ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator
Author(s): Charles R. Nelson and Richard Startz
Source: *Econometrica*, Vol. 58, No. 4 (Jul., 1990), pp. 967-976
Published by: The Econometric Society
Stable URL: https://www.jstor.org/stable/2938359
Accessed: 25-07-2019 10:04 UTC

Econometrica, Vol. 58, No. 4 (July, 1990), 967–976

SOME FURTHER RESULTS ON THE EXACT SMALL SAMPLE PROPERTIES OF THE INSTRUMENTAL VARIABLE ESTIMATOR

BY CHARLES R. NELSON AND RICHARD STARTZ[1]

IN THIS PAPER we present new results on the exact small sample distribution of the instrumental variable estimator. In particular, we compare the small sample distribution to the asymptotic distribution.

Among our findings are:

- The central tendency of the instrumental variable estimator is biased away from the true value.
- The central tendency is biased in the direction of the probability limit of the ordinary least squares estimator.
- The distribution is bimodal.
- The probability density equals zero at a point between the modes.
- When the asymptotic variance is large, the distribution is concentrated around the point where the probability equals zero.
- When the right-hand side variable and the regression error are uncorrelated, that is when ordinary least squares is the appropriate estimator, then the asymptotic approximation to the distribution of the instrumental variable gives the exact distribution.
- The asymptotic distribution is a poor approximation to the true distribution where the instruments are poor, in the sense of not being highly correlated with the regressor, and when the number of observations is small.

While we prove each of these claims for the case of one independent variable and one instrument, we see no reason why the intuition should not be of more general applicability.

Consider the regression equation $y = \beta x + u$, to be estimated with instrument z. To facilitate discussion, consider the following model describing generation of the data matrix:

$$y = \beta x + u,$$

$$x = \varepsilon + \lambda^{-1} u,$$

$$z = \upsilon + \gamma \varepsilon,$$

where, without loss of generality, $\beta = 0$. Assume u_i to be i.i.d. normal with zero mean and variance σ_u^2. Assume further that $(1/n)\Sigma[\varepsilon_i \upsilon_i]'[\varepsilon_i \upsilon_i]$ converges to a diagonal matrix. We let σ_ε^2 denote the first diagonal entry and σ_υ^2 the second diagonal entry. Within this sample there is only a single random variable, u, which accounts for the stochastic behavior of both y and x. Note that $\text{plim}(z'x/n) = \gamma\sigma_\varepsilon^2$ and $\text{plim}(z'z/n) = \sigma_\upsilon^2 + \gamma^2\sigma_\varepsilon^2$. So long as $\gamma \neq 0$, the variance of the limiting distribution of $n^{1/2}(\beta_{\text{IV}} - \beta)$ is $V^A = \sigma_u^2(\sigma_\upsilon^2 + \gamma^2\sigma_\varepsilon^2)/\gamma^2\sigma_\varepsilon^4, \beta_{\text{IV}} - \beta \sim AN(0, V^A/n)$.

In this paper we focus on the finite sample distribution of $\beta_{\text{IV}} - \beta$. The distribution we will derive is conditioned on ε and υ, and we impose the further restriction that the sample cross matrix is diagonal. We let $m_{\varepsilon\varepsilon}$ and $m_{\upsilon\upsilon}$ denote the sample second moments for ε and υ respectively and m_{zu} be the sample cross product between z and u. If $\gamma = 0$, z is uncorrelated with x and therefore is not a valid instrument—and the derivation of the asymptotic distribution breaks down. If γ is small, the normal approximation is asymptotically valid, but a poor approximation to the true distribution in finite samples.

[1] Suggestions from Neil Ericsson, seminar participants at the University of Maryland and the Federal Reserve Board, and from the referees and editor are greatly appreciated.

The instrumental variable estimator is

$$\beta_{\mathrm{IV}} = \frac{\Sigma zy}{\Sigma zx} = \beta + \frac{\frac{1}{n}\Sigma zu}{\frac{1}{n}\Sigma zx}.$$

The derivation of the asymptotic distribution is, loosely, that the distribution of $\Sigma zu/n$ approaches a bell curve centered upon zero and that $\Sigma zx/n$ goes to some nonzero constant; since a normal variable divided by a constant is still normal, the IV estimator is approximately bell-shaped around the true parameter.

Now expand out the denominator as

$$\frac{1}{n}\sum zx = \frac{1}{n}\sum z\varepsilon + \lambda^{-1}\frac{1}{n}\sum zu.$$

We can rewrite the IV estimator as

$$\beta_{\mathrm{IV}} = \beta + \frac{m_{zu}}{\gamma m_{\varepsilon\varepsilon} + \lambda^{-1} m_{zu}}.$$

The instrument being "poor" means the covariance between the instrument and regressor is small, which amounts to saying that $\gamma m_{\varepsilon\varepsilon}$ is close to zero. But if $\gamma m_{\varepsilon\varepsilon}$ is negligible, then $\beta_{\mathrm{IV}} - \beta$, rather than being dispersed around zero, will *typically be close to* λ. Therefore, for $\gamma = 0$, where the derivation of the asymptotic distribution breaks down, the finite sample distribution of $\beta_{\mathrm{IV}} - \beta$ collapses around the point λ. For this reason, we call λ a "point of concentration."

The small sample properties of the instrumental variable estimator for this particular problem have been considered by, among others, Basmann (1974), who summarizes a large body of work with particular respect to Haavelmo's model of the marginal propensity to consume, by Mariano and McDonald (1979), who give the pdf for $\hat{\beta}$, and by Anderson (1982), who discusses approximations to the cdf. Basmann and Mariano and McDonald point out that if x and u are normal, then $\hat{\beta}$ is the ratio of two correlated normal random variables and so its distribution may be studied using Fieller's (1932) results. (See also Johnson and Kotz (1972, pp. 123–124), Hinkley (1969), and Marsaglia (1965).) Phillips (1983) *Handbook of Econometrics* article provides a general review of available small sample results on the instrumental variable estimator. Early solutions for pdf of the instrumental variable estimator were provided by Richardson (1968) and Sawa (1969). Phillips (1980) provides the exact pdf for the general case of an equation with $n + 1$ included endogenous variables.

In this paper, we restrict our attention to the case in which there is one right-hand-side endogenous variable and one instrument. In return, we achieve considerable simplification in both the derivation and form of the pdf. Expressions for the pdf given by previous authors require summations of infinite series. As Phillips (1980, p. 870) says of the fully general case, "[While this] gives us a general representation of the exact joint density function of instrumental variable estimators... this type of series representation of the density is not as easy to interpret as we would like." The simplification here permits us to extend the work just cited by characterizing the pdf and cdf of instrumental variables, and comparing them to the asymptotic approximations, as the "quality" of the instruments varies. The most important finding for empirical work is that when the instruments are poor, the distribution of the instrumental variable estimator concentrates at a point away from the true parameter. (See Nelson and Startz (forthcoming).) In addition, we present an intuitive explanation of why the instrumental variable estimator may be a very poor one in small samples.

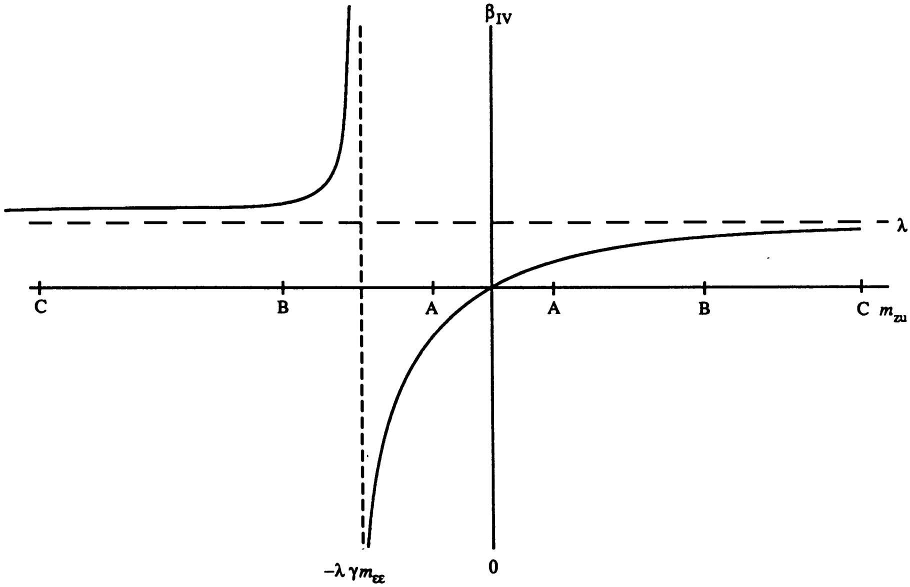

FIGURE 1.—$\hat{\beta}$ as a function of m_{zu}.

The body of the paper consists of two sections. In the first, we derive the exact density function and cumulative distribution function of the instrumental variable estimator. In the second, we consider the behavior of the distribution for various parameter values and prove the propositions stated above.

1. THE EXACT DENSITY AND DISTRIBUTION FUNCTIONS OF INSTRUMENTAL VARIABLES

The instrumental variable estimator can be written as the ratio of two normal random variables, $z'u$ and $z'x$. For convenience, write the sample moment of $z'u$ as

$$m_{zu} = \frac{1}{n} \sum_{i=1}^{n} z_i u_i .$$

Note that

$$m_{zu} \sim N\left(0, \frac{\sigma_u^2 m_{zz}}{n}\right) = N\left(0, (\gamma m_{\varepsilon\varepsilon})^2 V/n\right),$$

where $V = \sigma_u^2 m_{zz}/(\gamma m_{\varepsilon\varepsilon})^2$. Since $z'x = z'\varepsilon + \lambda^{-1} z'u$ we can write the instrumental variable estimator as

$$\hat{\beta} = \frac{m_{zu}}{\gamma m_{\varepsilon\varepsilon} + \lambda^{-1} m_{zu}} .$$

Figure 1 displays a graph of $\hat{\beta}$ as a function of m_{zu}. Note that while $\hat{\beta}$ is neither continuous nor monotonic in m_{zu}, there is nonetheless a one-to-one and onto correspondence and the function is differentiable everywhere except at the single discontinuity.

Since m_{zu} has a normal density, it is straightforward to derive the density of $\hat{\beta}$ by change of variables. If m_{zu} has the density function $f_m(m_{zu})$, then $\hat{\beta}$ has the density function $f_m(m_{zu}) dm/d\hat{\beta}$, where $dm/d\hat{\beta} = \gamma m_{\varepsilon\varepsilon}/(1 - \hat{\beta}/\lambda)^2$. The density function for

m_{zu} is given by

$$f(m_{zu}) = \left(2\pi\gamma^2 m_{\varepsilon\varepsilon}^2 V/n\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\gamma^2 m_{\varepsilon\varepsilon}^2 V/n} m_{zu}^2\right].$$

Therefore, the exact density of $\hat{\beta}$ is given by:

$$(1) \qquad f(\hat{\beta}) = \frac{1}{(1-\hat{\beta}/\lambda)^2} \frac{1}{\sqrt{2\pi V/n}} \exp\left[-\frac{1}{2V/n}\left(\frac{\hat{\beta}}{1-\hat{\beta}/\lambda}\right)^2\right]$$

as compared to the asymptotic approximation given by:

$$(2) \qquad f^A(\hat{\beta}) = \frac{1}{\sqrt{2\pi V^A/n}} \exp\left[-\frac{1}{2V^A/n}\hat{\beta}^2\right].$$

Note that while the asymptotic distribution depends only on the asymptotic variance, V/n, the true distribution depends on two parameters, V/n and λ.

Turn now to the derivation of the cumulative distribution function for $\hat{\beta}$. Since the mapping between $\hat{\beta}$ and m_{zu} is one-to-one and onto, the probability of $\hat{\beta}$ lying in a given interval is just the probability of m_{zu} lying in the corresponding interval of the normal. The cdf of $\hat{\beta}$ is defined piecewise according to whether $\hat{\beta}$ lies to the right of the singularity in Figure 1 ($\hat{\beta} > \lambda$) or to the left ($\hat{\beta} < \lambda$). If we write $\hat{\beta} = g(m_{zu})$, then, for $\beta < \lambda$, $\text{prob}(\hat{\beta} < \theta) = \text{prob}(m_{zu} < g^{-1}(\theta)) - \text{prob}(m_{zu} < -\lambda\gamma m_{\varepsilon\varepsilon})$. For $\hat{\beta} > \lambda$, $\text{prob}(\hat{\beta} < \theta) = \text{prob}(m_{zu} > -\lambda\gamma m_{\varepsilon\varepsilon}) + \text{prob}(m_{zu} < g^{-1}(\theta))$. The cdf of m_{zu} is

$$\Phi\left(\frac{m_{zu}}{\sqrt{\gamma^2 m_{\varepsilon\varepsilon}^2 V/n}}\right)$$

where Φ is the standard normal cdf. Making the appropriate substitutions, the cdf of the instrumental variables estimator is as given in (3):

$$(3) \qquad \begin{aligned} &\text{For } \beta < \lambda, \quad F(\hat{\beta}) = \Phi\left(\frac{\hat{\beta}}{1-\hat{\beta}/\lambda}\frac{1}{\sqrt{V/n}}\right) - \Phi\left(\frac{-\lambda}{\sqrt{V/n}}\right), \\ &\text{for } \beta > \lambda, \quad F(\hat{\beta}) = 1 - \Phi\left(\frac{-\lambda}{\sqrt{V/n}}\right) + \Phi\left(\frac{\hat{\beta}}{1-\hat{\beta}/\lambda}\frac{1}{\sqrt{V/n}}\right). \end{aligned}$$

In comparison, the cdf of the asymptotic distribution is $F^A(\hat{\beta}) = \Phi(\hat{\beta}/\sqrt{V^A/n})$.

It is useful in thinking about the bias of instrumental variables to look at the relation between λ and the asymptotic bias of ordinary least squares. Note that λ^{-1} is the (population value of) the regression coefficient of x on u. When the regressor and the error term are uncorrelated, $\lambda^{-1} = 0$, ordinary least squares is asymptotically unbiased. More generally,

$$\text{plim}\left(\hat{\beta}_{\text{OLS}} - \beta\right) = \lambda \cdot \frac{\sigma_u^2}{\lambda^2\sigma_\varepsilon^2 + \sigma_u^2}.$$

Thus $0 < |\,\text{plim}(\hat{\beta}_{\text{OLS}}) - \beta\,| < |\lambda|$. Note that as σ_u^2 grows large, V grows large and the asymptotic bias of ordinary least square converges to λ.

$\lambda = 1, \qquad V/n = 4$

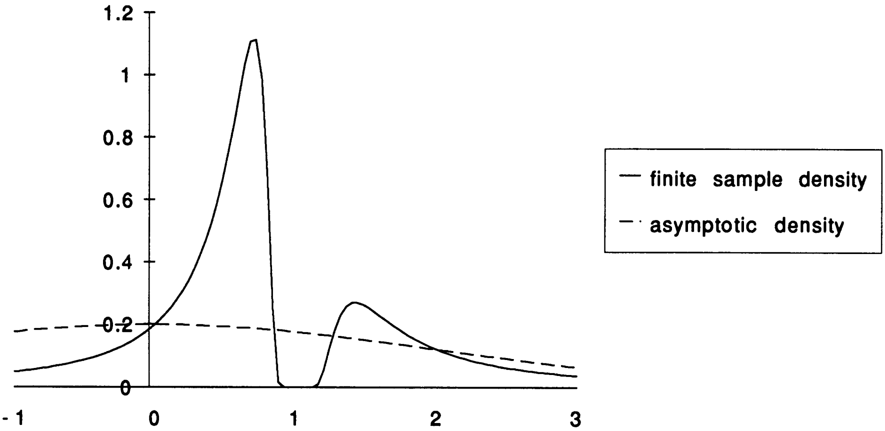

FIGURE 2.—Finite sample and asymptotic density functions for instrumental variables.

2. CHARACTERIZATION OF THE DISTRIBUTION OF THE INSTRUMENTAL VARIABLES ESTIMATOR

In this section we characterize the shape of the distribution of the instrumental variable estimator. The characterization takes two forms. We compare the actual distribution to the familiar centered-around-zero bell curve of the asymptotic distribution. We also prove a number of propositions which describe how the instrumental variable distribution changes when the parameters λ and γ change.

Figure 2 shows a representative sample of the density of $\hat{\beta}$, with the corresponding asymptotic approximation drawn for comparison. The region shown is ± 2 asymptotic standard deviations around λ. The first striking characteristic about the picture of the true distribution is that it is bimodal and has a minimum at λ.

The derivative of the density of $\hat{\beta}$ is given in (4):

$$f'(\hat{\beta}) = \frac{f(\hat{\beta})}{1 - \hat{\beta}/\lambda} \cdot \left[\frac{2}{\lambda} - \frac{\hat{\beta}}{V(1 - \hat{\beta}/\lambda)^2} \right]. \tag{4}$$

The density $f(\,)$ has three critical points. There are two maxima, so the distribution is bimodal. (Phillips and Hajivassiliou (1987, p. 2) also notes the bimodality of this estimator.) The modes are given by setting $f' = 0$ and applying the quadratic formula

$$\text{modes}\left(\frac{\hat{\beta}}{\lambda}\right) = 1 + \frac{\lambda^2}{4V/n} \pm \sqrt{\left(1 + \frac{\lambda^2}{4V/n}\right)^2 - 1}. \tag{5}$$

Note, from inspection of (5), that one mode occurs between zero and λ and the other to the right of λ. As γ goes to 0, V/n goes to ∞, the right-side of (5) goes to one, so both modes approach λ. As $V/n \to 0$, $n \to \infty$ for example, the modes go to zero and ∞ respectively. The third critical point is a minimum at $\hat{\beta} = \lambda$.

If the position of the modes relative to λ depends on V/n, what does V/n depend on? First, for any fixed parameters, as the sample size grows, V/n goes to zero. If one envisions watching the modes spreading towards zero and ∞ as the sample size grows, then one sees the process of convergence in distribution of $f(\hat{\beta})$ to the asymptotic distribution. Second, in a finite sample, if the instrument is poorly correlated with the regressor, then V/n will be large. For example, in our parameterization as γ becomes

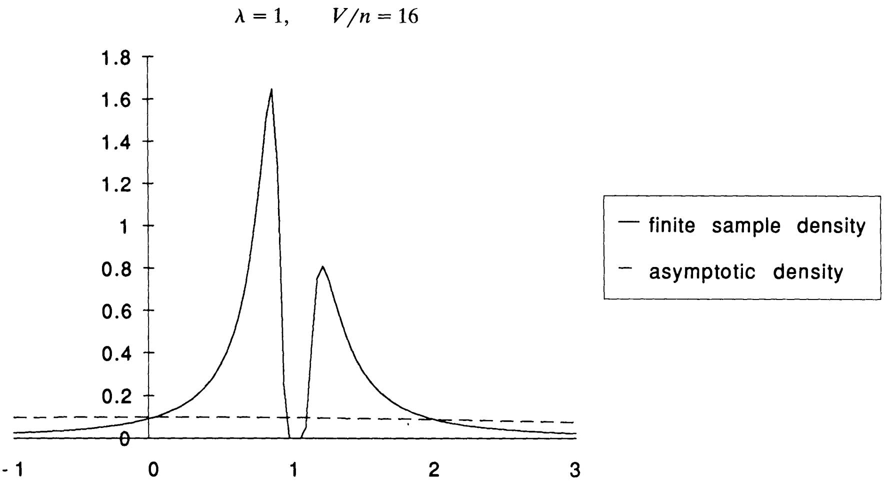

FIGURE 3.—Finite sample and asymptotic density functions for instrumental variables.

smaller in absolute value, V/n grows without limit. The conventional wisdom is that with poor instruments V/n is large so that the asymptotic distribution of $\hat{\beta}$ is dispersed, as illustrated in Figure 2. However, with large V/n, the asymptotic approximation is a poor approximation. The distribution of $\hat{\beta}$ may be quite concentrated, though it has fat tails, around a point away from the true parameter value.

Figure 3 shows the distribution of $\hat{\beta}$ for a case in which V/n is large relative to λ. V/n is set to 16 in Figure 3, as compared to 4 in Figure 2. The region of $\hat{\beta}$ shown is the same in Figures 2 and 3. Asymptotic distribution theory suggests that the fraction of the probability mass shown in Figures 2 and 3 falls from .62 to .37. In fact, the fraction rises from .82 to .90. The small sample distribution is becoming more concentrated, not less, as V/n rises.

We now turn the question around and ask how much of the mass lies within ± 1.96 asymptotic standard deviations of zero. For the asymptotic distribution, the answer is always .95. The correct answer depends on both λ and V/n. In Figure 2 the true mass is .91, while in Figure 3 the true mass is .97.

Figure 4 shows the distribution of $\hat{\beta}$ for a small asymptotic variance ($V/n = \frac{1}{2}$). Here, the asymptotic approximation is better than in Figure 2. According to asymptotic distribution theory, the fraction of the probability mass shown rises from .62 to .92. In fact, the fraction falls from .82 to .78. In Figure 4, .79 of the true mass lies within ± 1.96 asymptotic standard deviations of zero.

In general, the accuracy of "confidence intervals" based on the asymptotic distribution depends on $\lambda, V/n$, and the location of the region under consideration.

We turn now to several lemmas which give a more precise characterization of the distribution of $\hat{\beta}$.

LEMMA 1: *When the regressor and the regression error are uncorrelated, the asymptotic approximation is actually the true density.*

PROOF: This is the case where $\lambda^{-1} = 0$. The proof is by inspection of (1) and (2). Of course, in this case one would prefer ordinary least squares to IV estimation.

LEMMA 2: *At the point of concentration, λ, the true density equals zero.*

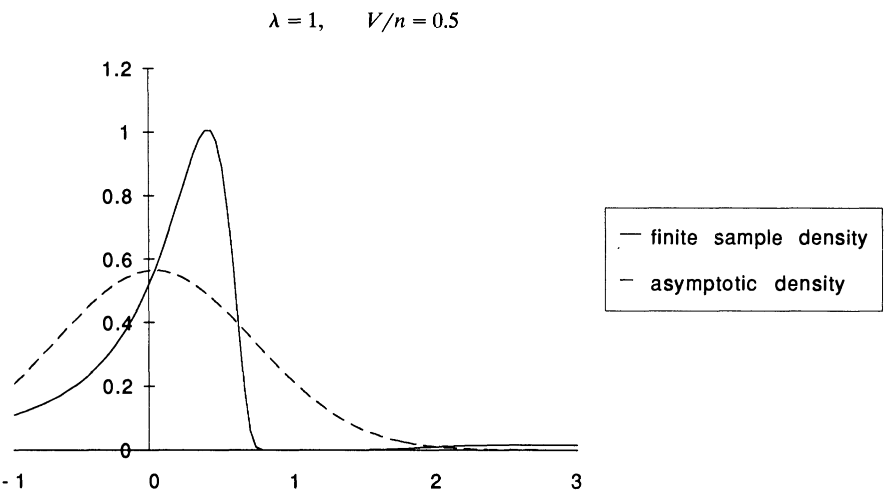

FIGURE 4.—Finite sample and asymptotic density functions for instrumental variables.

PROOF: As $\hat{\beta} \to \lambda$, $f(\hat{\beta}) \to 0$, by l'Hôpital's rule applied to equation (1).

LEMMA 3: *At the point of concentration, the true and asymptotic cdfs are equal,* $F(\lambda) = F^A(\lambda)$.

PROOF: Consider the first line of (3). As $\hat{\beta} \to \lambda$ from the left,

$$\Phi\left(\frac{\hat{\beta}}{1-\hat{\beta}/\lambda}\frac{1}{\sqrt{V/n}}\right) - \Phi\left(\frac{-\lambda}{\sqrt{V/n}}\right) \to \Phi(\infty) - \Phi\left(\frac{-\lambda}{\sqrt{V/n}}\right) = \Phi\left(\frac{\lambda}{\sqrt{V/n}}\right).$$

COROLLARY 3.1: $0 < |Median\ (\hat{\beta})| < |\lambda|$.

PROOF: For $\lambda > 0$, $F(0) = \Phi(0) - \Phi(-\lambda/\sqrt{V/n}) < \frac{1}{2}$ from the first line of (3) and, by the preceding lemma, $F(\lambda) = \Phi(\lambda/\sqrt{V/n}) > \frac{1}{2}$. For $\lambda < 0$, $F(0) = 1 - \Phi(-\lambda/\sqrt{V/n}) + \Phi(0) > \frac{1}{2}$, from the second line of (3), and $F(\lambda) = \Phi(\lambda/\sqrt{V/n}) < \frac{1}{2}$.

COROLLARY 3.2: *As* $\gamma \to 0$, *the median of* $\hat{\beta} \to \lambda$.

PROOF: From Lemma 3, $F(\lambda) = F^A(\lambda) = \Phi(\lambda/\sqrt{V/n})$. As $\gamma \to 0$, $V/n \to \infty$, $F(\lambda) \to \Phi(0) = \frac{1}{2}$.

LEMMA 4: *As* $\gamma \to 0$, *the distribution of* $\hat{\beta}$ *becomes concentrated around* λ *in the sense that* $\lim_{\gamma \to 0} \mathrm{Prob}(\lambda - \theta < \hat{\beta} < \lambda + \theta) = 1$, *for all* $\theta > 0$.

PROOF:

$$\begin{aligned}\operatorname{Prob}(\lambda-\theta<\hat{\beta}<+\theta) &= F_{\hat{\beta}>\lambda}(\lambda+\theta)-F_{\hat{\beta}<\lambda}(\lambda-\theta)\\ &=\left\{1-\Phi\left(\frac{-\lambda}{V/n}\right)+\Phi\left(\frac{\lambda+\theta}{1-(\lambda+\theta)/\lambda}\frac{1}{\sqrt{V/n}}\right)\right\}\\ &\quad-\left\{\Phi\left(\frac{\lambda-\theta}{1-(\lambda-\theta)/\lambda}\frac{1}{\sqrt{V/n}}\right)-\Phi\left(\frac{-\lambda}{\sqrt{V/n}}\right)\right\}\\ &=1+\Phi\left(-\left(\frac{\lambda^2}{\theta}+\lambda\right)\frac{1}{\sqrt{V/n}}\right)-\Phi\left(\left(\frac{\lambda^2}{\theta}-\lambda\right)\frac{1}{\sqrt{V/n}}\right).\end{aligned}$$

But as $\gamma \to 0$, the values of each of the cdfs in the last expression go to one-half, cancelling one another, so the probability goes to one.

LEMMA 5: *The first moment of the distribution of $\hat{\beta}$ is infinite and of sign opposite to λ. (The absence of a finite first moment is a special case of the proofs by Richardson* (1968) *and Sawa* (1969) *that the number of finite moments for 2SLS equals the number of overidentifying restrictions. Thus, unlike our other results, the intuition of Lemma* 5 *does not extend to more general instrumental variable estimators. It is nonetheless interesting to note that mean of the bias is infinitely negative in the case in which any other reasonable measure of central tendency is positive.)*

PROOF: Some intuition may be gained by looking at Figure 1. Just to the left of the discontinuity, $\hat{\beta}$ takes on infinite positive values. Just to the right of the discontinuity, $\hat{\beta}$ takes on infinite negative values. Since there is more mass of m_{zu} closer to zero, the negative values outweigh the positive ones. (This assumes $\lambda > 0$. For $\lambda < 0$, the same line of argument leads to an infinitely positive first moment.)

The mean of $\hat{\beta}$ is $\int_{-\infty}^{\infty}\hat{\beta}f(\hat{\beta})d\beta$. Define $\delta=\beta/(1-\beta/\lambda)$. Noting that $\beta=\delta\lambda/(\lambda+\delta)$, that $d\beta=[\lambda/(\lambda+\delta)]^2 d\delta$, and appropriately rearranging the limits of integration, we can make the change of variables to find the mean of $\hat{\beta}$ equals

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi V/n}}\frac{\lambda\delta}{\lambda+\delta}\exp\left[\frac{-1}{2V/n}\delta^2\right]d\delta. \tag{6}$$

Assume $\lambda > 0$. (The proof for $\lambda < 0$ is analogous.) Divide the definite integral in (6) into three regions: $(-\infty,-2\lambda)$, $(-2\lambda,0)$, and $(0,\infty)$. Over the first and third region, the integrand is everywhere positive and $\exp[-\delta^2/(2V/n)]$ goes to zero rapidly, so the integral in these regions is positive and finite. Subdivide the middle region into the subregions $(-2\lambda,-\lambda)$ and $(-\lambda,0)$. Over the first subregion perform the change of variable $s=-\lambda-\delta$ and over the second subregion perform the change of variable $s=-\lambda+\delta$. The integral becomes

$$\begin{aligned}&\int_{\lambda}^{0}\frac{1}{\sqrt{2\pi V/n}}\frac{\lambda(s+\lambda)}{s}\exp\left[-\frac{1}{2V/n}(s+\lambda)^2\right](-1)ds\\ &\quad+\int_{0}^{\lambda}\frac{1}{\sqrt{2\pi V/n}}\frac{\lambda(s-\lambda)}{s}\exp\left[-\frac{1}{2V/n}(s-\lambda)^2\right]ds\\ &=\frac{\lambda}{\sqrt{2\pi V/n}}\int_{0}^{\lambda}\frac{s+\lambda}{s}\exp\left[-\frac{1}{2V/n}(s+\lambda^2)\right]\\ &\quad+\frac{(s-\lambda)}{s}\exp\left[-\frac{1}{2V/n}(s-\lambda)^2\right]ds.\end{aligned} \tag{7}$$

The first term in the integrand of (7) is positive within the limits of integration and the second term is negative. The first term is less than $((s+\lambda)/s)\ \exp[-(1/(2V/n))\lambda^2]$, while the second term is greater in absolute value than $((s-\lambda/s)\exp[-(0/(2V/n))]$. Thus, the second, negative, integrand dominates. The definite integral of $-\lambda/s$ is

$$[-\lambda \log s]_0^{\lambda} = -\lambda[\log \lambda - \log 0] = -\infty. \qquad Q.E.D.$$

3. INTUITION

We have, with regard to the location of the probability mass of $\hat{\beta}$, three sets of results which appear to be at odds with one another. First, asymptotic distribution theory asserts that the distribution of $\hat{\beta}$ is approximately bell-shaped and centered around zero. Second, the absence of finite moments of $\hat{\beta}$ suggests that the distribution is fat-tailed. Third, we have shown that as $\gamma \to 0$, the mass concentrates around λ. How can these three statements be reconciled?

Return to Figure 1, which shows the correspondence between m_{zu} and $\hat{\beta}$. Since m_{zu} is normal, its mass *is* bell-shaped and centered around zero. Suppose γ is large, so the variance of m_{zu} is small; then most of the mass will be close to zero, say in the region marked *AA*. In this region, the mapping from m_{zu} to $\hat{\beta}$ is approximately linear, so a normal distribution on m_{zu} induces a normal distribution, centered around zero, on $\hat{\beta}$. Thus for a small asymptotic variance, the asymptotic distribution is a good approximation.

Suppose that γ is somewhat smaller, so that most of the mass of m_{zu} falls in the region marked *BB*. In this case, a significant portion of the mass of m_{zu} lies near the singularity, inducing values of $\hat{\beta}$ lying far out in the tails. This explains why the moments do not exist.

Finally, suppose that γ is smaller still, so that most of the mass of m_{zu} falls in the region marked *CC*. In this case, $\hat{\beta}$ is almost always close to the point of concentration, λ. Thus as γ grows small, most of the mass of $\hat{\beta}$ concentrates around λ.

4. CONCLUSION

We have shown that the true distribution of the instrumental variable estimator looks very little like the asymptotic approximation. In the case we study, the distribution is bimodal, fat-tailed, and may be heavily concentrated around a point closer to the probability limit of least squares than to the true parameter estimate. In a companion paper, Nelson and Startz (forthcoming), we use some of the analytical results presented here together with Monte Carlo studies to look at the distribution of test statistics based on instrumental variable estimation.

Department of Economics, University of Washington, Seattle, WA 98195, U.S.A.

Manuscript received June, 1988; final revision received July, 1989.

REFERENCES

ANDERSON, T. W. (1982): "Some Recent Developments on the Distributions of Single-Equation Estimators," in *Advances in Econometrics*, ed. by W. Hildenbrand. Amsterdam: North Holland Publishing Co., pp. 109–121.

BASMANN, R. L. (1974): "Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal," in *Frontiers of Quantitative Economics*, Vol. II, ed. by M. D. Intrilligator and D. A. Kendrick. Amsterdam: North Holland Publishing Co., pp. 209–288.

FIELLER, E. C. (1932): "The Distribution of the Index in a Normal Bivariate Population," *Biometrica*, 24, 428–440.

HINKLEY, D. V. (1969): "On the Ratio of Two Correlated Normal Random Variables," *Biometrika*, 56, 635–639.

JOHNSON, N. L., AND S. KOTZ (1972): *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley and Sons, Inc., 1972.

MARIANO, R. S., AND J. B. MCDONALD (1979): "A Note on the Distribution Functions of LIML and 2SLS Structural Coefficient in the Exactly Identified Case," *Journal of the American Statistical Association*, 74, 847–848.

MARSAGLIA, G. (1965): "Ratios of Normal Variables and Ratios of Sums of Uniform Variables," *Journal of the American Statistical Association*, 60, 193–204.

NELSON, C. R., AND R. STARTZ (1990): "The Distribution of the Instrumental Variable Estimator and Its t-Ratio When the Instrument is a Poor One," *Journal of Business*, forthcoming January, 1990.

PHILLIPS, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," *Handbook of Econometrics*, Vol. 1, Chapter 8. Amsterdam: North-Holland.

——— (1980): "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing $n+1$ Endogenous Variables," *Econometrica*, 48, 861–878.

PHILLIPS, P. C. B., AND V. A. HAJIVASSILIOU (1987): "Bimodal t-Ratios," Cowles Foundation Discussion Paper No. 842.

RICHARDSON, D. H. (1968): "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214–1226.

SAWA, T. (1969): "The Exact Finite Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–936.