The asymptotic distribution of the instrumental variable estimators when the instruments are not correlated with the regressors

Chirok Han, Peter Schmidt*

Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA

Received 19 March 2001; accepted 13 June 2001

Abstract

When the instruments are irrelevant, the IV estimator is neither consistent nor asymptotically normal. This paper calculates the mean of its asymptotic distribution and shows that it equals the probability limit of the OLS estimator. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Instrumental variables; Irrelevant instruments; Weak instruments

JEL classification: C1; C3

1. Introduction

A number of recent papers, including Bound et al. (1995) and Staiger and Stock (1997), have considered instrumental variable (IV) estimators when the instruments are weak, in the sense that the correlation between the instruments and the regressors is low. In this paper, we consider the extreme case that the instruments are completely irrelevant. In this case we can prove the following interesting result: the mean of the asymptotic distribution of the IV estimator is the same as the probability limit of the OLS estimator. Thus, as might be expected, irrelevant instruments do not remove the least-squares bias.

To be specific, consider the linear model $y = X\beta + \varepsilon$ (in matrix notation) where ε is a $T \times 1$ random vector with mean zero, X is a $T \times K$ random matrix of regressors, and β is a $K \times 1$ parameter. It is well known that when X_t is correlated with ε_t, the ordinary least-squares (OLS) estimator is not consistent. More specifically, under the regularity conditions that ensure the convergence of the

---

*Corresponding author. Tel.: +1-517-355-8381; fax: +1-517-432-1068.

E-mail address: schmidtp@msu.edu (P. Schmidt).

statistics $T^{-1}X'X$ and $T^{-1}X'\varepsilon$ in probability, the OLS estimator converges in probability as $T \to \infty$ to $\beta_0 + (EX_t X_t')^{-1}EX_t\varepsilon_t$, which is different from β_0, the true parameter, unless $EX_t\varepsilon_t = 0$.

To obtain a consistent estimator, one possibility is instrumental variable estimation. Good instruments Z $(T \times L)$ are those which satisfy:

(i) $T^{-1}Z'Z$ converges in probability to a nonrandom, nonsingular matrix;

(ii) $T^{-1}Z'X$ converges in probability to a nonrandom matrix with full column rank;

(iii) $T^{-1/2}Z'\varepsilon$ converges in distribution to a normal random vector with zero mean.

When the instruments are good, the IV estimator is consistent and asymptotically normal.

Here we are concerned with the case that condition (ii) fails. Suppose that $L \geq K$, so that there are enough instruments, but the instruments (Z) are not strongly correlated with the regressors (X). Specifically, let the reduced form for X be:

$$X = Z\Pi + V \tag{1}$$

Staiger and Stock (1997) consider the case that $\Pi = \Pi_T = C/\sqrt{T}$, with C a $L \times K$ matrix of constants. They call this the case of *weak instruments*. In this case the correlation between X_t and Z_t is of order $T^{-1/2}$, and condition (ii) fails. Staiger and Stock show that with weak instruments $\hat{\beta}_{IV}$, the IV estimator, does not have a probability limit but rather $\hat{\beta}_{IV} - \beta_0$ converges to a non-normal random variable. The mean of the asymptotic distribution of $\hat{\beta}_{IV} - \beta_0$ is non-zero, so that with weak instruments there is asymptotic bias. This bias is in the same direction as the bias of OLS.

In this paper we consider the case of *irrelevant instruments*, which are uncorrelated with the regressors. This is a special case of Staiger and Stock, corresponding to $C = 0$ so that $\Pi = 0$ in the reduced form (1) for all T. In this case we show that the mean of the asymptotic distribution of $(\hat{\beta}_{IV} - \beta_0)$ is the same as $(\text{plim}\hat{\beta}_{OLS} - \beta_0)$, the asymptotic bias of the OLS estimator.

2. The limit distribution

Consider a linear model in matrix notation

$$y = X^\circ \beta + W\gamma + \varepsilon \tag{2}$$

where y and X° are, respectively, a $T \times 1$ vector of dependent variables and a $T \times K$ matrix of the endogenous regressors, W is a $T \times G$ matrix of exogenous regressors, the first column of which is a vector of ones, ε is the vector of errors, and β and γ are the parameters to be estimated.

Consider a $T \times L$ random matrix Z° of 'instruments'. For any matrix A with full column rank, let $P_A = A(A'A)^{-1}A'$. Let $X = (I - P_W)X^\circ$ and $Z = (I - P_W)Z^\circ$. Thus X is the part of the endogenous regressors not explained by the exogenous regressors, and similarly Z is the part of the 'instruments' not explained by the exogenous regressors.

We make the following 'high level' assumptions.

Assumption 1. $T^{-1}X'X$, $T^{-1}\varepsilon'\varepsilon$, and $T^{-1}Z'Z$ converge in probability to finite, nonrandom, nonsingular matrices, and $T^{-1}X\varepsilon$ converges to a nonrandom matrix.

Let $\Sigma = \text{plim}T^{-1}(X,\varepsilon)'(X,\varepsilon)$. It has submatrices Σ_{XX}, $\Sigma_{X\varepsilon}$, and $\sigma_{\varepsilon\varepsilon}$, which are the probability limits

of $T^{-1}X'X$, $T^{-1}X'\varepsilon$, and $T^{-1}\varepsilon'\varepsilon$, respectively. Also let $\Omega = \mathrm{plim}\, T^{-1}Z'Z$. Assumption 1 can be regarded as the implication of a law of large numbers under more primitive assumptions on the sequences. For example, when the sequence $(\varepsilon_t, X_t^{\circ\prime}, Z_t^{\circ\prime})'$ is i.i.d. and its second moment exists, $\Sigma_{XX} = EX_t^\circ X_t^{\circ\prime} - EX_t^\circ W_t'(EW_t W_t')^{-1}EW_t X_t^{\circ\prime}$, $\Sigma_{X\varepsilon} = EX_t^\circ \varepsilon_t - EX_t^\circ W_t'(EW_t W_t')^{-1}EW_t \varepsilon_t = EX_t^\circ \varepsilon_t$, $\sigma_{\varepsilon\epsilon} = E\varepsilon_t^2$, and $\Omega = EZ_t^\circ Z_t^{\circ\prime} - EZ_t^\circ W_t'(EW_t W_t')^{-1}EW_t Z_t^{\circ\prime}$.

Let $\rho = \Sigma_{XX}^{-1/2} \Sigma_{X\varepsilon} \sigma_{\varepsilon\epsilon}^{-1/2}$ which is a multivariate correlation coefficient. A key assumption is the irrelevance of Z as instruments for X, as follows:

Assumption 2. $T^{-1/2}Z'(X, \ \varepsilon) \overset{d}{\to} \Omega^{1/2}(\xi \Sigma_{XX}^{1/2}, \ \eta \sigma_{\varepsilon\epsilon}^{1/2})$ where $\mathrm{vec}(\xi, \ \eta)$ is a multivariate centered normal with $E\,\mathrm{vec}(\xi)\,\mathrm{vec}(\xi)' = I$, $E\eta\eta' = I$ and $E\,\mathrm{vec}(\xi)\eta' = \rho \otimes I$.

Note that Assumption 2 implies $T^{-1}Z'X \overset{p}{\to} 0$, which may agree with an intuitive definition of irrelevant instruments. Also, this assumption can be regarded as the implication of a central limit theorem under more primitive assumptions, as above.

Now let $\hat{\beta}_{\mathrm{IV}}$ be the estimate of β in Eq. (2), when estimation is by IV using (Z°, W) as instruments. It is readily shown that

$$\hat{\beta}_{\mathrm{IV}} - \beta_0 = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon \tag{3}$$

By dividing $Z'Z$ by T, and $Z'X$ and $Z'\varepsilon$ by $T^{1/2}$, we observe that $\hat{\beta}_{\mathrm{IV}} - \beta_0$ is a function φ of $(T^{-1}Z'Z, T^{-1/2}Z'X, T^{-1/2}Z'\varepsilon)$, where $\varphi: \mathbb{R}^{L \times L} \times \mathbb{R}^{L \times K} \times \mathbb{R}^{L \times 1} \to \mathbb{R}^{K \times 1}$ is defined by $\varphi(\Omega, A, b) = (A'\Omega^{-1}A)^{-1}A'\Omega^{-1}b$. Obviously, φ is measurable and is almost surely continuous in the limit. Here continuity is assured by the nonsingularity of the limit of $T^{-1}Z'Z$ and the almost sure full column rank of the limit of $T^{-1/2}Z'X$. Therefore, we apply the continuous mapping theorem to get the following result.

Theorem 1. *Under Assumptions* 1 *and* 2,

$$\hat{\beta}_{\mathrm{IV}} \overset{d}{\to} \tilde{\beta}_{\mathrm{asy}} = \beta_0 + \Sigma_{XX}^{-1/2}(\xi'\xi)^{-1}\xi'\eta\sigma_{\varepsilon\epsilon}^{1/2} \tag{4A}$$

or equivalently,

$$\hat{\delta} = \Sigma_{XX}^{1/2}(\hat{\beta}_{\mathrm{IV}} - \beta_0)\sigma_{\varepsilon\epsilon}^{-1/2} \overset{d}{\to} \tilde{\delta}_{\mathrm{asy}} = (\xi'\xi)^{-1}\xi'\eta \tag{4B}$$

We note that the result in (4A) is the same as Eq. (2.5) of Staiger and Stock (1997, p. 562) when $C = 0$ (and therefore $\lambda = 0$ in (2.3a) and (2.3b)).

We now calculate the density of $\tilde{\delta}_{\mathrm{asy}}$ as follows.

Theorem 2. *Under Assumptions* 1 *and* 2, *the density of* $\tilde{\delta}_{\mathrm{asy}}$ *is*

$$f(d) = C_{K,L} \cdot (1 - \rho'\rho)^{-K/2} \left| (I\ d) \begin{pmatrix} I & \rho \\ \rho' & 1 \end{pmatrix}^{-1} \begin{pmatrix} I \\ d' \end{pmatrix} \right|^{-(L+1)/2} \tag{5}$$

where

$$C_{K,L} = 2^{-(L-1)(K-1)/2}\, \pi^{-K/2}\, \Gamma\!\left(\frac{L+1}{2}\right) \Gamma\!\left(\frac{L-K+1}{2}\right)^{-1}.$$

Proof. See Appendix A.

Given K (the dimension of X_t) and L (the dimension of Z_t), the density depends upon ρ only. As is mentioned in Phillips (1980), this density is similar to the multivariate t distribution. The first moment of $\tilde{\delta}_{asy}$ exists as long as L is strictly greater than K, and more generally its integer moments exist up to the degree of over-identification (see Phillips, 1980, p. 870).

3. The relationship with the OLS estimator

We are now in a position to prove our main result.

Theorem 3. *Suppose* $L > K$. *Then under Assumptions* 1 *and* 2, *the mean of* $\tilde{\beta}_{asy}$ *is equal to the probability limit of the OLS estimator.*

Proof. We observe that the density of $\tilde{\delta}_{asy}$ in (4B) is symmetric around ρ, the correlation coefficient of the endogenous regressors and the error. Furthermore, if $L > K$, the mean of $\tilde{\delta}_{asy}$ exists. Therefore, if $L > K$, $E\tilde{\delta}_{asy} = \rho$. Then

$$
\begin{aligned}
E\tilde{\beta}_{asy} &= \beta_0 + \Sigma_{XX}^{-1/2} E\tilde{\delta}_{asy} \sigma_{\varepsilon\epsilon}^{1/2} \\
&= \beta_0 + \Sigma_{XX}^{-1/2} \rho \sigma_{\varepsilon\epsilon}^{1/2} \\
&= \beta_0 + \Sigma_{XX}^{-1} \Sigma_{X\varepsilon} \\
&= \text{plim } \hat{\beta}_{OLS}.
\end{aligned}
\tag{6}
$$

An alternative proof that does not depend on the exact form of the density of $\tilde{\delta}_{asy}$ is as follows. When the mean of $\tilde{\delta}_{asy} = (\xi'\xi)^{-1}\xi'\eta$ exists,

$$
E(\xi'\xi)^{-1}\xi'\eta = E(\xi'\xi)^{-1}\xi'E(\eta|\xi)
\tag{7}
$$

by the law of iterated expectations. But since $E \text{ vec}(\xi) \text{ vec}(\xi)'$, $E \text{ vec}(\xi)\eta'$, and $E\eta\eta'$ are, respectively, equal to $I \otimes I$, $\rho \otimes I$, and I,

$$
E\eta|\text{vec}(\xi) = (\rho' \otimes I)(I \otimes I)^{-1}\text{vec}(\xi) = \xi\rho.
\tag{8}
$$

(For the operations involved with the Kronecker product and vec operators, see Magnus and Neudecker (1988, Ch. 2).) Hence, $E(\xi'\xi)^{-1}\xi'\eta = \rho$. It follows that $E\tilde{\beta}_{asy} = \beta_0 + \Sigma_{XX}^{-1/2} E\tilde{\delta}_{asy} \sigma_{\varepsilon\varepsilon}^{1/2} = \beta_0 + \Sigma_{XX}^{-1}\Sigma_{X\varepsilon}$, which is equal to the probability limit of the OLS estimator, as in the original proof.

4. Conclusion

In this paper, we answered some questions about the IV estimator using irrelevant instruments in linear models. We saw that the IV estimator is not consistent but converges to a nondegenerate distribution which is similar to a multivariate t distribution. When the number of instruments

(excluding the exogenous regressors) is strictly greater than the number of endogenous regressors, the mean of the asymptotic distribution exists and is equal to the probability limit of the OLS estimator.

Appendix A. Proof of Theorem 2

First, observe that the rows of the $L \times (K+1)$ matrix (ξ, η) are a random sample from $N(0, J)$ where

$$J = \begin{pmatrix} I & \rho \\ \rho' & 1 \end{pmatrix}.$$

Thus, $(\xi, \eta)'(\xi, \eta)$ has a $K+1$ dimensional central Wishart distribution with L degrees of freedom on the covariance matrix J. When $L \geq K+1$, its density at the point $\xi'\xi = B_1$, $\xi'\eta = b_2$, and $\eta'\eta = b_3$ is

$$g(B_1, b_2, b_3) = 2^{-L(K+1)/2} \Gamma_{K+1} \left(\frac{L}{2} \right)^{-1} (1 - \rho'\rho)^{-L/2} |B|^{(L-K-2)/2} \exp\{-1/2 \operatorname{tr} J^{-1} B\} \tag{A.1}$$

where

$$B = \begin{pmatrix} B_1 & b_2 \\ b_2' & b_3 \end{pmatrix} \text{ and } \Gamma_n$$

is the multivariate gamma function defined as

$$\Gamma_n(a) = \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma \left(a - \frac{j-1}{2} \right). \tag{A.2}$$

See Johnson and Kotz (1972, p. 162).

Following Phillips (1980), consider the one-to-one mapping ψ on the set of $K+1$ dimensional, real, symmetric, positive definite matrices defined as

$$\psi : \begin{pmatrix} B_1 & b_2 \\ b_2' & b_3 \end{pmatrix} \to \begin{pmatrix} B_1 & B_1^{-1} b_2 \\ b_2' B_1^{-1} & b_3 - b_2' B_1^{-1} B_3 \end{pmatrix} \tag{A.3}$$

Then the inverse ψ^{-1} is

$$\psi^{-1} : \begin{pmatrix} A_1 & d \\ d' & a_3 \end{pmatrix} \to \begin{pmatrix} A_1 & A_1 d \\ d' A_1 & a_3 + d' A_1 d \end{pmatrix} \tag{A.4}$$

whose Jacobian turns out to be $|A_1|$. Therefore, by the change-of-variable technique, the density of the symmetric random matrix, which is defined such that the upper-left $K \times K$ diagonal block is $\xi'\xi$, the lower-right 1×1 diagonal block is $\eta'\eta - \eta'\xi(\xi'\xi)^{-1}\xi'\eta$, and the upper-right $K \times 1$ off-diagonal block is $\tilde{\delta}_{\text{asy}} = (\xi'\xi)^{-1}\xi'\eta$, evaluated at the point such that

$$\xi'\xi = A_1, (\xi'\xi)^{-1}\xi'\eta = d, \text{ and } \eta'\eta - \eta'\xi(\xi'\xi)^{-1}\xi'\eta = a_3, \tag{A.5}$$

where A_1 is symmetric, positive definite and a_3 is positive, becomes

$$
\begin{aligned}
h(A_1,d,a_3) \quad &= g(A_1,A_1 d,a_3 + d' A_1 d)\cdot|A_1| \\
&= 2^{-L(K+1)/2}\, \Gamma_{K+1}\!\left(\frac{L}{2}\right)^{-1}(1-\rho\rho')^{-L/2}\cdot H_1(A_1)\cdot H_3(a_3)
\end{aligned}
\tag{A.6}
$$

where

$$
H_1(S) = |S|^{(L-K)/2}\exp\{-\tfrac{1}{2}\,\mathrm{tr}\,S[I+(1-\rho'\rho)^{-1}(d-\rho)(d-\rho)']\}
\tag{A.7}
$$

and

$$
H_3(x) = x^{(L-K)/2-1}\exp\{-\tfrac{1}{2}\,x(1-\rho'\rho)^{-1}\}
\tag{A.8}
$$

The density of $\tilde{\delta}_{\mathrm{asy}}$ at d is obtained by integrating out A_1 (symmetric and positive definite) and a_3 (positive) from (A.6). From the definition of the $\Gamma(\cdot)$ function, the integral of $H_3(a_3)$ in (A.6) over all positive a_3 is equal to

$$
\int_0^\infty H_3(x)\mathrm{d}x = 2^{(L-K)/2}(1-\rho'\rho)^{(L-K)/2}\Gamma\!\left(\frac{L-K}{2}\right)
\tag{A.9}
$$

The integral of the matrix argumented function $H_1(S)$ over all symmetric, positive definite matrices is obtained from the results in James (1964). Eqs. (25), (26), and (28) of James (1964, pp. 479–480) imply that for any nonsingular real symmetric $K\times K$ matrix D,

$$
\int_{S>0} |S|^{a-(K+1)/2}\exp\{-\mathrm{tr}\,SD\}\mathrm{d}S = \Gamma_K(a)|D|^{-a}
\tag{A.10}
$$

where the integral is taken over all symmetric, positive definite $K\times K$ matrices. Thus, we have the evaluation

$$
\int_{S>0} H_1(S)\mathrm{d}S = 2^{(L+1)/2}\Gamma_K\!\left(\frac{L+1}{2}\right)\cdot|I+(1-\rho'\rho)^{-1}(d-\rho)(d-\rho')|^{-(L+1)/2}
\tag{A.11}
$$

The desired density (5) is obtained by combining Eqs. (A.6), (A.9), and (A.11).

References

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

James, A.T., 1964. Distributions of matrix variates and latent roots derived from normal samples. Annals of Mathematical Statistics 35, 475–501.

Johnson, N.L., Kotz, S., 1972. In: Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

Magnus, J.R., Neudecker, H., 1988. In: Matrix Differentiation Calculus with Applications in Statistics and Econometrics. Wiley, New York.

Phillips, P.C.B., 1980. The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables. Econometrica 48, 861–878.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.