A REMARK ON BIMODALITY AND WEAK INSTRUMENTATION IN STRUCTURAL EQUATION ESTIMATION

BY

PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1171

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2006

http://cowles.econ.yale.edu/

Econometric Theory, **22**, 2006, 947–960. Printed in the United States of America.
DOI: 10.1017/S0266466606060439

A REMARK ON BIMODALITY AND WEAK INSTRUMENTATION IN STRUCTURAL EQUATION ESTIMATION

Peter C.B. Phillips
Cowles Foundation, Yale University
University of Auckland
and
University of York

In a simple model composed of a structural equation and identity, the finite-sample distribution of the instrumental variable/limited information maximum likelihood (IV/LIML) estimator is always bimodal, and this is most apparent when the concentration parameter is small. Weak instrumentation is the energy that feeds the secondary mode, and the coefficient in the structural identity provides a point of compression in the density that gives rise to it. The IV limit distribution can be normal, bimodal, or inverse normal depending on the behavior of the concentration parameter and the weakness of the instruments. The limit distribution of the ordinary least squares (OLS) estimator is normal in all cases and has a much faster rate of convergence under very weak instrumentation. The IV estimator is therefore more resistant to the attractive effect of the identity than OLS. Some of these limit results differ from conventional weak instrument asymptotics, including convergence to a constant in very weak instrument cases and limit distributions that are inverse normal.

1. INTRODUCTION

Some recent attention has been given to the fact that structural equation estimators may have bimodal finite-sample distributions. Phillips and Hajivassiliou (1987) explicitly mentioned the phenomenon, Nelson and Startz (1990) brought the property into prominence, and Maddala and Jeong (1992) provided some further analysis. There has since been a good deal of interest, recent contributions being Woglom (2001), Hillier (2006), Forchini (2006), and Kiviet and Niemczyk (2005). Forchini (2006) studies conditions for the finite-sample

My thanks to Richard Smith and two referees for comments on an earlier version. Section 2 of the paper is based on lectures given to students over the 1970s and 1980s at Essex, Birmingham, and Yale. Partial support is acknowledged from a Kelly Fellowship at the University of Auckland School of Business and the NSF under Grant SES 04-142254. Address correspondence to Peter Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA; e-mail: peter.phillips@yale.edu.

© 2006 Cambridge University Press 0266-4666/06 $12.00

distribution of the instrumental variable (IV) estimator to be bimodal, using the factorization of the standard expression of the density (Phillips, 1980) into a leading term and complementary term to find the parameter configurations where bimodality occurs.

A demonstration that the distributions of simultaneous equations estimators were not always unimodal appeared many years ago as a solved (and largely forgotten) exercise in Phillips and Wickens (1978, Sol. 6.19, pp. 351–355), which analyzed the finite-sample distribution of the limited information maximum likelihood/instrumental variable (LIML/IV) estimator in a simple Keynesian structural model. In fact, in this example the finite-sample distribution is *always* bimodal. The property holds for all parameter values and all sample sizes, although its magnitude is not always practically important. The present paper briefly revisits the Phillips and Wickens example and adds some further analysis and asymptotics to cover weak and very weak instrument cases, where some of the outcomes differ from the results in the conventional weak instrument literature. The model studied by Phillips and Wickens is, in fact, formally equivalent to the model that was considered much later by Nelson and Startz (1990), so there are some interesting linkages to the subsequent literature on the topic.

The Keynesian model is a case of strong endogeneity, where there is a structural behavioral equation and an identity. The identity is another structural relation, and its role is important in the distribution theory because it provides a magnet for an alternative centering, pulling consistent estimators like IV and LIML away from the relevant parameter in the behavioral relation and thereby naturally inducing a bimodality. In fact, it is the identity that is the source of the bimodality. This situation differs in some important ways from the standard case that has been studied intensively in the recent research. The simplicity of the example also means that the key properties of the distribution can be demonstrated analytically in a straightforward way without the use of special functions. For the conditions under which their analysis proceeds, the Nelson and Startz (1990) model is formally equivalent to a structural equation with a companion structural identity, as indicated earlier, and it is the identity that explains the bimodality noted in their paper. This corroborates the comments by Maddala and Jeong (1992) on the role played by strong endogeneity in the occurrence of bimodality.

The fact that identities are common in structural systems makes results for this simple model of more than passing theoretical interest. These results reveal elements in the earlier work on finite-sample theory that have implications on weak instrumentation and weak IV limit theory.

2. STRUCTURAL ESTIMATION WITH AN IDENTITY

The model considered here is based on the simplest Keynesian model with a single structural equation involving two measured endogeneous variables y_t,

x_t, a stochastic disturbance u_t, and a parameterized structural identity involving an observed IV z_t that is assumed to be exogenous. The system is

$$y_t = \beta x_t + u_t, \tag{1}$$

$$x_t = y_t + \gamma z_t, \tag{2}$$

where the (spending propensity) parameter β is assumed to satisfy $\beta \neq 1$, so that an equilibrium solution exists. Of course, this condition is needed for the existence of a reduced form and a proper data generating mechanism for the sample data $\{y_t, x_t : t = 1, \ldots, n\}$. The parameter γ controls the relevance of the instrument z_t in the system and is convenient to use as a scale coefficient in this equation, but it could readily be absorbed into z_t and its effects measured in terms of the signal from the instrument. When $\gamma \to 0$, the instrument z_t becomes irrelevant to the determination of y_t and x_t, and we end up with the identity $x_t = y_t$ in place of (2). On the other hand, when $\gamma \to \infty$, the system is dominated by the signal from z_t. In view of the identity (2) and the exogeneity of z_t, the degree of endogeneity as measured by the correlation coefficient of x_t and u_t is unity, so that there is strong endogeneity in the system.

Sometimes it is convenient to extend the model by an array formulation and index one or the other of the parameters γ and β by the sample size n. Use of an indexed sequence γ_n for γ opens up the study of weak instrument cases, where γ_n may be passed to zero at certain rates, and use of β_n for β enables the model to be analyzed for spending propensities in the vicinity of unity, where $\beta_n \to 1$.

For a finite-sample development, the errors $\{u_t : t = 1, \ldots, n\}$ in (1) may be assumed to be *iid* $N(0, \sigma^2)$, although Gaussianity is unnecessary for the asymptotics. Phillips and Wickens, like Bergstrom (1962), who first considered finite-sample distributions in this system, allow for an intercept in (1), which is inconsequential, and they did not parameterize (2), setting $\gamma = 1$.

Define $s_{zz} = \sum_{t=1}^{n} z_t^2$. The reduced form is

$$y_t = \pi_y z_t + \frac{1}{1-\beta} u_t, \qquad \pi_y = \frac{\beta\gamma}{1-\beta}, \tag{3}$$

$$x_t = \pi_x z_t + \frac{1}{1-\beta} u_t, \qquad \pi_x = \frac{\gamma}{1-\beta}, \tag{4}$$

and β is identified by

$$\beta = \pi_y / \pi_x = 1 - \gamma / \pi_x. \tag{5}$$

The IV or LIML estimator of β is $\hat{\beta} = \hat{\pi}_y / \hat{\pi}_x$, where $\hat{\pi}_y$ and $\hat{\pi}_x$ are the reduced form least squares estimates. Analogous to Bergstrom (1962), Phillips and Wickens (1978) gave the exact density of $\hat{\beta}$:

$$\text{pdf}(b) = \frac{\lambda_n^{1/2}}{\sqrt{2\pi}} \frac{|1-\beta|}{\sigma} \frac{1}{(1-b)^2} \exp\left\{ -\frac{\lambda_n}{2\sigma^2} \left(\frac{b-\beta}{1-b} \right)^2 \right\}, \tag{6}$$

where $\lambda_n = \gamma^2 s_{zz}$ is the noncentrality parameter. Note that λ_n depends on the sample size through the sample moment s_{zz} but also through the parameter γ when $\gamma = \gamma_n$ depends on n. The exact density (6) is the same as that studied in Nelson and Startz (1990), after notational translation. This is because the model studied by Nelson and Startz (1990) is observationally equivalent to a structural equation with a parameterized identity under the conditioning and zero covariance assumptions that are made in that paper.[1]

As shown in Phillips and Wickens (1978), the density pdf(b) is continuous and has a zero at $b = 1$, the same result later being given in Lemma 2 of Nelson and Startz (1990). Rather obviously, the tails are $O(b^{-2})$, or Cauchy-like (as pointed out in Sargan, 1970/1988, and Phillips, 1983, 1984, 1985, 1986, for structural full information maximum likelihood [FIML] and LIML estimators), and the distribution therefore has modes on either side of the zero at $b = 1$. Simple calculations reveal that there are two modes located at

$$1 + \frac{\lambda_n}{4\sigma^2}(1-\beta) \pm \frac{(1-\beta)\lambda_n^{1/2}}{\sqrt{2}\sigma} \left\{ 1 + \frac{\lambda_n}{8\sigma^2} \right\}^{1/2}$$

$$\sim \begin{cases} 1 + \dfrac{\lambda_n^{1/2}}{\sqrt{2}\sigma}(1-\beta) + \dfrac{\lambda_n(1-\beta)}{4\sigma^2} + O(\lambda_n^{3/2}) & \text{as } \lambda_n \to 0 \\ 1 - \dfrac{\lambda_n^{1/2}}{\sqrt{2}\sigma}(1-\beta) + \dfrac{\lambda_n(1-\beta)}{4\sigma^2} + O(\lambda_n^{3/2}) & \text{as } \lambda_n \to 0 \end{cases}$$

$$\sim \begin{cases} 2 - \beta + \dfrac{\lambda_n}{2\sigma^2}(1-\beta) - \dfrac{2\sigma^2(1-\beta)}{\lambda_n} + O\left(\dfrac{1}{\lambda_n^2}\right) & \text{as } \lambda_n \to \infty \\ \beta + \dfrac{2\sigma^2(1-\beta)}{\lambda_n} + O\left(\dfrac{1}{\lambda_n^2}\right) & \text{as } \lambda_n \to \infty, \end{cases}$$

where the last two expressions are asymptotic expansions of the modal locations as $\lambda_n \to 0$ and $\lambda_n \to \infty$, respectively. For small λ_n, the modes are placed nearly symmetrically on either side of unity at $1 \pm (\lambda_n^{1/2}/\sqrt{2}\sigma)(1-\beta)$, and at these points the density is of $O(\lambda_n^{-1})$ as $\lambda_n \to 0$. For large λ_n, one mode is located near β around the value $\beta + 2\sigma^2(1-\beta)/\lambda_n$, and the second mode is located around $2 - \beta + (\lambda_n/2\sigma^2)(1-\beta)$. At this second mode, the density is $O(\lambda_n^{-2})$ as $\lambda_n \to \infty$, and so it is negligible in magnitude for large λ_n.

These properties hold for $\beta \neq 1$ and for all sample sizes n, or all values of the noncentrality parameter λ_n provided $\gamma \neq 0$. The bimodality disappears asymptotically when $\lambda_n \to \infty$ (as happens when γ is fixed and nonzero and

$n \to \infty$), because then the distribution of $\hat{\beta}$ is asymptotically Gaussian. In this event we have the limit theory

$$\sqrt{\lambda_n}(\hat{\beta} - \beta) \Rightarrow N(0, \sigma^2(1-\beta)^2),$$

which we can write in standardized form as

$$\frac{\sqrt{\lambda_n}(\hat{\beta} - \beta)}{1-\beta} \Rightarrow N(0, \sigma^2),$$

which covers the case where $\beta = \beta_n \to 1$. In the latter case, it is apparent that the presence of a spending propensity in the vicinity of unity raises the convergence rate above $\sqrt{\lambda_n}$, which is explained by the fact that both the structural equation and the identity work to attract the estimator to unity. Nelson and Startz (1990) argue that the usual Gaussian limit theory is often a very poor approximation to the finite-sample distribution and that it is particularly bad when the instrument is weak. This follows earlier arguments made in Phillips (1989) for the case of irrelevant instruments. However, as we see later, there is an alternative limiting inverse normal theory in this case that provides a very satisfactory approximation to the finite-sample distribution, including its bimodality.

Some typical shapes of the finite-sample density (6) are shown in Figures 1 and 2. For $\lambda_n = 50$, the distribution is close to symmetric, is centered on β, and

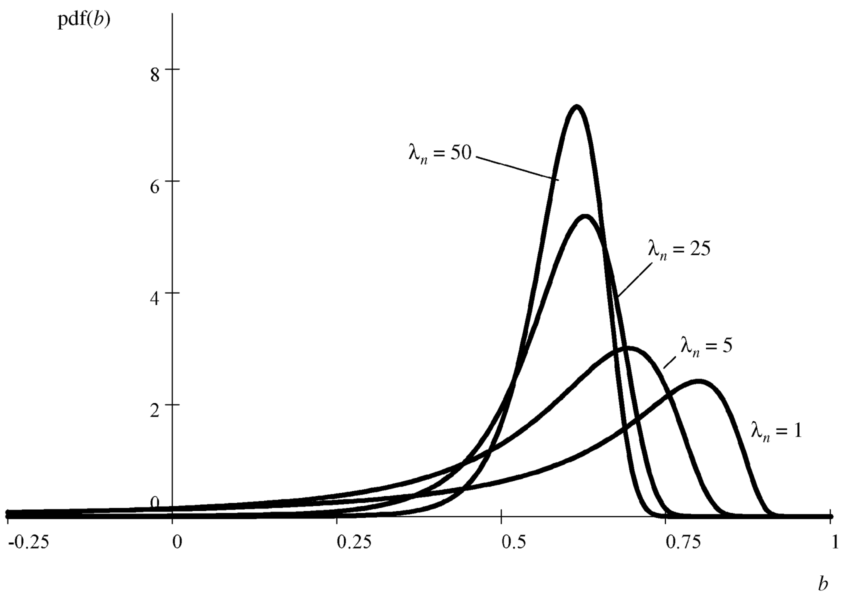

FIGURE 1. Finite-sample distributions of $\hat{\beta}$ for $\beta = 0.6$ and $\lambda_n = 1, 5, 25, 50$.

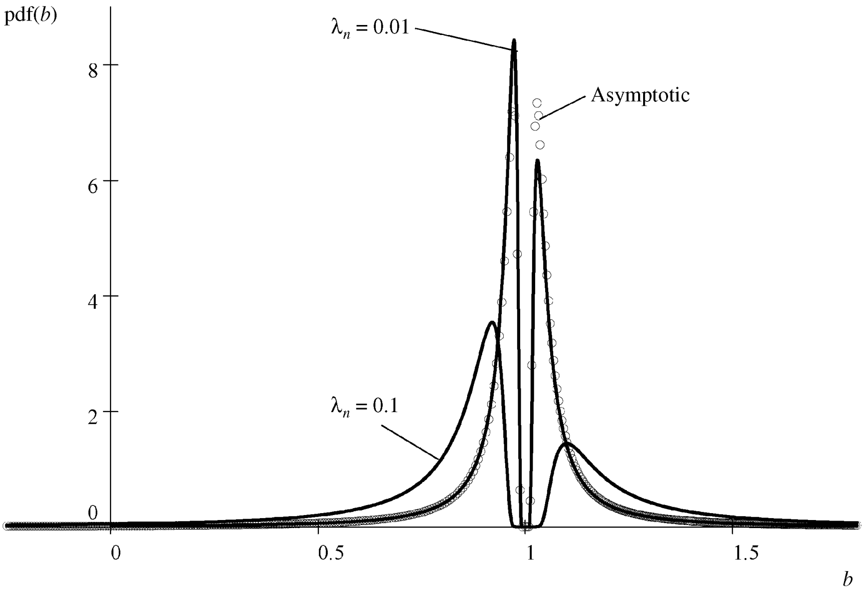

FIGURE 2. Finite-sample distributions of $\hat{\beta}$ for $\beta = 0.6$ and $\lambda_n = 0.1$, 0.01, and the (very weak) instrument asymptotic distribution (circles).

is well approximated by the Gaussian limit. For $\lambda_n = 5,1$ the distributions are decidedly non-Gaussian in form and show substantial bias and asymmetry, with the mode shifting away from β toward unity as λ_n decreases. In Figure 1 the mode on the right side of unity is so small that it cannot be seen on this scale, so the right tail of the distribution is omitted from view.

Figure 2 shows the densities when $\lambda_n = 0.1, 0.01$. Both have marked bimodality. As γ and λ_n decrease, the pull of the structural identity becomes stronger, and the primary mode shifts closer to unity. As γ and λ_n decrease further, the secondary mode on the right side of unity becomes more accentuated, and the primary mode shifts even closer to unity. The following intuition explains the bimodality and the location of the modes. Because the IV density is held to zero at $b = 1$, the secondary mode occurs to the right of unity, much in the same way as a balloon (representing density) is squeezed and a new bubble (mode) arises outside the point of compression to accommodate the density that has been squeezed out elsewhere. Here, the identity serves to provide the point of compression (delivered by the coefficient of the structural identity), and the weak instrumentation provides the energy of compression. Interestingly, the distribution becomes more symmetric again as γ and λ_n decrease (just as it is for large λ_n), but now about the value $b = 1$. Also, although the modes are more peaked and move closer to unity from the right and left sides, the density still descends to zero at unity for all n.

As is clear from the functional form (6), the density pdf(b) does not tend to a proper probability density as $\lambda_n \to 0$. In fact, pdf$(b) \to 0$ for all values of b as $\lambda_n \to 0$, but not uniformly. So the probability mass escapes at infinity in this case as we take the limit of the density. Nonetheless, we still get convergence of $\hat{\beta}$ in this case, but to unity, not to β. In fact, $\hat{\beta} \to_p 1$ as $\lambda_n \to 0$. Correspondingly, the modes in the density become more peaked and move closer together.

In the limit where $\gamma = 0$, we have $\pi_y = \pi_x = 0$, and the reduced form coefficients contain no information[2] about the parameter β. In this case, the instrument z_t is irrelevant for x_t in estimating β in the structural equation (1). Nonetheless, when $\gamma = 0$, the identity (2) becomes $x_t = y_t$, and it is apparent that in this event $\hat{\beta} = 1$, *a.s.* for all n. The IV estimator also takes the same value as OLS in this case. This result is different from the distribution in the irrelevant instrument case considered in Phillips (1989), where the IV estimator differs from OLS, is random for all n, and converges to a random variable as $n \to \infty$, unlike OLS, which has a finite nonrandom probability limit.

The result also differs from the limiting case studied in Forchini (2006). Forchini's model is a standard two variable simultaneous system[3] without an identity and with its degree of endogeneity measured by the correlation coefficient $\rho = -\beta^*/(1 + \beta^{*2})^{1/2}$, where β^* is the standardized form (see Phillips, 1983) of the structural coefficient β to be estimated by IV. Forchini takes the limit of the finite-sample density of the IV estimator of β^* as the concentration parameter $\lambda_n \to 0$ along the parameterized path $\lambda_n = \theta(1 - \rho^2) + o(1 - \rho^2) = \theta(1 + \beta^{*2})^{-1} + o(\beta^{*-2})$ for some $\theta \geq 0$. In this case, the limit of the finite-sample density is a proper probability density, which in the just-identified case has the following form (Forchini, 2006, Thm. 2):

$$\text{pdf}(b) = \frac{e^{-\theta/2}}{\pi(1+b^2)}\, {}_1F_1\left(1, \frac{1}{2}; \frac{\theta}{2}\frac{b^2}{1+b^2}\right), \tag{7}$$

where ${}_1F_1$ denotes the confluent hypergeometric function. Forchini shows this limit distribution to be bimodal when $\theta > 1$.

When standardizing transformations are employed so that $n^{-1}Z'Z = I_k$, the parameter θ has the form

$$\theta = \lambda_n(1 + \beta^{*2}) = \frac{n}{\sigma_v^2}\gamma'\gamma(1 + \beta^{*2}) = \frac{n\gamma'\gamma}{\sigma_v^2(1 - \rho^2)}$$

and may therefore be interpreted as the *total* concentration coefficient because it represents the total amount of "energy" in the reduced form coefficients (γ and $\beta^*\gamma$). Note that $\lambda_n \to 0$ requires $\gamma'\gamma \to 0$ and therefore weak instrumentation. The limiting density (7) includes the totally unidentified case where there is no energy ($\theta = 0$) as in Phillips (1989) and cases where there is some energy ($\theta > 0$) in the coefficients.

As θ increases, we might expect the density to be somewhat more revealing about β^*, even when there is weak instrumentation ($\gamma'\gamma$ is small), and this is precisely what happens. As θ becomes large, we can use the following simple approximation:

$$\mathrm{pdf}(b) = \frac{\theta^{1/2}|b|e^{-(\theta/2)(1/(1+b^2))}}{\sqrt{2\pi}(1+b^2)^{3/2}}\{1+O(\theta^{-1})\}, \tag{8}$$

based on the asymptotic expansion of the confluent hypergeometric function ${}_1F_1$ for large θ (e.g., Lebedev, 1972, p. 271). By a straightforward calculation, the modes of (8) are located at

$$b = \pm\left(\frac{\theta-1}{2}\right)^{1/2} + O\left(\frac{1}{\theta}\right).$$

This approximation is sufficiently accurate, in fact, to capture correctly the criterion for bimodality, namely, $\theta > 1$, given in Forchini (2006).

Figure 3 shows the density (8) for various values of θ. Interestingly, as θ increases, the modes in the density (8) move further away from the origin (the point of attraction provided by the least squares estimate and the mode of the IV estimator in the zero energy case), and they diverge with θ. Thus, the modes are drawn toward the two possible limiting values of the standardized struc-

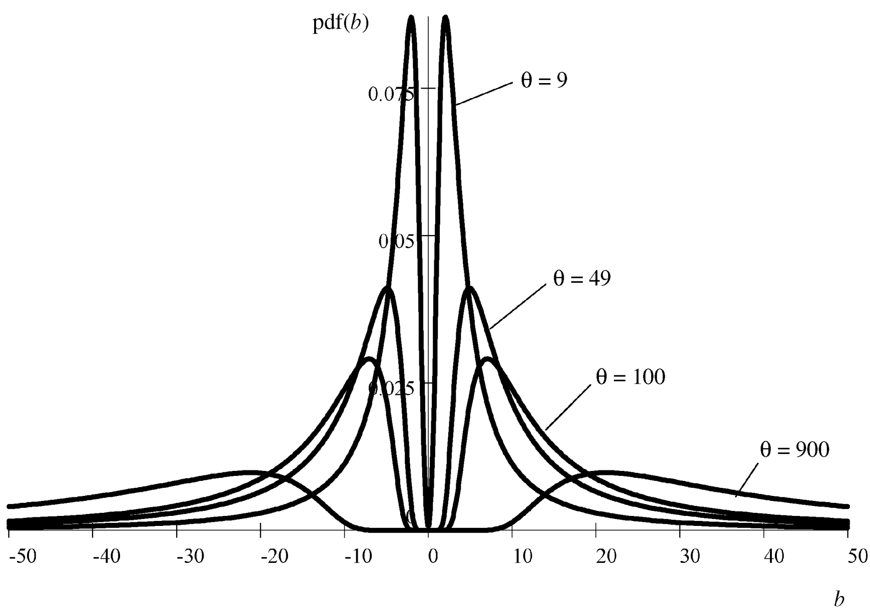

FIGURE 3. Finite-sample densities (8) of IV estimator for θ = 9, 49, 100, 900.

tural coefficient, namely, $\beta^* = \pm\infty$, which serve as dual attractors based on the point at infinity $\beta^{*2} = \infty$, for which we have $\rho^2 = 1$. Correspondingly, the density becomes less peaked and the modes are less well defined as θ increases. Indeed, mass escapes from the density and $\mathrm{pdf}(b) \to 0$ for all b as $\theta \to \infty$. The situation here is precisely the opposite of what happens in the case of the density (6), where the modes become more accentuated and move toward the value $b = 1$, a single attractor provided by the coefficient in the structural identity. In that case, weak instrumentation ($\gamma \to 0$) is the force behind the movement in the modes and the concentration of the distribution. In the case of the density (8), however, the modes move apart and are drawn to points at infinity corresponding to $\beta^* = \pm\infty$, mass escapes to infinity, and the IV estimator of β^* diverges. In this event, the force behind the movement in the modes and the divergence of the IV estimator is the energy ($\theta \to \infty$) in the reduced form coefficients. Thus, even in the presence of weak instrumentation where $\gamma'\gamma \to 0$, the degree of endogeneity increases sufficiently fast (i.e., $\rho^2 \to 1$) that the total energy in the reduced form coefficients diverges ($\theta \to \infty$), thereby revealing information about β^*.

3. SOME LIMIT THEORY FOR WEAK INSTRUMENTS

Next consider the case where γ_n depends on n and $\gamma_n \to 0$ as $n \to \infty$. The simplest situation of this kind occurs when a local $\sqrt{n}$ to zero sequence $\gamma_n = d/\sqrt{n}$ is used, an approach that has become conventional in the study of weak instrumentation. Some recent reviews of this literature are given in Andrews and Stock (2005), Hahn and Hausman (2003), and Stock, Wright, and Yogo (2002). Cases of many weak instruments are studied in Chao and Swanson (2005) and Han and Phillips (2006).

The particular $\sqrt{n}$ local to zero sequence has no special significance, and limit results may be obtained for other cases. However, we may usefully start with this case, because of its popularity in the literature, and compare results with what happens when $\gamma_n \to 0$ at a faster rate. Accordingly, set $\gamma_n = d/\sqrt{n}$, require $d \neq 0$, and suppose that the sample second moment of the instrument converges to a positive constant, so that $n^{-1}s_{zz} \to m_{zz} > 0$. In this event, the noncentrality parameter tends to a positive constant

$$\lambda_n = \gamma_n^2 s_{zz} \to d^2 m_{zz} = \lambda > 0, \tag{9}$$

and it is apparent that the density (6) has the following limit as $n \to \infty$:

$$\mathrm{pdf}(b) = \frac{\lambda^{1/2}}{\sqrt{2\pi}} \frac{|1-\beta|}{\sigma} \frac{1}{(1-b)^2} \exp\left\{-\frac{\lambda}{2\sigma^2}\left(\frac{b-\beta}{1-b}\right)^2\right\}. \tag{10}$$

The limit theory in this weak instrument case, analogous to Staiger and Stock (1997), simply reproduces the finite-sample distribution under Gaussianity, a phenomenon that reflects the fact that as n becomes large, the data (and the instrument in particular) become less informative about the true value of the parameter of interest, β. As originally shown in Phillips (1989), it is a straightforward matter to convert the limit result (10) into an invariance principle. All that is required is an appeal to a central limit theorem for the reduced form coefficients on which $\hat{\beta}$ depends. In the present case, we have the singular normal limit

$$\sqrt{n}\begin{bmatrix} \hat{\pi}_y - \pi_y \\ \hat{\pi}_x - \pi_x \end{bmatrix} \Rightarrow N\left(0, \frac{\sigma^2}{(1-\beta)^2 m_{zz}}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right), \tag{11}$$

and the limit distribution (10) follows in a straightforward manner that is analogous to the finite-sample development. Observe that the limit distribution (10), like the finite-sample distribution, has asymmetric bimodality and Cauchy-like tails.

In place of the conventional weak instrument condition (9), we may consider the very weak instrument case where $\gamma_n = o(n^{-1/2})$ and

$$\lambda_n = \gamma_n^2 s_{zz} \to 0. \tag{12}$$

In this case, observe that $\sqrt{n}\pi_x = \sqrt{n}\gamma_n/(1-\beta) = o(1)$, so that the systematic part of the reduced form is sufficiently small that the limit theory (11) can be replaced by

$$\sqrt{n}\begin{bmatrix} \hat{\pi}_y \\ \hat{\pi}_x \end{bmatrix} \Rightarrow N\left(0, \frac{\sigma^2}{(1-\beta)^2 m_{zz}}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\eta, \tag{13}$$

where $\eta = (\sigma/(|1-\beta| m_{zz}^{1/2}))\xi$ and ξ is standard normal. We deduce that

$$\hat{\beta} = 1 - \frac{\gamma_n}{\hat{\pi}_x} = 1 - \frac{\sqrt{n}\gamma_n}{\sqrt{n}\hat{\pi}_x} \to_p 1, \tag{14}$$

so that $\hat{\beta}$ is inconsistent and converges to unity in this case. Some further elementary manipulations[4] reveal that the limit distribution is given by

$$\delta_n(\hat{\beta} - 1) \Rightarrow -\frac{1}{\eta}, \tag{15}$$

the inverse of a central Gaussian variate, and the rate of convergence is given by $\delta_n = 1/\sqrt{n}\gamma_n$. In this very weak instrument case, the systematic part of the reduced form is so small that we do not accumulate information fast enough from the observations (the rate here is still the conventional $\sqrt{n}$ rate in view of the behavior of s_{zz}) to learn enough about π_x and π_y to distinguish these parameters from zero. In effect, what we learn from large samples of data is that $\hat{\pi}_x$ and $\hat{\pi}_y$ are distributed about the origin. Relative to the rate at which $\gamma_n \to 0$, we then deduce from the relationship between $\hat{\beta}$ and $\hat{\pi}_x$ in (14) that the estimate $\hat{\beta}$ is centered on unity and behaves, after scaling, like the inverse normal variate $1/\sqrt{n}\hat{\pi}_x$.

The asymptotic distribution in the very weak instrument case is shown in Figure 2. As is apparent, the asymptotic distribution provides a good approximation to the finite-sample distribution. Unlike the limit distributions given in (10) and in other weak instrument cases considered in the recent literature, this asymptotic distribution is symmetrically bimodal. It also has zero density at the probability limit of $\hat{\beta}$ and continues, of course, to have Cauchy-like tails.

At the point in the parameter space where $\gamma = 0$, the data generating mechanism becomes simply $y_t = x_t = u_t/(1-\beta)$. This mechanism incorporates the identity (2) and the reduced form equations (3) and (4) with $\gamma = 0$. It applies whatever the value of β, and the structural equation (1) is essentially irrelevant to the data generating mechanism. Similarly, when $\gamma_n \to 0$ fast enough, the same effects occur. The asymptotic distribution of $\hat{\beta}$ is centered about unity (the coefficient in the structural identity $y_t = x_t$), the distribution converges at the rate $1/\sqrt{n}\gamma_n$, which reflects the manner in which the concentration in the data about the structural identity occurs, and the limit distribution provides no information about β when σ^2 is unknown.

On the other hand, as remarked earlier, the value of β does affect the reduced form error variance and thereby affects the limit distribution of $\hat{\beta}$ about unity. Because this error variance $\sigma_v^2 = \sigma^2/(1-\beta)^2 \to \infty$ as $\beta \to 1$, it follows that $1/\eta \to_p 0$, as $\beta \to 1$. We therefore may write (15) in the standardized form

$$\frac{\delta_n \sigma}{m_{zz}^{1/2}(1-\beta)}(\hat{\beta}-1) \Rightarrow \frac{1}{\xi},$$

where ξ is standard normal. This result covers the case of a localizing sequence β_n for which $\beta_n \to 1$ and then

$$\frac{\sqrt{n}\sigma}{\sqrt{\lambda_n}(1-\beta_n)}(\hat{\beta}-1) \Rightarrow \frac{1}{\xi}, \tag{16}$$

where the rate of convergence is correspondingly accelerated.

These limit results may be compared with those for the least squares estimator β^*, for which standard calculations lead to the following limit theory for the $\sqrt{n}$ local to zero sequence $\gamma_n = d/\sqrt{n}$:

$$\frac{\sqrt{n}}{\gamma_n}(\beta^* - 1) \Rightarrow N\left(\frac{d(1-\beta)m_{zz}}{\sigma^2}, \frac{(1-\beta)^2 m_{zz}}{\sigma^2}\right) \tag{17}$$

and for the very weak instrument case where $\gamma_n = o(n^{-1/2})$:

$$\frac{\sqrt{n}}{\gamma_n}(\beta^* - 1) \Rightarrow N\left(0, \frac{(1-\beta)^2 m_{zz}}{\sigma^2}\right). \tag{18}$$

In both cases $\beta^* \to_p 1$ and the limit distribution is normal. There is a noncentrality in the limit distribution for the $\sqrt{n}$ local to zero sequence, and the convergence rate is $O(n)$ in this case. In the very weak instrument case, the limit distribution is central normal and the convergence rate is $O(n\delta_n)$, which is faster than $O(n)$. Again, these results extend to the case of a localizing sequence β_n for which $\beta_n \to 1$, and then (18), for example, becomes

$$\frac{n\delta_n}{(1-\beta_n)}(\beta^* - 1) \Rightarrow N\left(0, \frac{m_{zz}}{\sigma^2}\right).$$

It follows from these results that, although the IV and OLS estimators have the same limit in probability in the very weak instrument case, their limit distributions and rates of convergence are quite different, so the estimators are not asymptotically equivalent, in contrast to the case where the degree of apparent overidentification is large enough to make the estimators equivalent. In the present case, both IV and OLS estimators are attracted to the same limit point but in a very different manner and at different rates.

In weak identification situations, it is sometimes argued that the IV estimator is drawn toward the least squares estimator.[5] This can be a misleading representation of the phenomenon, as the present example shows. Rather than IV being drawn to OLS, it is the coefficient in the structural identity in the model (or, equivalently, the regression coefficient implied by the presence of strong endogeneity) that acts as an alternative point of attraction in estimation for both estimators. IV is drawn to this point of attraction from both the left and the right because of its bimodal distribution, in contrast to OLS. In situations of weak identification, both estimators tend to be attracted to this point in the parameter space, and the magnetic force is stronger the weaker the instruments. But, as the preceding results show, the attraction is far greater for OLS than it is for IV, so that the former estimator has a faster rate of convergence and the latter estimator is much more resistant to the attractor. Indeed, for $\sqrt{n}$ local to zero sequences like $\gamma_n = d/\sqrt{n}$, OLS converges at an $O(n)$ rate to the attractor, whereas IV converges to a random variable with the bimodal distribution (10), so it puts probability mass away from and on both sides of the attractor. On the other hand, when the instruments are very weak, we have $\delta_n(\hat{\beta} - \beta^*) \Rightarrow -(1/\eta)$, so that the IV estimator has a symmetric bimodal limiting distribution about the least squares value and IV is "drawn to OLS" but in an ambivalent way. Clearly, this is very different from the two estimators being the same in the limit or IV simply being biased toward the least squares value.

4. FURTHER DISCUSSION

In both strong and very weak instrument cases with a structural identity the IV/LIML estimator converges in probability to a constant. In one case, the constant is the true value of the structural parameter β. In the other, it is the coefficient in the structural identity. In conventional weak instrument and irrelevant instrument cases, the estimator converges weakly to a random quantity whose distribution reflects the uncertainty associated with a finite sample of data, as in Phillips (1989). So the presence of a structural identity can lead to substantial changes in weak instrument limit theory.

The finite-sample distribution (and the implied limit distribution under conventional weak instrumentation) is bimodal for all values of the parameters. The source of the bimodality is the presence of an alternative point of attraction provided by the coefficient in the structural identity in the system. The attraction process applies also to OLS but is unidirectional and stronger in the case of OLS and bidirectional and weaker for IV.

The bimodality is of a magnitude to be practically important in cases where the concentration parameter λ_n is small. As λ_n becomes very small, the distribution becomes strongly bimodal about the coefficient in the identity, and if $\lambda_n \to 0$ the estimator converges in probability to that coefficient at an $O(\lambda_n^{-1/2})$ rate and has a limit distribution that is proportional to the inverse of a standard normal variate and is symmetrically bimodal. By contrast, the OLS estimator converges at an $O(n\lambda_n^{-1/2})$ rate and has a limiting normal distribution.

These results show that, even for the exceedingly simple structural system considered here, weak instrument limit theory has a richer range of possible outcomes than are contained in the present literature.

NOTES

1. Maddala and Jeong (1992) observed that the Nelson and Startz (1990) model is formally equivalent under their stated conditions to the model

$$y_t = \beta x_t + u_t,$$

$$x_t = \theta(z_t - v_t) + \phi u_t$$

for certain parameters θ and ϕ and where the analysis is conditioned on the supplementary variable v_t. The second equation may be rewritten as the parametrized identity

$$x_t = \frac{\phi}{1+\beta\phi} y_t + \frac{\theta}{1+\beta\phi}(z_t - v_t),$$

so that the Nelson and Startz (1990) model is equivalent to (1) and (2) after rescaling x_t and β and conditioning on v_t.

2. The reduced form error variance $\sigma_v^2 = \sigma^2/(1-\beta)^2$ also contains no recoverable information about β, as σ^2 is unknown.

3. The model is

$$y_t = \beta x_t + u_t,$$

$$x_t = \gamma' z_t + v_t,$$

where (u_t, v_t) have covariance matrix $\Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$ and the concentration parameter is $\lambda_n = \gamma' Z' Z\gamma/\sigma_v^2$.

4. $\hat{\beta} - 1 = -\dfrac{\sqrt{n}\gamma_n}{\sqrt{n}\hat{\pi}_x} = -\dfrac{\sqrt{n}\gamma_n}{\dfrac{\sqrt{n}\gamma_n}{1-\beta} + \eta_n} = -\dfrac{\sqrt{n}\gamma_n}{\eta_n} + o_p(1), \quad \text{where } \eta_n \Rightarrow \eta.$

5. For example, one of the referees refers to "the result that under weak instruments and strong endogeneity the 2SLS estimator is biased towards the least squares value, as discussed by Staiger and Stock (1997)."

REFERENCES

Andrews, D.W.K. & J.H. Stock (2005) Inference with Weak Instruments. Cowles Foundation Discussion paper 1530, Yale University.

Bergstrom, A.R. (1962) The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume. *Econometrica* 30, 480–490.

Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1712.

Forchini, G. (2006) On the bimodality of the exact distribution of the TSLS estimator. *Econometric Theory* 22, 932–946 (this issue).

Hahn, J. & J. Hausman (2003) Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review* 93, 118–125.

Han, C. & P.C.B. Phillips (2006) GMM with many moment conditions. *Econometrica* 74, 147–192.

Hillier, G.H. (2006) Yet more on the exact properties of IV estimators. *Econometric Theory* 22, 913–931 (this issue).

Kiviet, J.F. & J. Niemczyk (2005) The Asymptotic and Finite Sample Distributions of OLS and IV in Simultaneous Equations. UVA Econometrics Discussion paper 2005/01.

Lebedev, N.N. (1972). *Special Functions and Their Applications*. Dover.

Maddala, G.S. & J. Jeong (1992) On the exact small sample distribution of the instrumental variable estimator. *Econometrica* 60, 181–183.

Nelson, C.R. & R. Startz (1990) Some further results on the small sample properties of the instrumental variable estimator. *Econometrica* 58, 967–976.

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

Phillips, P.C.B. (1983) Exact small sample theory in the simultaneous equations model. In M.D. Intriligator & Z. Griliches (eds.), *Handbook of Econometrics*, 449–516. North-Holland.

Phillips, P.C.B. (1984) The exact distribution of LIML, part I. *International Economic Review* 25, 249–261.

Phillips, P.C.B. (1985) The exact distribution of LIML, part II. *International Economic Review* 26, 21–36.

Phillips, P.C.B. (1986) The distribution of FIML in the leading case. *International Economic Review* 27, 239–243.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Phillips, P.C.B. & V.A. Hajivassiliou (1987) Bimodal *t* Ratios. Cowles Foundation Discussion paper 842.

Phillips, P.C.B. & M.R. Wickens (1978) *Exercises in Econometrics*, vol. 2. Ballinger and Philip Allan.

Sargan, J.D. (1970/1988) The Finite Sample Distribution of FIML Estimators. Paper presented at the Second World Congress of the Econometric Society. Reprinted in E. Maasoumi (ed.), *Contributions to Econometrics*, vol. 2, pp. 57–75 (Cambridge, 1988).

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

Stock, J.H., J. Wright, & M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20, 518–529.

Woglom, G. (2001) More results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 69, 1381–1389.