NBER TECHNICAL WORKING PAPER SERIES

THE DISTRIBUTION OF THE INSTRUMENTAL VARIABLES ESTIMATOR AND ITS t-RATIO WHEN THE INSTRUMENT IS A POOR ONE

Charles R. Nelson

Richard Startz

Technical Working Paper No. 69

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1988

This research is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Technical Working Paper #69
September 1988

THE DISTRIBUTION OF THE INSTRUMENTAL VARIABLES ESTIMATOR AND ITS t-RATIO WHEN THE INSTRUMENT IS A POOR ONE

ABSTRACT

When the instrumental variable is a poor one, in the sense of being weakly correlated with the variable it proxies, the small sample distribution of the IV estimator is concentrated around a value that is inversely related to the feedback in the system and which is often further from the true value than is the plim of OLS. The sample variance of residuals similarly becomes concentrated around a value which reflects feedback and not the variance of the disturbance. The distribution of the t-ratio reflects both of these effects, stronger feedback producing larger t-ratios. Thus, in situations where OLS is badly biased, a poor instrument will lead to spurious inferences under IV estimation with high probability, and generally perform worse than OLS.

Charles R. Nelson
Department of Economics, DK-30
University of Washington
Seattle, WA 98195

Richard Startz
Department of Economics, DK-30
University of Washington
Seattle, WA 98195

1. Introduction

As motivation for the investigation of the properties of the instrumental variables (IV) estimator when the instrument is a poor one (by which we mean weakly correlated with explanatory variables), consider an example in the spirit of the recent literature on the consumption function. Hall (1978) pointed out that when future income is uncertain, maximization of expected utility implies

$$(1.1) \qquad U'(C_{t+1}) = [\frac{(1+\delta)}{(1+r)}]U'(C_t) + \epsilon_{t+1}$$

which is a stochastic Euler equation where $U(\cdot)$ is the one period flow utility function, δ is the rate of time preference, r is the real interest rate, and ϵ_{t+1} is a stochastic error that is uncorrelated with C_t. If utility of consumption is quadratic then this equation is linear and can be estimated by least squares. Suppose, however, that marginal utility is quadratic (or is to be approximated as a quadratic) in which case we have, assuming $\delta = r$,

$$(1.2) \qquad C_{t+1} + \beta C_{t+1}^2 = C_t + \beta C_t^2 + \epsilon_{t+1}$$

so the equation to be estimated is

$$(1.3) \qquad C_{t+1} - C_t = \beta(C_{t+1}^2 - C_t^2) + \epsilon_{t+1}.$$

Now C_{t+1}^2 is contemporaneous with ϵ_{t+1} and IV estimation seems an obvious procedure to ensure consistency in the estimate of β, with the lagged value of the right–hand–side variable being the obvious instrument. Application of *IV* in this context can be thought of as a simple example of the method of Euler equation estimation proposed in Hansen and Singleton (1982).

Under the null hypothesis $\beta = 0$ we would have $C_{t+1}0 = C_t + \epsilon_{t+1}$ which is essentially Hall's empirical model for consumption. Using this as the generating mechanism we calculated a realization of length 120 (roughly Hall's sample period) with $\epsilon \sim i.i.d.\ N(0,1)$ and $C_0 = 100$. The IV results for model (1.3) with a constant included both in the regression and instrument list were:

$$
\begin{aligned}
b_{IV} &= .0059 \\
&\quad (.0005) \\
t(b_{IV}) &= 12.69 \\
R^2 &= .99 \\
S.E. &= .08 \\
D.W. &= 1.95
\end{aligned}
\tag{1.4}
$$

The t–ratio tells us that β is not zero when in fact it is zero. The R^2 tells us that the first difference of C^2 explains almost all of the variation in the change in consumption, when in fact it explains none. Accordingly, the standard error of the regression is far less than the true value of 1.0. These misleading results are not due to an aberrant sample. In 10,000 replications of this experiment, the median t was 11.5 and the median standard error of the regression was .06. Rather, the problem can be traced to the fact that the instrument is only weakly correlated with the right–hand side variable (the sample squared correlation is .009).

In the remainder of the paper we study the influence of weak correlation between the instrument and the right–hand side variable on the finite sample distribution of the IV estimator b_{IV}, the standard error of the regression, the estimated variance of b_{IV}, and on its t–ratio. We find that the effect of weak correlation is to concentrate the small sample distribution of the IV estimator around a value that is inversely related to the feedback from the dependent variable to the explanatory variable. That value is frequently further from the true parameter than is the plim of OLS. The sample variance of residuals also becomes concentrated around a value that reflects the strength of feedback and not the actual variance we wish to estimate. The distribution of the t–ratio reflects both of these effects, and its central tendency is therefore determined by the feedback in the system, stronger feedback producing larger sample t–values. In other words, in those cases where OLS is a poor estimator, IV with a poor instrument will be even worse. The spurious results seen in the example above are to be expected when the instrument is poor.

2. Concentration of the IV Estimator

Consider a simple system in which variables x and z are cross-correlated with representation

$$x = \lambda u + \epsilon \tag{2.1}$$

$$z = \gamma \epsilon + v$$

where λ and γ are fixed parameters and ϵ, u, and v are serially random disturbances, not cross–correlated, and are normal with mean zero and variances $\sigma_\epsilon^2, \sigma_u^2$ and σ_v^2 respectively. The relationship of interest to the investigator is

$$(2.2) \qquad y = \beta x + u.$$

We would like to estimate β and test the null hypothesis $\beta = \beta^*$.

I will be useful to keep in mind some basic characteristics of the system. The least squares estimator of β is b_{OLS} and

$$(2.3) \qquad plim(b_{OLS} - \beta) = \frac{\lambda \sigma_u^2}{(\lambda^2 \sigma_u^2 + \sigma_\epsilon^2)}$$

so of course it is the feedback from u to x through the parameter λ that creates the need for the IV estimator. For expository purposes it will be helpful to consider the special case of unit disturbance variances for which

$$(2.3a) \qquad plim\ (b_{OLS} - \beta) = \frac{\lambda}{(\lambda^2 + 1)}$$

The suitability of z as an instrument for x depends on its being uncorrelated with u (true by assumption) but correlated with x. The squared correlation between x and z is

$$(2.4) \qquad \rho_{xz}^2 = \frac{(\gamma \sigma_\epsilon^2)^2}{(\lambda^2 \sigma_u^2 + \sigma_\epsilon^2)(\gamma^2 \sigma_\epsilon^2 + \sigma_v^2)}$$

which of course will be nonzero if and only if γ is nonzero. In the unit variance case we have

$$(2.4a) \qquad \rho_{xz}^2 = \frac{\gamma^2}{(\lambda^2 + 1)(\gamma^2 + 1)}$$

The IV estimator of β is the ratio

(2.5) $$b_{IV} = \frac{m_{zy}}{m_{zx}}$$

where m denotes the sample second moment between the indicated variables. The small sample properties of the instrumental variable estimator 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 b_{OLS}, and by Anderson (1982), who discusses approximations to the *cdf*. Basmann and Mariano and McDonald point out that b_{OLS} is the ratio of two correlated normal random variables an so that its distribution may be studied using Fieller's (1932) results. (See also Johnson and Kotz (1972), pp. 123–124, Hinkley (1969) and Marsaglia (1965)). Nelson and Startz (1988) extends the work just cited by characterizing the *pdf* and *cdf* of the instrumental variables estimator, and comparing them to the asymptotic approximations, as the "quality" of the instruments varies. In this paper, we focus on the limiting effects of low correlation between the instrument and the explanatory variable.

Substituting for the observed variables using the equation of the system we have

(2.6) $$b_{IV} = \beta + \frac{m_{zu}}{\lambda m_{zu} + m_{z\epsilon}}$$

Now think of ϵ and v as being drawn once and then fixed as we sample from the distribution of u. This fixes z in repeated samples, but not x since x includes the random element u. Equivalently, consider the distribution of y and x conditional on ϵ and v. Thus $m_{z\epsilon}$ in (2.6) is fixed and nonstochastic, but m_{zu} is a linear combination of the random u's with the fixed z's as weights and therefore the mean of m_{zu} is zero. The particular value of $m_{z\epsilon}$ will depend on the ϵ's and v's drawn, but will be within sampling error of the population moment which is $\gamma \cdot \sigma_\epsilon^2$. For purposes of exposition we assume this is the value of $m_{z\epsilon}$ so we have

$$(2.7) \qquad (b_{IV} - \beta) \approx \frac{m_{zu}}{\lambda m_{zu} + \gamma \cdot \sigma_\epsilon^2} = \frac{1}{\lambda + (\gamma \cdot \sigma_\epsilon^2)/m_{zu}}$$

For the unit variance case we have

$$(2.7a) \qquad b_{IV} - \beta \approx \frac{1}{\lambda + \gamma / m_{zu}}$$

Consider now how the sampling error for b_{IV} depends on the realization of m_{zu} in (2.7) and (2.7a). When m_{zu} takes on its expected value of zero then $(b_{IV} - \beta)$ is zero. When m_{zu} is very large relative to $(\gamma \cdot \sigma_\epsilon^2)$ then $(b_{IV} - \beta)$ is approximately λ^{-1}. When m_{zu} has the value $-\frac{\gamma \cdot \sigma_\epsilon^2}{\lambda}$ then $(b_{IV} - \beta)$ is infinite, and for values of m_{zu} in the neighborhood of that value we will get very large sampling errors. These features are readily seen in the graph of $(b_{IV} - \beta)$ as a function of m_{zu} in Figure 1.

For illustrative purposes in Figure 1 the value of λ is unity, γ is .05, and σ_ϵ^2 is unity. The singularity value for m_{zu} at which $(b_{IV} - \beta)$ diverges to $\pm\infty$ is -0.05. As the value of m_{zu} moves away from the singularity value in either direction, the error function $(b_{IV} - \beta)$ approaches λ^{-1} asymptotically. The error function is symmetric around the singularity and around the asymptote. To the right of the singularity, the error function crosses the zero line from below (at which point $b_{IV} = \beta$) where $m_{zu} = 0$. The rate of convergence of the error function toward the asymptote also depends on γ, λ, and σ_ϵ^2. In particular, at a value of m_{zu} that is, say, δ from the singularity the error function differs from its asymptotic value λ^{-1} by the amount $(\gamma \cdot \sigma_\epsilon^2)/(\lambda^2 \cdot \delta)$. The smaller is γ or σ_ϵ^2 or the larger is λ, the closer the singularity will be to zero and the more rapidly will the error function converge to λ^{-1}.

The sampling range of m_{zu} is centered on its mean at zero, and its variance is

$$
\begin{aligned}
var(m_{zu}) &= E(\frac{1}{n}\sum zu)^2 \\
&= \sigma_u^2 \cdot m_{zz} \cdot \frac{1}{n} \qquad (2.8) \\
&\approx \sigma_u^2 \cdot \frac{\gamma^2\sigma_\epsilon^2 + \sigma_v^2}{n}
\end{aligned}
$$

where n is sample size and the last line is an approximation which replaces the sample variance of z with its population variance. For the unit variance case we have

$$
var\ (m_{zu}) \approx \frac{\gamma^2 + 1}{n}. \qquad (2.8a)
$$

The range depicted in Figure 1, $\pm.10$, is therefore roughly ± 1 standard deviation for the unit variance case with $n = 100$ observations. Clearly, sampling errors will often be large if the region around the singularity is visited often by m_{zu}. As a result, the moments of $(\beta_{IV} - \beta)$ do not exist, as shown by Nelson and Startz (1988).

A smaller value of γ or σ_ϵ^2 or a larger value of λ will shift the singularity closer to zero and increase the rate of convergence of the error function toward λ^{-1}. Any of these changes would correspond to a smaller ρ_{xz}^2 (the "first stage" r^2 in the two stage least squares interpretation of IV) and therefore to z being a poorer instrument. In Figure 2 the value of γ is reduced to 0.01 from 0.05 in Figure 1. The value of ρ_{xz}^2 drops from 1.25×10^{-3} to 5.0×10^{-5}. Note that the value of m_{zu} at the singularity in Figure 2 is much closer to its mean (-0.01 instead of -0.05) and that the error function converges more rapidly to its asymptote. Most realizations of m_{zu} will correspond to a value of $(b_{IV} - \beta)$ closer to λ^{-1} than in Figure 1. The range of m_{zu} depicted in Figure 2 is again ± 1 standard deviation for $n = 100$.

It is clear that in the situation depicted in Figure 2 most realizations of m_{zu} will result in a value of b_{IV} close to λ^{-1} and that this can be thought of as the central tendency of the sampling distribution. It is not the mean, which does not exist, and it is not the mode since the probability that $(b_{IV} - \beta) = \lambda^{-1}$ is zero. It may be useful to think of λ^{-1} as the *point of concentration* of the error distribution as $\gamma \cdot \sigma_\epsilon^2/\lambda$ gets small. Occasionally, very large outliers in $(b_{IV} - \beta)$ will occur. *Rarely will the sampling error $(b_{IV} - \beta)$ be closer to zero than to λ^{-1}.*

We conclude then that realizations of $(b_{IV} - \beta)$ will cluster around λ^{-1} if the value of m_{zu} at the singularity is small relative to the standard deviation of m_{zu}. This condition requires that

$$|\frac{-m_{z\epsilon}}{\lambda}| \ll SD(m_{zu}) \tag{2.9}$$

where the left–hand side is the value of m_{zu} at the singularity. From (2.7) and (2.8) and using the population moments corresponding to m_{zu} and $m_{z\epsilon}$ we have the condition

$$(\frac{\gamma \cdot \sigma_\epsilon^2}{\lambda})^2 \ll \sigma_u^2 \cdot \frac{\gamma^2\sigma_\epsilon^2 + \sigma_v^2}{n}, \tag{2.10}$$

After substitution and rearrangement (2.1) becomes

$$\frac{1}{\rho_{xz}^2} \cdot \frac{\lambda^2 \cdot \sigma_u^2}{\lambda^2 \cdot \sigma_u^2 + \sigma_\epsilon^2} \gg n \tag{2.11}$$

Recall that ρ^2_{xz}, given by equation (2.4), is the population R^2 in the first stage regression of x on z in the two–stage least squares interpretation of IV. A small value of ρ^2_{xz} will tend to satisfy (2.11), so a low first stage R^2 is certainly an important element in producing spurious IV estimates. (Note, however, that the sample R^2 may have a significant upward bias. See Section 5.) For the unit error variance case with $\lambda = 1$, $\gamma = .01$, and $n = 100$ illustrated in Figure 2 we have

$$(2.12) \qquad \frac{1}{0.00005} \cdot \frac{1}{2} = 10^4 \gg 100 = n$$

so condition (2.11) is satisfied.

Note also that using equation (2.3), condition (2.11) can be rewritten as

$$(2.13) \qquad \frac{1}{\rho^2_{xz}} \cdot \lambda \cdot plim(b_{OLS} - \beta) >> n$$

so the more serious the bias in b_{OLS} the stronger the tendency for sampling errors in b_{IV} to cluster around λ^{-1}. In other words, the worse are the results of OLS, the stronger the tendency of IV to give a consistently spurious estimate.

It is interesting to compare the biases of the IV and OLS estimators in the poor instrument case, using the point of concentration and *plim* respectively as measures of central tendency. Referring to (2.3) and the above discussion of Figure 2, we have

$$(2.14) \qquad \frac{Bias\ OLS}{Bias\ IV} \approx \frac{\lambda^2}{\lambda^2 + \sigma^2_\epsilon/\sigma^2_u} \leq 1$$

which depends on the feedback from u to x measured by λ and on the ratio of the variances of ϵ and u. Regardless of the specific values of the parameters, the bias in OLS is *less* than that of IV. Less surprisingly, the larger is λ (stronger feedback) the closer the bias ratio is to unity. When feedback is weak, OLS may be much less biased than IV. Using $\lambda = 1$ and unit variances as in Figure 2 the bias in OLS is only half that of IV.

The concentration of the sampling distribution of b_{IV} around the value λ^{-1} as a function of λ and γ is seen in the results of a Monte Carlo experiment shown in Table 1. Briefly, the data are generated by system (2.1) - (2.2) with unit variances and $\beta = 0$. Conditional on a realization of z, ϵ such that $m_{z\epsilon}$ is close to its population value, repeated samples of u's are drawn resulting in random x and y. Each sample is of length $n = 100$. Fractiles of b_{IV} are reported in Table 1 along with those of b_{OLS} for comparison. The asymptotic bias of b_{OLS} is 0.5 for $\lambda = 1$, and indeed that appears to be roughly the median of its distribution. For $\lambda = 1$ and $\gamma = .05$ (a moderately poor instrument), the values illustrated in Figure 1, the sampling distribution of b_{IV} is not concentrated around $\lambda^{-1} = 1$ but has median .63. The asymptotic variance formula would predict the fractiles given in the next column labeled "Asy." The striking difference is in the 1 and 99 fractiles which are much farther from zero than the asymptotic formula would suggest. These fat tails reflect values of m_{zu} close to the singularity.

When we reduce γ to 0.01 in the next column (a very poor instrument), the concentration effect becomes evident. This is the case illustrated in Figure 2. The sampling distribution is much less disperse than the asymptotic formula would predict, since the asymptotic variance is larger for a poorer instrument. Finally, for $\gamma = .001, b_{IV}$ is tightly distributed around $\lambda^{-1} = 1$, only the occasional outlier being evident in the 99 fractile. Asymptotic theory, in contrast, predicts a more disperse distribution as γ is reduced. Note that there is little overlap between the distributions of b_{IV} and b_{OLS}, with the latter being concentrated around $\lambda/(\lambda^2 + 1) = .5$.

3. Concentration of the residual Variance

In order to do statistical inference about β we need to estimate the variance of the regression disturbance, u. The estimate of σ_u^2 based on the sample variance of residuals for any estimated coefficient b is

$$
\begin{aligned}
s^2 &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - x_i b)^2 \\
&= [1 - (b-\beta)\lambda]^2 \cdot m_{uu} \\
&+ 2(b-\beta)[\lambda(b-\beta) - 1] \cdot m_{u\epsilon} \\
&+ (b-\beta)^2 \cdot m_{\epsilon\epsilon},
\end{aligned}
\tag{3.1}
$$

using $x = \lambda u + \epsilon$. Conditional on a realization of the ϵ's, $m_{\epsilon\epsilon}$ is fixed and $m_{u\epsilon}$ is a linear combination of the random u's. However, in the case that b is the IV estimator under condition (2.11) where it is close to λ^{-1} with high probability, then $(b - \beta) \approx \lambda^{-1}$ and we get

$$
\begin{aligned}
s_{IV}^2 &\approx \lambda^{-2} \cdot m_{\epsilon\epsilon} \\
&\approx \lambda^{-2} \cdot \sigma_\epsilon^2,
\end{aligned}
\tag{3.2}
$$

replacing the sample statistic $m_{\epsilon\epsilon}$ by its population counterpart in the second line.

Evidently the residual variance does not involve the variance of disturbances u in the case of a poor instrument, since (3.2) tells us that s_{IV}^2 will depend on σ_ϵ^2, the variance of the exogenous part of x, rather than on σ_u^2, the variance we would like to estimate. The reason is that the spurious value of b_{IV} tends to cancel u out of the residuals, leaving ϵ. To see this, note that we have,

$$
(3.3) \qquad \begin{aligned}
(y_i - x_i b_{IV}) &= (x_i\beta + u_i - x_i b_{IV}) \\
&\approx (u_i - x_i\lambda^{-1}) \\
&= (u_i - (u_i + \lambda^{-1}\epsilon)) \\
&= \lambda^{-1}\epsilon.
\end{aligned}
$$

As a consequence, the residual variance will tell us little about the value of σ_u^2.

Concentration of the sampling distribution of s_{IV}^2 around the value $\lambda^{-2} \cdot \sigma_\epsilon^2$ can be seen by the Monte Carlo results reported in Table 2. As in Table 1, the variances are all unity and we consider the effect of successively smaller values of γ. Using $\lambda = 10$ we expect that the sampling distribution of s_{IV}^2 will concentrate around $\lambda^{-2} = 0.01$ instead of the true value of $\sigma_u^2 = 1$. With $\gamma = .05$ the median is 0.013 but sometimes much larger estimates are obtained, corresponding to outliers in b_{IV}. As γ is reduced to .01 and then .001 in Table 2, the concentration effect is evident.

The consequence of concentration in the distribution of s_{IV}^2 is that when we come to calculate the standard error for b_{IV} using the classical formula we will be replacing σ_u^2 with an estimate that is spurious. It may be biased upward or downward depending on the relative magnitudes of σ_u^2 and $\lambda^{-2} \cdot \sigma_\epsilon^2$, but will not reflect the true value of σ_u^2 if the instrument is a poor one.

4. The Central Tendency of the t–Ratio for b_{IV}

The asymptotic variance of b_{IV} is given by

$$
(4.1) \qquad Asy\ \ Var(b_{IV}) = \frac{\sigma_u^2 \cdot m_{zz}}{n \cdot m_{zx}^2}
$$

which suggests that the variance of b_{IV} should be estimated by replacing σ_u^2 with s_{IV}^2. The square root of the resulting statistic provides a standard error for b_{IV} and one would hope that the ratio

$$t = \frac{(b_{IV} - \beta)}{\sqrt{s_{IV}^2 \cdot \frac{1}{n} \cdot \frac{m_{zz}}{m_{zx}^2}}} \tag{4.2}$$

would be approximately distributed as Student's t with $(n-1)$ degrees of freedom. This "t–ratio" can also be written as

$$t = \frac{(b_{IV} - \beta)}{\sqrt{s_{IV}^2 \cdot \frac{1}{n} \cdot \frac{1}{r_{zx}^2 \cdot m_{xx}}}} \tag{4.3}$$

where again r_{zx}^2 is the "first stage" r^2 in the regression of x on the instrument z.

In the case of a poor instrument we know that $(b_{IV} - \beta)$ will be concentrated on the value λ^{-1} and s_{IV}^2 on the value $\lambda^{-2} \cdot \sigma_\epsilon^2$. Therefore in this case we have

$$\begin{aligned} t &\approx \frac{\lambda^{-1}}{\sqrt{\lambda^{-2} \cdot \sigma_\epsilon^2 \cdot \frac{1}{n} \cdot \frac{1}{r_{zx}^2 \cdot m_{xx}}}} \\ &\approx \sqrt{\frac{n \cdot r_{zx}^2 \cdot m_{xx}}{\sigma_\epsilon^2}} \end{aligned} \tag{4.4}$$

Now, the sample statistic r_{zx}^2 is biased upward by approximately $\frac{1}{n}$. Its mean is

$$E(r^2) \approx \rho^2 + \frac{(1-\rho^2)}{n-1} = \frac{1}{n-1} + \frac{n-2}{n-1} \cdot \rho^2 \tag{4.5}$$

where ρ is the population correlation (see Johnson and Kotz, p. 244). If condition (2.11) holds it implies that $\rho^2 \ll (n-1)^{-1}$ in which case

$$E(r^2) \approx (n-1)^{-1}. \tag{4.5a}$$

The sample variance m_{xx} has mean $(\lambda^2\sigma_u^2 + \sigma_\epsilon^2)$. We surmise then that very roughly

$$E(t) \approx \sqrt{\frac{\lambda^2 \cdot \sigma_u^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2}} \tag{4.6}$$

which is the unit variance case becomes

$$E(t) \approx \sqrt{\lambda^2 + 1}. \tag{4.6a}$$

Evidently, the value of the t–ratio will reflect the size of the feedback coefficient λ when the instrument is a poor one. If feedback is strong then the t–ratio will tend to be large even if the null hypothesis is true. Further, note that the t–ratio will be positive if λ is positive and negative if λ is negative since the sign of t in (4.4) is given by the sign of λ. These effects are apparent in the Monte Carlo estimates of the fractiles of b_{IV} and the t–ratio reported in Table 3. Data are generated with $\lambda = 10$ and $n = 100$ as before. The first value of γ considered is $\gamma = 1.0$ in which case z is not poor enough an instrument to concentrate b_{IV} around $\lambda^{-1} = 0.1$. The median of the t–ratio is 0.42 and the distribution is strongly skewed to the right. For $\gamma = 0.01$ the concentration effect in b_{IV} is becoming more apparent, and the median of t rises to 3.68. Negative values of t become very infrequent. At $\gamma = 0.01$ and 0.001 the concentration effect on b_{IV} is strongly apparent, the median of t is about 7.3, the probability of a negative t has become negligible, and the probability of t being larger than $+2.0$ is greater than 0.75.

It is far from clear that the mean or other moments of t exist, but estimated means and standard deviations are given in Table 3. As expected, the estimated mean of t increases as γ declines, evidently approaching a value somewhat less than $\sqrt{\lambda^2+1}$. The standard deviation of t also increases as the instrument becomes poorer, and is over 6 instead of being unity.

Under what circumstances will the distribution of $t(b_{IV})$ approach Student's t? It should be intuitive from the discussion in this paper and the geometry of Figure 1 that when condition (2.11) is reversed b_{IV} becomes well behaved in the sense of having properties closer to those suggested by asymptotic theory. In the unit variance set–up those will be cases of larger γ and smaller λ. If λ is small, then the potential advantage of IV over OLS disappears. The value of λ that maximizes $plim(b_{OLS} - \beta)$ is $\lambda = 1$, so this is a case where we would hope IV would be an improvement over OLS. Table 4 shows how the sampling distributions of b_{IV} and $t(b_{IV})$ move toward the asymptotic distributions as we consider larger values of γ which make z a better instrument.

Along with the use of nonlinear instrumental variable methods for the estimation of Euler equations, Hansen and Singleton popularized testing a model's overidentifying restrictions as an overall check of model validity. In the 2SLS' context, the test statistic for overidentifying restrictions is SSR/s^2, which is distributed asymptotically $\chi^2(q-k)$, where SSR is the sum square residuals of the "second- stage" regression of the fitted Y on the fitted X (see Basmann (1960) and Startz (1983).) Just as for the t–statistic, if s^2 underestimates the true variance, the test for overidentifying restrictions will give too many false rejections.

We ran a Monte Carlo experiment to illustrate the problem. While our model to this point has been limited to the just identified case, a simple extension provides for two instruments, Z_1 and Z_2, where

$$(4.7) \qquad \begin{aligned} Z_1 &= \gamma\epsilon + v_1 + \delta v_2 \\ Z_2 &= \gamma\epsilon + \delta v_1 + v_2 \end{aligned}$$

Table 5 shows the results of the Monte Carlo. The variances were all set to unity and $\lambda = 10, \gamma = 0.001$, and $\delta = 0.999$. To illustrate that the problem is not limited to small "small samples," we set $n = 1200$. In the case shown in Table 5, the investigator will almost certainly falsely reject the overidentifying restrictions, rejecting over one-fourth of the time at the 1 percent critical value.

5. Suggestions for Practitioners

If the authors' own experience is any guide, the general impression among practitioners of econometrics is that the consequence of having a poor instrument is a large standard error and a low t–ratio. This paper leads us to the conclusion that the consequences of having a poor instrument are much more insidious than that; namely the bias in the estimated coefficient will be large relative to its calculated standard error if the feedback is strong. In these cases the bias of OLS will be smaller than that of IV in the sense defined in this paper.

The only protection against erroneous inference is to look directly at the correlation between the instrument and the explanatory variable, using inequality (2.11) to judge whether the correlation is too low. We would like to know whether

$$\frac{1}{\rho_{xz}^2} \cdot \frac{\lambda^2 \cdot \sigma_u^2}{\lambda^2 \cdot \sigma_u^2 + \sigma_\epsilon^2} \gg n \tag{5.1}$$

which depends on unknown parameters, but we can make use of the fact that

$$\frac{1}{\rho_{xz}^2} > \frac{1}{\rho_{xz}^2} \cdot \frac{\lambda^2 \cdot \sigma_u^2}{\lambda^2 \cdot \sigma_u^2 + \sigma_\epsilon^2} \tag{5.2}$$

since ρ_{xz}^2 can be estimated directly from the data on x and z. The approximate bias in sample r^2 is given by (4.5), so a bias–corrected estimate of ρ_{xz}^2 is given by

$$\hat{\rho}_{xz}^2 = \frac{n-1}{n-2} r_{xz}^2 - \frac{1}{n-2}; \tag{5.3}$$

for ρ^2 small. If we find that

$$\frac{1}{\hat{\rho}_{xz}^2} \gg n \tag{5.4}$$

we should consider that inequality (2.11) may apply and therefore be wary of using the instrument. For example, in the artificial data used in the introductory section of the paper, we have

$$\begin{aligned} \hat{\rho}_{xz}^2 &= \frac{119}{118} \cdot (.009) - \frac{1}{118} = .0006 \\ \frac{1}{.006} &= 1662 \gg 120, \end{aligned} \tag{5.6}$$

suggesting that the results reported in (1.4) are spurious, which indeed they are.

If we return to the 2SLS interpretation of instrumental variables, (5.4) is essentially equivalent to saying that the coefficients on the instruments are not very significant in the first–stage regression. Formula (5.4) is approximately the same as the statement that the first–stage $TR^2 \ll 2$. So in the one right–hand–side variable, one instrument, linear, instrumental variable problem, checking the significance of the first–stage regression provides an easy to implement safety check against the concentration phenomenon, although a less satisfactory test against the "fat–tail" problem.

In the context of estimating stochastic Euler equations, we would particularly caution against the use of lagged changes in consumption or lagged stock returns as instruments for current values, as advocated by Hansen and Singleton (1982). The quality of these instruments depends on the degree of serial correlation in these variables which is well known to be low.

Finally, we are obliged to caution that when the model includes multiple explanatory and instrumental variables, high first–stage R^2's do not provide an adequate warning against spurious inference. (It is easy to produce Monte Carlo results for the model $k = 2$ in which each first–stage R^2 is above .98, but where all the qualitative problems discussed above recur.) Our preliminary investigations suggest that correlation among the instruments is also relevant, a phenomenon which we plan to investigate in further research. The only safe strategy appears to be Monte Carlo investigation of the distribution of test statistics under the null hypothesis one wants to test.

REFERENCES

Anderson, T.W. "Some Recent Developments on the Distributions of Single–Equation Estimators." *Advances in Econometrics*, W. Hildebrand, ed., pp. 109–121, 1982.

Basmann, R.L., "On the Asymptotic Distribution of Generalized Linear Estimators," *Econometrica* 28, pp. 97–107, 1960.

Basmann, R.L., "Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal," in *Frontiers of Quantitative Economics*, Vol. II, M.D. Intrilligator and D.A. Kendrick, eds., pp. 209–288.

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

Hall, Robert E., "Stochastic Implications of the Life Cycle–Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, No. 6, Vol. 86 (December 1978), pp. 971–987.

Hansen, Lars Peter; and Singleton, Kenneth J., "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, No. 5, Vol. 50 (September 1982), pp. 1269–1286.

Hinkley, D.V., "On the Ratio of Two Correlated Normal Random Variables," *Biometrika* 56, 3, pp. 635–639, 1969.

Johnson, N.L. and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons, Inc., New York, 1972.

Mariano, R.S. and J.B. McDonald, "A Note on the Distribution Functions of LIML and 2SLS Structural Coefficient in the Exactly Identified Case," *Journal of the American Statistical Association*, No. 368, Vol. 74, pp. 847–848.

Marsaglia, G., "Ratios of Normal Variables and Ratios of Sums of Uniform

Variables," *Journal of the American Statistical Association* 60, pp. 193–204, 1965.

Nelson, C.R. and R. Startz, "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," University of Washington Discussion Paper, May 1988.

Startz, R. "Testing Rational Expectations by the Use of Overidentifying Restrictions," *Journal of Econometrics* 23 (1983), 343–395.

FIGURE 1

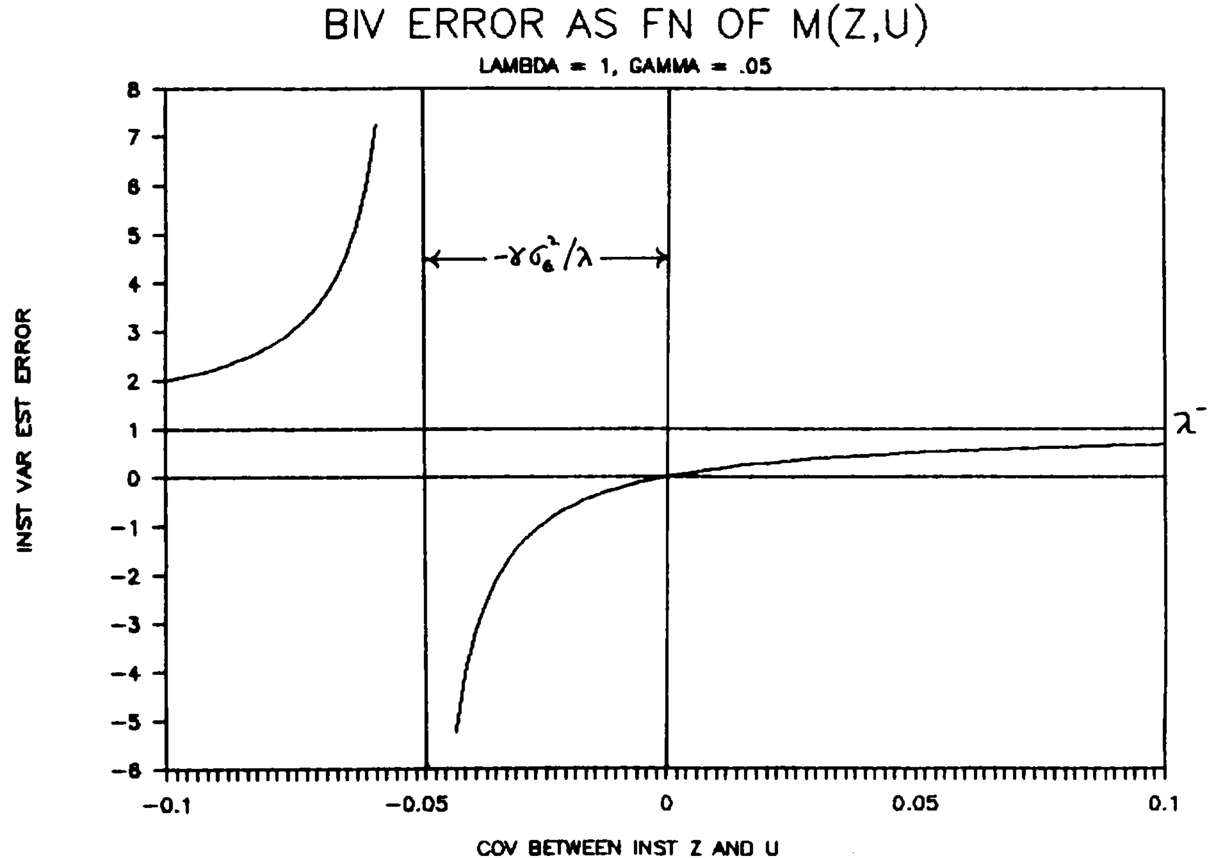
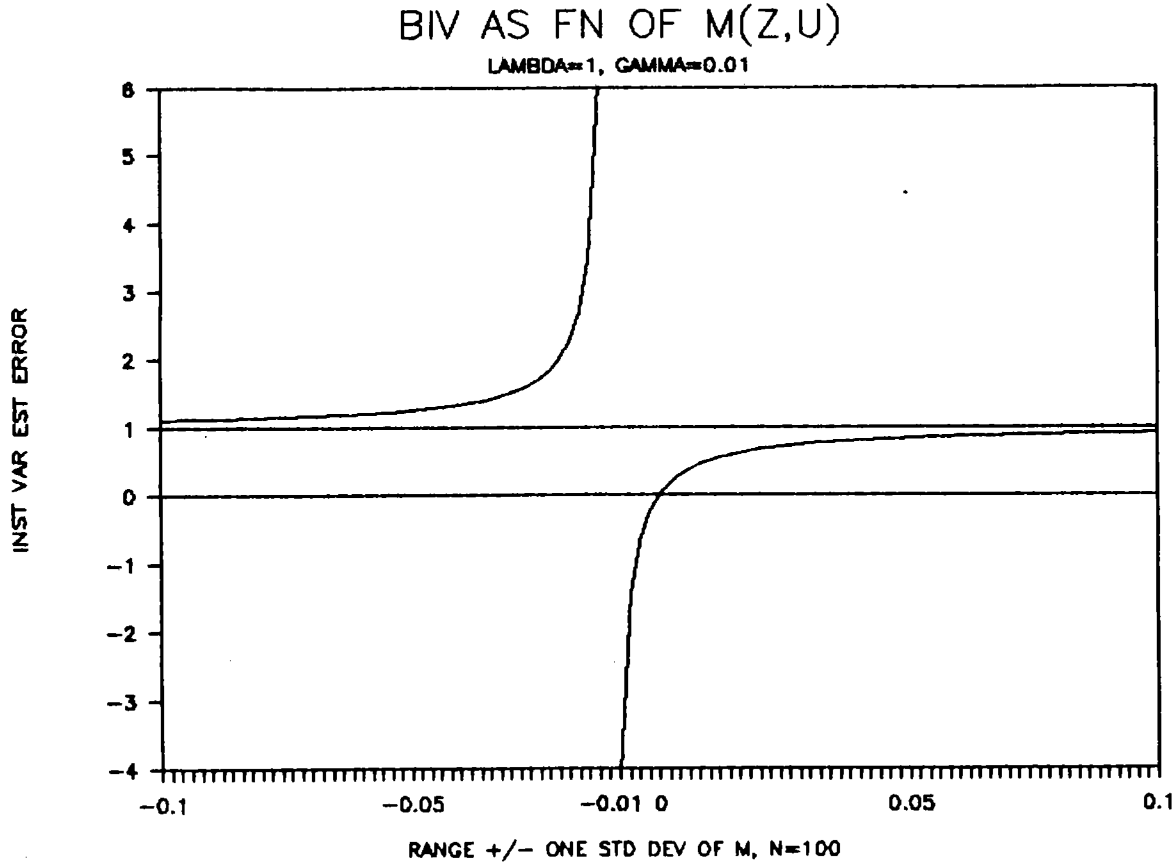

BIV ERROR AS FN OF M(Z,U)
LAMBDA = 1, GAMMA = .05
INST VAR EST ERROR
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-0.1
-0.05
0
0.05
0.1
COV BETWEEN INST Z AND U
$-\gamma\sigma_6^2/\lambda$
λ^{-1}

FIGURE 2


BIV AS FN OF M(Z,U)
LAMBDA=1, GAMMA=0.01
INST VAR EST ERROR
6
5
4
3
2
1
0
-1
-2
-3
-4
-0.1
-0.05
-0.01 0
0.05
0.1
RANGE +/- ONE STD DEV OF M, N=100