Sigma_{VV}^{-1/2} \xrightarrow{p} T \Sigma_{VV}^{-1/2} C^\top \Sigma_{ZZ} C \Sigma_{VV}^{-1/2}, \tag{6}$$

which clearly grows as T increases. When both N and T tend to infinity, the concentration parameter also increases to infinity. While in a cross sectional model, the concentrate parameter converges to a constant when N tends to infinity. This is the reason of inconsistency for the instrumental variables estimation. On the other hand, to see how fast the asymptotic bias shrinks to zero, one can show that when $p = 1$, the asymptotic bias for the two-stage least squares estimator ($k = 1$) is given by

$$E[\widehat{\beta}_{TSLS} - \beta] = q T^{-1} \Sigma_{Vu} \left[C^\top \Sigma_{ZZ} C\right]^{-1} + O(T^{-2}), \tag{7}$$

which has the order $O(T^{-1})$. Note that the proof of (7) is similar to that in Cai, Fang and Li (2010) and omitted here. Therefore, when $T \to \infty$, intuitively, we can achieve the consistent estimation of the two-stage least squares estimator and the limited information maximum likelihood estimator. The consistent result and asymptotic normality of the two-stage least squares 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) $\widehat{\beta}_{TSLS} \xrightarrow{p} \beta$; *and* *(b)* $\sqrt{T}\left[\widehat{\beta}_{TSLS} - \beta\right] \xrightarrow{d} N(0, \sigma_u^2 (C^\top \Sigma_{ZZ} C)^{-1})$.

Note that the asymptotic distribution depends on C which is never identified under Assumption 1. Therefore, the above asymptotic normality can not be used in testing the coefficient β. To make a statistical inference under weak instruments in a panel data model, the reader is referred to the paper by Cai, Fang and Li (2010).

When both N and T tend to infinity, the limited information maximum likelihood estimator is asymptotically equivalent to the two-stage least squares estimator. To show the asymptotic equivalence of both estimators, it suffices to prove that $NT(k_{LIML} - 1) \xrightarrow{p} 0$ (Schmidt, 1976), where k_{LIML} is the smallest root of the determinantal equation given in (5).

Theorem 2. Under Assumptions 1-4, as both N and T tend to infinity, one has

$$NT(k_{LIML} - 1) \xrightarrow{p} 0.$$

Therefore, the limited information maximum likelihood estimator and the two-stage least squares estimator are asymptotically equivalent.

4 A Monte Carlo Simulation Study

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

$$y_{it} = \alpha_i + 9\,Y_{it} + u_{it}, \qquad Y_{it} = \lambda_i + (0.7/\sqrt{N})Z_{it} + v_{it}, \quad 1 \le i \le N,\ 1 \le t \le T,$$

where Z_{it} is generated from a uniform distribution $U(2, 10)$, α_i and γ_i are generated from independent standard normal distributions. u_{it} and v_{it} are generated jointly from a bivariate normal distribution with the correlation coefficient $\rho = 0.7$[4]. Note that in the above data generating process, $\Pi = 0.7/\sqrt{N}$ which reflects the setup of weak instruments. Clearly, $\{Z_{it}\}$ is independent of u_{it} and v_{it}. 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 two-stage least squares estimators and the limited information maximum likelihood estimators respectively, and the medians of absolute bias as well. 1000 replications are performed for each pair of N and T. All simulation results are summarized in Tables 1-3.

When T is fixed, as Table 1 shows, an increase of N can not reduce the bias neither of the two-stage least squares estimators or the limited information maximum likelihood estimators. When N is fixed but T grows, Table 2 shows clearly that the average absolute bias is reduced from 0.0714 (when T is 50) to 0.0235 (when T is 450) for the two-stage least squares estimators, and from 0.0685 to 0.0237 for the limited information maximum likelihood estimators. The median of the absolute bias also decreases significantly when T grows large. Table 3 demonstrates that the bias can be reduced when N and T grow proportionally. All these simulation results are consistent with our theoretical results in previous sections.

[4] ρ is used to control the degree of endogeneity of Y_{it}.

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

$T = 50$	Average Absolute Bias		Median of Absolute Bias	
	TSLS	LIML	TSLS	LIML
$N = 50$	0.0714	0.0697	0.0588	0.0596
$N = 150$	0.0724	0.0705	0.0600	0.0588
$N = 250$	0.0732	0.0691	0.0606	0.0609
$N = 350$	0.0715	0.0726	0.0583	0.0618
$N = 450$	0.0703	0.0726	0.0583	0.0611

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

$T = 50$	Average Absolute Bias		Median of Absolute Bias	
	TSLS	LIML	TSLS	LIML
$T = 50$	0.0714	0.0685	0.0582	0.0584
$T = 150$	0.0402	0.0401	0.0344	0.0339
$T = 250$	0.0310	0.0317	0.0274	0.0274
$T = 350$	0.0264	0.0271	0.0213	0.0228
$T = 450$	0.0235	0.0237	0.0198	0.0202

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

$T = 50$	Average Absolute Bias		Median of Absolute Bias	
	TSLS	LIML	TSLS	LIML
$T = 20$	0.1133	0.1229	0.0933	0.1005
$T = 40$	0.0805	0.0810	0.0675	0.0644
$T = 60$	0.0649	0.0654	0.0549	0.0567
$T = 80$	0.0536	0.0565	0.0463	0.0489
$T = 100$	0.0490	0.0518	0.0407	0.0447

5 Conclusions

This paper shows that the asymptotic bias of an instrumental variables estimation arising from weak instruments shrinks when independently repeated cross-sectional data are available. As the number of independently repeated cross sections goes to infinity, we can achieve the consistent estimation of the two-stage least squares estimator and the limited information maximum likelihood estimator.

Acknowledgements

The authors would like to thank the Associate Editor and an anonymous referee for their valuable comments on the early version of this manuscript. Cai's research was supported, in part, by the National Nature Science Foundation of China grants 70871003 and 70971113, as well as funds provided by the "1000 Talents Plan" from Chinese Government, Fujian Province and Xiamen City, the "Cheung Kong Scholarship Project" from Chinese Ministry of Education, the "Minjiang Scholarship Project" from Fujian Province, China, and Xiamen University. Fang's research is partially supported by the National Natural Science Foundation of China 70971113 and by the Fundamental Research Funds for the Central Universities 2010221092.

Appendix

Proof of Theorem 1: The bias of the k-class of estimators is given by

$$\begin{aligned}&\widehat{\beta}_{k-class} - \beta_0 = [Y^{*\top}(I - kM_{Z^*})Y^*]^{-1}[Y^{*\top}(I - kM_{Z^*})u^*] \\ =\ & [Y^{*\top}P_{Z^*}Y^* - (\kappa_T/NT)Y^{*\top}M_{Z^*}Y^*]^{-1}[Y^{*\top}P_{Z^*}u^* - (\kappa_T/NT)Y^{*\top}M_{Z^*}u^*] + o_p(1),\end{aligned}$$

where $P_B = I - M_B$ and κ_T is given in Theorem 1. Clearly, we have the following results: as $N \to \infty$,

$$\begin{aligned}Y^{*\top}P_{Z^*}Y^* &= (Y^{*\top}Z^*/\sqrt{N})(Z^{*\top}Z^*/N)^{-1}(Z^{*\top}Y^*/\sqrt{N}) \\ &\to T[\Sigma_{ZZ}C + \frac{1}{\sqrt{T}}\Psi_{Zu}]^\top\Sigma_{ZZ}^{-1}[\Sigma_{ZZ}C + \frac{1}{\sqrt{T}}\Psi_{Zu}] \qquad \text{(A.1)}\end{aligned}$$

and

$$\begin{aligned}Y^{*\top}P_{Z^*}u^* &= (Y^{*\top}Z^*/\sqrt{N})(Z^{*\top}Z^*/N)^{-1}(Z^{*\top}u^*/\sqrt{N}) \\ &\to \sqrt{T}[\Sigma_{ZZ}C + \frac{1}{\sqrt{T}}\Psi_{ZV}]^\top\Sigma_{ZZ}^{-1}\Psi_{Zu}. \qquad \text{(A.2)}\end{aligned}$$

Note that $Y^{*\top}P_{Z^*}Y^*/NT \to 0$ and $Y^{*\top}P_{Z^*}u^*/NT \to 0$ as $N \to \infty$. The result of the theorem follows from (A.1), (A.2), and the facts that $V^{*\top}V^*/NT \to^p \Sigma_{VV}$ and $V^{*\top}u^*/NT \to^p \Sigma_{Vu}$. Q.E.D.

Proof of Theorem 2: k_{LIML} is the smallest root of the determinantal equation $|\bar{Y}^{*\top}\bar{Y}^* - k\bar{Y}^{*\top}M_{Z^*}\bar{Y}^*| = 0$. Let $J = \begin{pmatrix} 1 & 0 \\ -\beta & I \end{pmatrix}$ and note that $\bar{Y}^*J = \begin{pmatrix} u^* & Y^* \end{pmatrix}$. Since J is a nonsingular matrix, the roots of the modified determinantal equation $|NT(J^\top\bar{Y}^{*\top}\bar{Y}^*J/NT - kJ^\top\bar{Y}^{*\top}M_{Z^*}\bar{Y}^*J/NT)| = 0$ has the same roots of the original determinantal equation

$$\begin{aligned}& NT(J^\top\bar{Y}^{*\top}\bar{Y}^*J/NT - kJ^\top\bar{Y}^{*\top}M_{Z^*}\bar{Y}^*J/NT) \\ =\ & NT \times \{\begin{pmatrix} \frac{u^{*\top}u^*}{NT} & \frac{u^{*\top}Y^*}{NT} \\ \frac{Y^{*\top}u^*}{NT} & \frac{Y^{*\top}Y^*}{NT}\end{pmatrix} - \begin{pmatrix} \frac{u^{*\top}M_{Z^*}u^*}{NT} & \frac{u^{*\top}M_{Z^*}Y^*}{NT} \\ \frac{Y^{*\top}M_{Z^*}u^*}{NT} & \frac{Y^{*\top}M_{Z^*}Y^*}{NT}\end{pmatrix}\} \\ \to^p\ & NT\{\begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}^\top \\ \Sigma_{Vu} & \Sigma_{VV}\end{pmatrix} - k\begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}^\top \\ \Sigma_{Vu} & \Sigma_{VV}\end{pmatrix}\} \\ =\ & NT(k-1)\begin{pmatrix} \sigma_u^2 & \Sigma_{Vu}^\top \\ \Sigma_{Vu} & \Sigma_{VV}\end{pmatrix} = 0.\end{aligned}$$

It follows that $NT(k_{LIML} - 1) \to 0$. Q.E.D.

References

Andreoni, J., Payne, A., 2003, Do government grants to private charities crowd out giving or fund-raising?, American Economic Review, 792-812.

Andreoni, J., Payne, A., 2007, Crowding out both sides of the philanthropy market: evidence from a panel of charities, Working Paper, University of California, San Diego.

Arellano, M., 2003, Panel Data Econometrics, New York: Oxford University Press.

Berkowitz, D., Caner, M., Fang, Y., 2008, Are nearly exogenous instruments reliable? Economics Letters 101, 20-28.

Berkowitz, D., Caner, M., Fang, Y., 2011, The validity of instruments revisited, working paper.

Bound, J., Jaeger D.A., Baker, R.M., 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, Journal of the American Statistical Association 90, 443-450.

Cai, Z., Fang, Y., Li, H., 2010, Weak instrumental variables models for longitudinal data, Econometric Reviews, forthcoming.

Card, D., 2001, Estimating the return to schooling: Progress on some persistent econometric problems, Econometrica 69, 1127-1160.

Chao, J., Swanson, N., 2005, Consistent estimation with a large number of weak instruments, Econometrica 73, 1673-1692.

Fishback, P., Haines, M., Kantor, S., 2002, The welfare of children during the great depression, NBER working paper, No. 8902.

Gruber, J., Hungerman, D., 2007, Faith-based charity and crowd-out during the great depression, Journal of Public Economics 91, 1043-1069.

Nagar, A.L., 1959, The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations, Econometrica 27, 575-595.

Nelson, C., Startz, R., 1990a, Some further results on the exact small sample properties of the instrumental variable estimator, Econometrica 58, 967-976.

Nelson, C., Startz, R., 1990b, The distribution of instrumental variable estimator and its t-ratio when the instrument is a poor one, Journal of Business 63, 125-5140.

Schmidt, P., 1976, Econometrics, New York: Marcel Dekker.

Staiger, D., Stock, J., 1997, Instrumental variables regression with weak instruments, Econometrica 65, 556-586.

Stock, J., Wright, J., 2000, GMM with weak instruments, Econometrica 68, 1055-1096.

Stock, J., Wright, J., Yogo, M., 2002, A survey of weak instruments and weak identification in generalized method of moments, Journal of Business and Economic Statistics 20,518-529.

Theil, H., 1961, Economic Forecasts and Policy, 2nd edition, Amsterdam: North-Holland.