On the Bimodality of the Exact Distribution of the TSLS Estimator

G. Forchini[1 2]

Monash University

March 7, 2006

Abstract

We investigate the possible bimodality of the density of the TSLS estimator in a just/over-identified linear structural equation. By studying the interaction between weakness of instruments, degree of endogeneity and degree of over-identification we are able to identify conditions for its existence.

1. Introduction

Although the exact density of the two stage least squares (TSLS) estimator has been known for a few decades (e.g. the review by Phillips (1983)) some of its properties are still surprising for econometricians. Bimodality is one of these *unexpected* properties: Phillips and Wickens (1978) Solution 6.19, pp. 351-355, Nelson and Startz (1990), Maddala and Jeong (1992) and Woglom (2001) have shown that the density of the TSLS estimator may be bimodal in a just-identified structural equation.

This note investigates existence of bimodality in the density of the TSLS estimator in the context of a just/over-identified structural equation. First, we look at the claim that the exact finite sample distribution of the TSLS estimator cannot be easily interpreted (e.g. Woglom (2001) p. 1381). We show that the possible bimodality of the density of the TSLS estimator can be easily understood, and generalized to over-identified models, using the exact results reviewed by Phillips (1983). In fact, it is the outcome of the interaction of two components of the exact

---

[1] Address for correspondence: Giovanni Forchini, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia.
E-mail: Giovanni.Forchini@BusEco.monash.edu.au

[2] I thank Grant Hillier, Patrick Marsh, Don Poskitt, the Editor Peter Phillips and two anonymous referees for useful and encouraging comments.

density: one has only one mode and is symmetric around zero; the other has the shape of a pulse wave and depends on both the strength of the instruments and the degree of endogeneity. We prove a necessary condition for bimodality of the density of the TSLS estimator, and show that the bimodality does not occur if the degree of over-identification is large enough.

Next, we analyse the interaction between lack of identification and high degree of endogeneity. We argue that there are an infinite number of possible densities for the TSLS estimator when the model is unidentified, depending on the path along which the quality of the instruments goes to zero. Since these limit densities do not depend on the sample size, they can all be asymptotic distributions for the TSLS estimator when the concentration parameter is local to zero as, for instance, in the weak instruments literature.

Parts of our analysis are closely related to recent work of Hillier (2006) and Phillips (2006). Hillier (2006) studies the exact properties of the TSLS estimator by re-deriving the exact density using simple conditioning arguments. Hillier (2006) links bimodality to the normalization used in specifying the interest parameter in the structural equation (see also Hillier (1990)). In this note, we are only concerned with bimodality, and thoroughly study the interaction among noncentrality parameter, degree of endogeneity and degree of over-identification to determine precise conditions for its existence. Phillips (2006) also studies the density of the TSLS estimator as the noncentrality parameter tends to zero in the just-identified model with a structural identity considered by Phillips and Wickens (1978), Solution 6.19, pp. 351-355. The degree of endogeneity for his model is one, whereas, in this note, we study the interaction between noncentrality parameter tending to zero and degree of endogeneity tending to one. Phillips (2006) develops some asymptotic theory while we consider a fixed sample size only.

The rest of the paper is organized as follows. Section 2 presents the model. The properties of the exact density affecting bimodality of the TSLS estimator are considered in Section 3. Section 4 derives the limit densities as the correlation between right-hand-side endogenous variables and the instruments tends to zero. Section 5 concludes. Proofs are in the appendix.

2. The two-endogenous variables model

Consider the simple instrumental variables model:

(1) $$y_t = x_t \beta + u_t$$

(2) $$x_t = z_t ' \gamma + v_t ,$$

$t = 1, 2, ..., T$, where y_t and x_t are endogenous variables, z_t is a $(k \times 1)$ vector of exogenous variables, β and γ are unknown parameters of dimension (1×1) and $(k \times 1)$, respectively, and u_t and v_t are random errors. We assume that (u_t, v_t) are independent normal random variables with zero mean and covariance matrix

(3) $$\Sigma = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix},$$

where $-1 \le \rho \le 1$ denotes the correlation between u_t and v_t (i.e. the *degree of endogeneity*), a parameter which affects the presence of bimodality in the density of the TSLS estimator (e.g. Maddala and Jeong (1992) and Woglom (2001)). If we apply the *canonical transformations* described in Theorem 3.3.1 of Phillips (1983) to the structural equation we obtain a *canonical structural parameter* $\beta^* = -\rho / \sqrt{1-\rho^2}$ that is a bijective function of the degree of endogeneity (e.g. equations (3.32) and (3.33) of Phillips (1983)).

For the simple model we consider, the TSLS estimator of β is

$$\hat{\beta} = \frac{x' P_Z y}{x' P_Z x},$$

where y and x are $(T \times 1)$ vectors having components y_t and x_t respectively, $P_Z = Z(Z'Z)^{-1}Z'$, and Z is a $T \times k$ matrix having the vectors z_t' as rows. We focus on

(4) $$\begin{aligned} w &= \frac{\sigma_v(\hat{\beta}-\beta)}{\sigma_u\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} = \frac{\sigma_v}{\sigma_u\sqrt{1-\rho^2}}\left(\hat{\beta} - \left[\beta + \frac{\sigma_u}{\sigma_v}\rho\right]\right) \\ &= \frac{\sigma_v(\hat{\beta}-\beta)}{\sigma_u\sqrt{1-\rho^2}} + \beta^*, \end{aligned}$$

that can be interpreted as the TSLS estimator of the structural parameter β^* in the structural equation after reduction to canonical form. Hillier (2006) calls

$e = (\sigma_v / \sigma_u)(\hat{\beta} - \beta) / \sqrt{1-\rho^2}$ the *scaled estimation error* in $\hat{\beta}$. It follows from equation (4) that $e = w - \beta^*$, so that the estimation error is zero (i.e. $\hat{\beta} = \beta$) if and only if $w = \beta^*$. Similarly, the TSLS estimator equals $\beta + (\sigma_u / \sigma_v)\rho$, the probability limit of the OLS estimator, if and only if $w = 0$. We will see later on that both β and $\beta + (\sigma_u / \sigma_v)\rho$ are, in the terminology of Phillips (2006), *magnetic attractors* for the probability mass of the density of the TSLS estimator.

The density of w is given by equation (3.45) of Phillips (1983) as

$$pdf\,(w) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\pi^{1/2}\Gamma\left(\frac{k}{2}\right)\left(1+w^2\right)^{\frac{k+1}{2}}} \times \tag{5}$$
$$\exp\left\{-\frac{\lambda}{2}\left(1+\beta^{*2}\right)\right\} \sum_{j=0}^{\infty} \frac{\left(\frac{k-1}{2}\right)_j}{j!\left(\frac{k}{2}\right)_j} \left(\frac{\lambda\beta^{*2}}{2}\right)^j {}_1F_1\left(\frac{k+1}{2}; \frac{k}{2} + j; \alpha(w)\right)$$

where $\lambda = \gamma' Z' Z \gamma / \sigma_v^2$ is the concentration parameter,

$$\alpha(w) = \frac{\lambda}{2} \frac{\left(1+\beta^* w\right)^2}{1+w^2}, \tag{6}$$

and ${}_1F_1(b;c;x)$ denotes a confluent hypergeometric function

$${}_1F_1(b;c;x) = \sum_{j=0}^{\infty} \frac{(b)_j}{j!(c)_j} x^j ,$$

(e.g. Slater (1960) for details). In the function above $(b)_j = b(b+1)\cdots(b+j-1)$. For $b > 0$ and $c > 0$, ${}_1F_1(b;c;x)$ is a monotonically increasing function of x. This property will be very useful later on. Equation (5) is the same as equation (22) of Hillier (2006) who writes λ and η for, respectively, our $\lambda\left(1+\beta^{*2}\right)$ and $-\beta^*$.

Equation (5) depends on β^*, λ and k only, and although it looks complicated, we will show in the next section that the shape of the density of the TSLS estimator depends on simple properties of confluent hypergeometric functions. In the just identified model considered by Woglom (2001) (i.e. $k = 1$) the density of the TSLS estimator simplifies considerably (e.g. equation (14) of Phillips (1980), or equation (3.35) of Phillips (1983)). Hillier (2006) gives a simple derivation of the density of both the TSLS and LIML estimators, and discusses (conditional) measures of precision.

3. Properties of the exact density

The density of the TSLS estimator can be written as the product of two terms

$$pdf\,(w) = lt(w) \times nc(w).$$

The first term, $lt(w)$, usually called the "leading term" (e.g. Phillips (1983)), is obtained by replacing $\lambda = 0$ in $pdf\,(w)$, and corresponds to the first line of equation (5). The second term, $nc(w)$, is the non-centrality component and is given by the second line of equation (5). It can be written as

$$(7) \qquad nc(w) = \exp\left\{-\frac{\lambda}{2}\left(1+\beta^{*2}\right)\right\} E_s\left\{ {}_1F_1\left(k;\frac{k}{2};\left[s\frac{\lambda\beta^{*2}}{2} + (1-s)\alpha(w)\right]\right)\right\}$$

where E_s denotes the expected value with respect to $s \sim Beta\left((k-1)/2,(k+1)/2\right)$.

The bimodality of the density of the TSLS estimator is generated by the interaction between $lt(w)$ and $nc(w)$. Equation (7) shows that $nc(w)$ is a monotonically increasing function of $\alpha(w)$, and its shape it mainly determined by $\alpha(w)$. If $\beta^* \neq 0$, the function $\alpha(w)$ has the form of a pulse wave, and as β^* increases it tends to become v-shaped since the crest (highest part of the wave) becomes less noticeable.

Proposition 1. (1) $lt(w)$ *is symmetric around the origin;*

(2) *If* $\beta^* = 0$ *then* $nc(w)$ *is bell-shaped but if* $\beta^* \neq 0$ *then* $nc(w)$ *has the form of a pulse wave; the undisturbed level (the equilibrium level as* w *tends to infinity) is at*

$$(8) \qquad nc^U = \exp\left\{-\frac{\lambda}{2}\right\} {}_1F_1\left(-\frac{k}{2};\frac{k}{2};-\frac{\lambda\beta^{*2}}{2}\right),$$

the crest (highest part of the wave) is at $w = \beta^*$ *where* $nc(w)$ *equals*

$$(9) \qquad nc^C = \sum_{j=0}^{\infty} \frac{\left(\frac{k-1}{2}\right)_j}{j!\left(\frac{k}{2}\right)_j}\left(\frac{\lambda\beta^{*2}}{2}\right)^j {}_1F_1\left(j-\frac{1}{2};\frac{k}{2}+j;-\frac{\lambda\left(1+\beta^{*2}\right)}{2}\right),$$

and the trough (lowest part of the wave) is at $w = -1/\beta^*$ *where* $nc(w)$ *takes on the value*

(10) $$nc^T = \exp\left\{-\frac{\lambda}{2}\right\} {}_1F_1\left(\frac{1}{2};\frac{k}{2};-\frac{\lambda\beta^{*2}}{2}\right).$$

While the leading term tends to centre the probability mass of the TSLS estimator at the probability limit of the OLS estimator, the term $nc(w)$ tends to concentrate it around β. When one of these two terms dominates, the density of the TSLS estimator has only one relevant mode. However, when none prevails, bimodality may appear (c.f. Hillier (2006)). Note that the leading term tends to shift the location of the principal mode of the density away from the correct point β (cf. Hillier (1990)).

One would expect to observe bimodality in the density of the TSLS estimator when $\left(nc^C - nc^U\right)/\left(nc^U - nc^T\right)$ is approximately one and $nc^C - nc^T$ is large, because in this case $nc(w)$ has a deep through and a high crest of similar size. The following results suggest situations when this may happen.

Proposition 2. *(1) If either* $\beta^* = 0$ *or* $\lambda = 0$, *then both* $nc(w)$ *and* $pdf(w)$ *are bell-shaped.*

(2) If $\beta^* \neq 0$ *and* λ *is large, then* $nc(w)$ *has a high crest (*$nc^C - nc^U = O\left(\lambda^{1/2}\right)$*) and a shallow trough (*$nc^U - nc^T = O(1/\lambda)$*). There could be two modes in the density of* w *but one of them would be very small and, certainly, undetectable for large values of the concentration parameter* λ.

(3) If $\lambda \neq 0$ *and* $\left|\beta^*\right|$ *is large, then* $nc(w)$ *has a high crest (*$nc^C - nc^U = O\left(\left|\beta^*\right|^{k_2}\right)$*) and a deep trough (*$nc^U - nc^T = O\left(\left|\beta^*\right|^{k_2}\right)$*), so that* $pdf(w)$ *could present two relevant modes (one on each side of* $w = 0$*).*

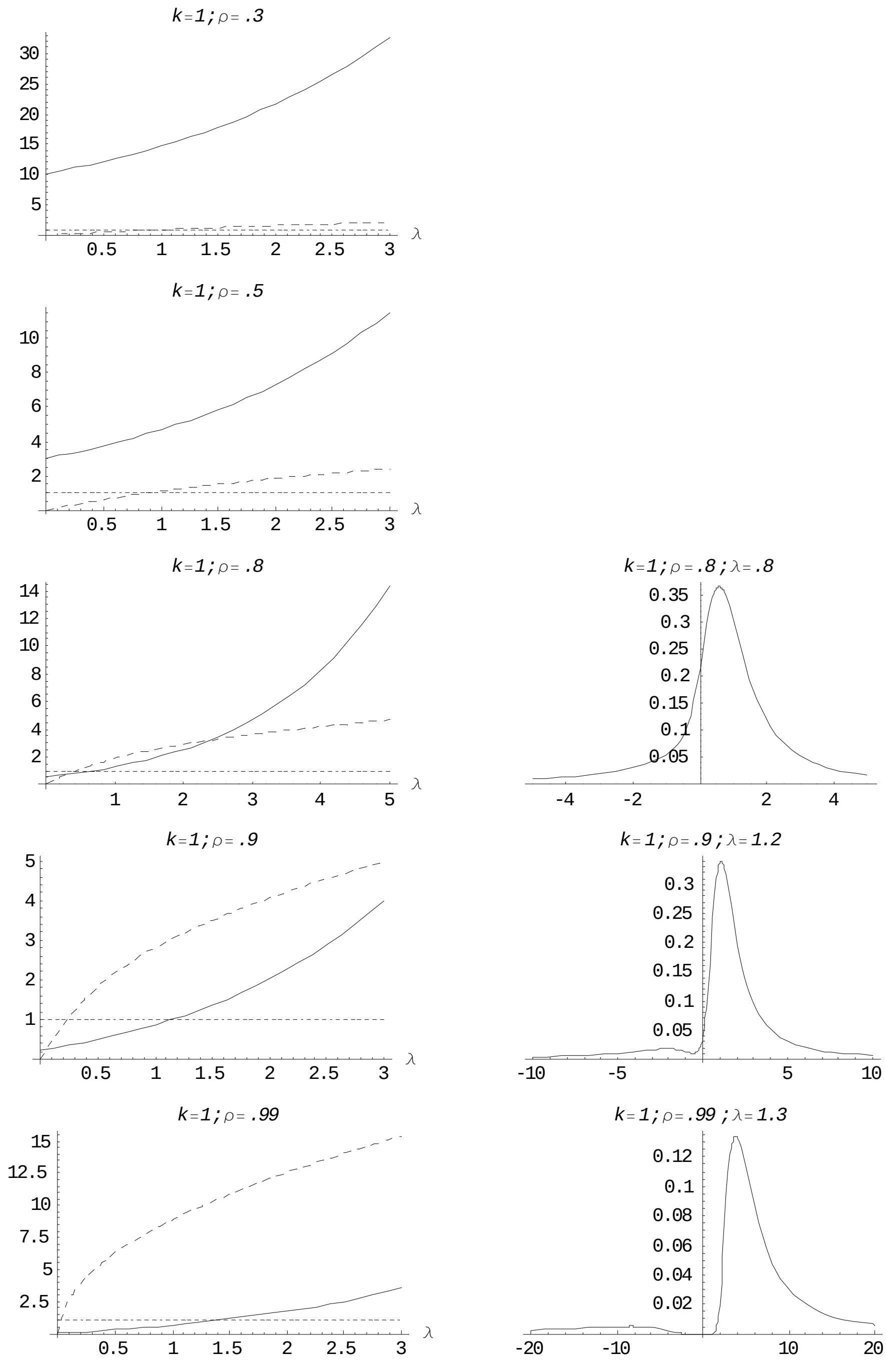

Figure 1. Graphs of $\left(nc^C - nc^U\right)/\left(nc^U - nc^T\right)$ (solid line), $nc^C - nc^T$ (dashed line) and 1 (dotted line) as functions of λ on the left-hand-side and densities of w on the right-hand-side, for $k = 1$ and different values of ρ. The graphs of the densities are shawn for values of λ yielding $\left(nc^C - nc^U\right)/\left(nc^U - nc^T\right) \simeq 1$. For $\rho = .3$ and $\rho = .5$, there are no values of λ for which $\left(nc^C - nc^U\right)/\left(nc^U - nc^T\right) \simeq 1$ and no density is shown on the right hand side.

Figure 1 shows some typical shapes of $\left(nc^{C}-nc^{U}\right)/\left(nc^{U}-nc^{T}\right)$ and $nc^{C}-nc^{T}$ as functions of λ for different values of ρ and $k=1$. When $|\rho|$ is small (e.g. $\rho=.3$ and $\rho=.5$ as in the first two graph from the top), $nc^{C}-nc^{U}$ is much larger than $nc^{U}-nc^{T}$ so that there cannot be any bimodality in the density of the TSLS estimator. When $\rho=.8$, $\left(nc^{C}-nc^{U}\right)/\left(nc^{U}-nc^{T}\right)\simeq 1$ for $\lambda=.8$, but $nc^{C}-nc^{T}$ is small around that point $\lambda=.8$, and the fluctuations of $nc\left(w\right)$ are not large enough to generate any bimodality in the density of the TSLS estimator. For $\rho=.9$, $\left(nc^{C}-nc^{U}\right)/\left(nc^{U}-nc^{T}\right)\simeq 1$ and $nc^{C}-nc^{T}$ is large and bimodality appears.

Regarding $nc\left(w\right)$ as the only factor determining bimodality of the density of the TSLS estimator is a simplification even if it gives an intuitive explanation for it. The leading term plays a very important role too, but the study of their interaction is very complicated. For example, for $k>1$, the graphs of $\left(nc^{C}-nc^{U}\right)/\left(nc^{U}-nc^{T}\right)$ and $nc^{C}-nc^{T}$ show similar patterns to those in Figure 1. However, the leading term tends to concentrate around zero and to dominate so that bimodality becomes less likely.

We now give a necessary condition for the existence of bimodality in the density of the TSLS estimator.

Theorem 1. *A necessary condition for the existence of bimodality in the density of the TSLS estimator is that $\Delta<0$ where*

$$\begin{aligned}\Delta=&\left(1+k\right)^{4}-6\left(1+k\right)^{3}\left(\beta^{*2}-1\right)\lambda+4\left(1+k\right)^{2}\left(3+5\beta^{*2}+3\beta^{*4}\right)\lambda^{2}\\&-8\left(1+k\right)\left(\beta^{*2}-1\right)\left(1+\beta^{*2}\right)^{2}\lambda^{3}-4\beta^{*2}\left(1+\beta^{*2}\right)^{2}\lambda^{4}.\end{aligned}\tag{11}$$

The shaded areas in the graphs in Figure 2 show the regions in the $\left(\lambda,\rho\right)$-plane where the necessary condition for bimodality is satisfied for different values of k. Figure 2 makes clear that high endogeneity is a necessary, although not sufficient, condition for bimodality. Moreover, since the shaded area (where $\Delta<0$) shrinks as the degree of over-identification increases, bimodality becomes less likely. This is formally stated in the following corollary.

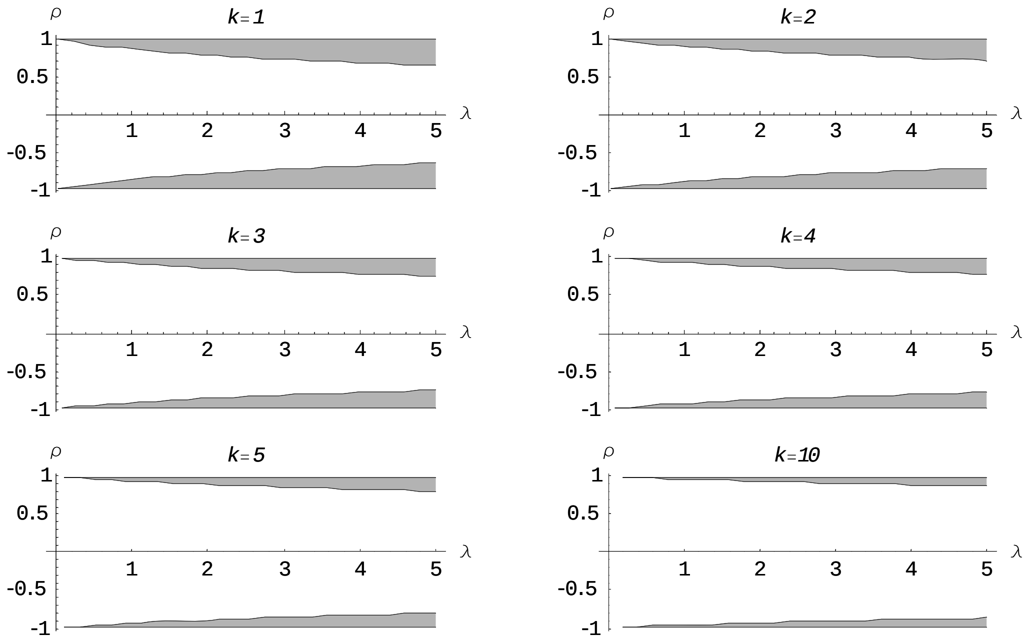

Figure 2. Regions in the (λ, ρ)-plane where the necessary condition for the existence of bimodality is satisfied (shaded area) for different values of k.

Corollary 1. *If k is large, $\beta^* \neq 0$ and finite, and $\lambda > 0$, then the density of the TSLS estimator has only one mode in a neighbourhood of $\beta + (\sigma_u / \sigma_v)\rho$.*

Therefore, as the number of instruments becomes large, the density of the TSLS estimator tends to have only one mode in the neighbourhood of $\beta + (\sigma_u / \sigma_v)\rho$. Intuitively, $(1+w^2)^{-(k+1)/2}$ in the leading term $lt(w)$ becomes concentrated around zero and dominates $nc(w)$ when k is large (c.f. Bekker (1994), Chao and Swanson (2005), Han and Phillips (2006) and Proposition 5 of Hillier (2006)).

The exact distribution of the OLS estimator of the canonical structural parameter β^* is given by equation (5) with k replaced by T (e.g. Phillips (1983)). The results of the recent analysis of Kiviet and Niemczyk (2005) about the small sample properties of the OLS and the TSLS estimators and their relative performance can be explained using Theorem 3 and Proposition 1. If T is large, and k and the noncentrality parameter are small, then the OLS estimator is very concentrated around its probability limit whereas the probability mass for the TSLS estimator may be split between $\beta + (\sigma_u / \sigma_v)\rho$ and β. In this situation, the OLS estimator may perform better than the TSLS estimator in terms of mean squared error (or analogous measures if $k = 1$). As k increases the difference between the densities of the OLS and the TSLS estimators vanishes (c.f. Bekker (1994), Remark 8 of Hillier (2006)). Hillier (2006) argues that the LIML estimator is better centred than the TSLS estimator when the degree of over-identification is large because its leading term does not depend on k.

In a just/over-identified structural equation, one may thus follow Woglom (2001) and conclude that “practically important bimodality [in the density of the TSLS estimator] requires high endogeneity [...] along with relatively small first stage correlation” (p. 1387). As the degree of over-identification increases bimodality becomes less likely.

4. The model with an unidentified structural parameter

Phillips and Wickens (1978), Nelson and Startz (1990) and Phillips (2006) consider a model with $k = 1$ consisting of a structural equation like the one in equation (1) and a structural identity (i.e. $x_t = y_t + z_t\gamma$ instead of equation (2)) with degree of exogeneity equal to one. They conclude that bimodality is always a feature of the exact density of the TSLS estimator. The model used in this note, specified in equations (1) and (2), does not contain any structural identity and it is not directly comparable to that of Phillips and Wickens (1978), Nelson and Startz (1990) and Phillips (2006). However, it can be used to illustrate the effects of strong endogeneity on the exact density of the TSLS estimator when the model is close to being unidentified. The differences between the models with and without a structural identity are discussed by Phillips (2006).

The partial Fisher information for β is $\lambda\sigma_v^2 / \sigma_u^2$, and it tends to zero as $\lambda \to 0$. In this case the non-centrality term vanishes (e.g. Phillips (1983)), unless the effect of $\lambda \to 0$ is cancelled out by the degree of endogeneity ρ^2 going to one (or, equivalently, $\beta^{*2} \to \infty$). We will study these limits along a path of the form $\lambda = \theta(1-\rho^2) + o(1-\rho^2)$, $0 \le \theta < \infty$. In this case w is not defined but the limiting density of the TSLS estimator is. In the notation of Section 2.4 of Hillier (2006), the direction φ tends to $(0,\pm 1)'$ when $\beta^{*2} \to \infty$, and these are points where the coordinates $\varphi = \left(1, -\beta^*\right)' / \sqrt{1+\beta^{*2}}$ are not defined.

If $\theta = 0$, the TSLS estimator has a Cauchy distribution for any $|\rho| < 1$ (e.g. Phillips (1983)). If $\theta > 0$, the situation is more complicated, and the following theorems give some insights about the shape of the limit density.

Theorem 2. *Suppose* $\lambda \to 0$ *and* $\rho^2 \to 1$ *on the path* $\lambda = \theta\left(1-\rho^2\right) + o\left(1-\rho^2\right)$, $0 \le \theta < \infty$, *then (i) the density of the TSLS estimator is*

$$pdf\left(w\right) = \frac{\Gamma\left(\frac{k+1}{2}\right)\exp\left\{-\frac{\theta}{2}\right\}}{\pi^{\frac{1}{2}}\Gamma\left(\frac{k}{2}\right)\left(1+w^2\right)^{\frac{k+1}{2}}} \sum_{j=0}^{\infty} \frac{\left(\frac{k-1}{2}\right)_j}{j!\left(\frac{k}{2}\right)_j}\left(\frac{\theta}{2}\right)^j {}_1F_1\left(\frac{k+1}{2}; \frac{k}{2}+j; \frac{\theta}{2}\frac{w^2}{1+w^2}\right).$$

(ii) If the model is just identified ($k = 1$) *then the limit of the density simplifies to*

$$pdf\left(w\right) = \frac{\exp\left\{-\frac{\theta}{2}\right\}}{\pi\left(1+w^2\right)} {}_1F_1\left(1; \frac{1}{2}; \frac{\theta}{2}\frac{w^2}{1+w^2}\right).$$

Theorem 3. *The limit density in Theorem 1 has the following properties:*
(i) if $k = 1$ *bimodality occurs for* $\theta > 1$*;*
(ii) if $k = 2$ *then bimodality occurs for* $\theta > 3.15991$*;*
(iii) if $k \ge 3$ *the density is always unimodal.*

Figure 3 shows the limit density in the just identified case for some values of parameter θ. Theorem 3 suggests that if the model is unidentified (or close to being unidentified) and the number of instruments is large then the distribution of the TSLS estimator is also concentrated around $\beta + \left(\sigma_u / \sigma_v\right)\rho$. This result holds true independently of the path chosen to calculate the limit density. Moreover, since the exact density given in Theorem 1 does not depend on the sample size, it is also the asymptotic density for the TSLS estimator.

For $k \le 2$, if the value of θ is sufficiently large so that λ is larger than $1-\rho^2$, bimodality arises because the TSLS estimator is simultaneously attracted towards the values of $\pm\beta$ and $\beta + \left(\sigma_u / \sigma_v\right)\rho$. Phillips (2006) provides an asymptotic expansion for the position of the modes in the just-identified case that captures the criterion in Theorem 3 *(i)*. If $k \geq 3$, then there is no bimodality because the attraction towards the probability limit of the OLS estimator prevails.

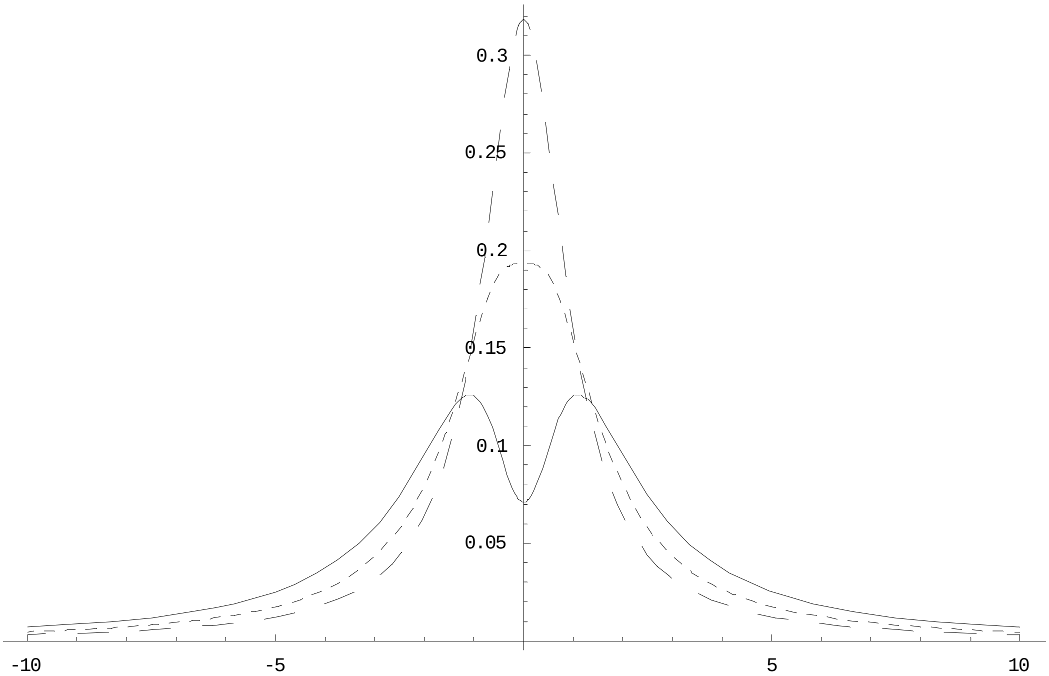

Figure 3. Graph of the limit densities of w in the just-identified case for $\theta = 0$ (dashed line), $\theta = 1$ (dotted line), and $\theta = 3$ (solid line).

5. Conclusions

Phillips and Wickens (1978), Nelson and Startz (1990), Maddala and Jeong (1992) and Woglom (2001) have shown that the density of the TSLS estimator may be bimodal in a just identified structural equation. This paper has looked further at this issue in a just/over-identified structural equation in order to provide a better understanding of the problem.

Our conclusions are as follows.

(1) Bimodality arises because of the complex interaction between two components of the exact density: one of these is symmetric and one has the shape of a pulse wave. Depending on the value of three key parameters, the noncentrality parameter, the degree of endogeneity and the degree of overidentification, we may or may not observe bimodality in the density of the TSLS estimator.

(2) As for the just-identified case, bimodality occurs if the noncentrality parameter is large. However, one of the modes would be surely undetectable in this case. Clear bimodality tends to occur when the degree of endogeneity is close to one and the

noncentrality parameter is relatively small.

(3) The limit density of the TSLS estimator as the noncentrality parameter tends to zero and the degree of endogeneity tends to one may have one or two modes when the degree of over-identification is less or equal to two. In all other cases it has only one mode centred around the probability limit of the OLS estimator.

(4) When the noncentrality parameter is finite, bimodality in the density of the TSLS estimator becomes less likely as the degree of over-identification increases (in this case the density has only one mode in the neighbourhood of the probability limit of the OLS estimator).

6. Technical appendix

Proof of Proposition 1

Equations (9) and (10) can be easily obtained from the second line of (5). Equation (8) follows easily from (7) after taking the limit as w tends to infinity.

Proof of Proposition 2

These results follow from Proposition 1 and approximations for the confluent hypergeometric function reviewed by Slater (1960).

Proof of Theorem 1

Bimodality exists when

$$\frac{dpdf\,(w)}{dw} = \frac{dlt(w)}{dw} \times nc(w) + lt(w) \times \frac{dnc(w)}{dw} = 0$$

has at least three real solutions in w. The last equation can be re-written as

(12) $$-\frac{\frac{dlt(w)}{dw}}{lt(w)} = \frac{\frac{dnc(w)}{dw}}{nc(w)}.$$

It can be easily checked that $-\left(dlt(w)/dw\right)/lt(w) = (1+k)\,w/\left(1+w^2\right)$. From (7) we find that

$$\frac{\frac{dnc(w)}{dw}}{nc(w)} = 2\frac{(\beta^* - w)(1+\beta^* w)\lambda}{(1+w^2)^2}\frac{E_s\left\{(1-s)\,{}_1F_1\left(k+1;\frac{k}{2}+1;\left[s\frac{\lambda\beta^{*2}}{2}+(1-s)\alpha(w)\right]\right)\right\}}{E_s\left\{{}_1F_1\left(k;\frac{k}{2};\left[s\frac{\lambda\beta^{*2}}{2}+(1-s)\alpha(w)\right]\right)\right\}}.$$

This is a continuous function of w, moreover, since for every $q \ge 0$ we have ${}_1F_1(k;k/2;q) \ge {}_1F_1(k+1;k/2+1;q)$, and $1-s \le 1$, the ratio of the two expectations in the above display is always between zero and one, so that the right-hand side of equation (12), as a function of w, is always between the horizontal axis and the function

$$2\frac{(\beta^* - w)(1+\beta^* w)\lambda}{(1+w^2)^2}.$$

Therefore, a necessary but not sufficient condition for the existence of bimodality is that

$$\frac{(1+k)w}{1+w^2} = 2\frac{(\beta^* - w)(1+\beta^* w)\lambda}{(1+w^2)^2}$$

has at least three real solutions. Simplifying and rearranging the last equation, we conclude that a necessary condition for the existence of bimodality is that the cubic equation

$$(1+k)w^3 + 2\beta^*\lambda w^2 + (1+k+2\lambda-2\beta^{*2}\lambda)w - 2\beta^*\lambda = 0$$

has three real solutions. The theorem follows from standard results on the number of real solutions of cubic equations.

Proof of Corollary 1

We have seen in the proof of Theorem 1 that a necessary condition for the existence of bimodality is that

(13) $$(1+k)w = 2\frac{(\beta^* - w)(1+\beta^* w)\lambda}{1+w^2}.$$

As $k \to \infty$ the left-hand side increases but the right-hand side stays the same for fixed β and λ. If k is large enough the above equation will have only one real solution between $\min\{\beta^*, -1/\beta^*\}$ and $\max\{\beta^*, -1/\beta^*\}$. The slope of $(1+k)w$ tends to infinity as $k \to \infty$, so that $(1+k)w$ tends to a vertical line going though the origin and intersects the curve representing the right-hand side of equation (13) in a neighbourhood of zero only.

Proof of Theorem 2

Using (5) we can write $\lambda = \theta/\left(1+\beta^{*2}\right)+o\left(\left(\beta^{*}\right)^{-2}\right)$ and the statement of the theorem follows easily from the continuity of the exponential and of the hypergeometric functions.

Proof of Theorem 3

It is easily verified that the limit density can have only two modes, and that if a trough exists it must occur at $w=0$. Moreover, one can easily show that

$$\frac{d^2}{dw^2}\frac{{}_1F_1\left(\frac{k+1}{2};\frac{k}{2}+j;\frac{\theta}{2}\frac{w^2}{1+w^2}\right)}{\left(1+w^2\right)^{\frac{k+1}{2}}}\Bigg|_{w=0}=\frac{k+1}{4}\Gamma\left(\frac{k}{2}\right)\left(\frac{k}{2}\right)_j\left[-\frac{4}{\Gamma\left(\frac{k}{2}\right)\left(\frac{k}{2}\right)_j}+\frac{2\theta}{\Gamma\left(\frac{k}{2}+1\right)\left(\frac{k}{2}+1\right)_j}\right]$$

so that

$$\frac{d^2 pdf\left(w\right)}{dw^2}\Bigg|_{w=0}=\frac{k+1}{4}\frac{\Gamma\left(\frac{k+1}{2}\right)\exp\left\{-\frac{\theta}{2}\right\}}{\pi^{\frac{1}{2}}\left(1+w^2\right)^{\frac{k+1}{2}}}\sum_{j=0}^{\infty}\frac{\left(\frac{k-1}{2}\right)_j}{j!}\left(\frac{\theta}{2}\right)^j\left[-\frac{4}{\Gamma\left(\frac{k}{2}\right)\left(\frac{k}{2}\right)_j}+\frac{2\theta}{\Gamma\left(\frac{k}{2}+1\right)\left(\frac{k}{2}+1\right)_j}\right].$$

After using equation (2.2.4) of Slater (1960) and simplifying, one obtains

$$\frac{d^2 pdf\left(w\right)}{dw^2}\Bigg|_{w=0}=-\frac{\left(k+1\right)\Gamma\left(\frac{k+1}{2}\right)\exp\left\{-\frac{\theta}{2}\right\}}{\Gamma\left(\frac{k}{2}+1\right)\pi^{\frac{1}{2}}\left(1+w^2\right)^{\frac{k+1}{2}}}\,{}_1F_1\left(\frac{k-3}{2};\frac{k}{2};\frac{\theta}{2}\right)$$

and $d^2 pdf\left(w\right)/dw^2\big|_{w=0}>0$ if and only if ${}_1F_1\left(\left(k-3\right)/2;k/2;\theta/2\right)<0$. For $k=1$ one has ${}_1F_1\left(-1;1/2;\theta/2\right)=1-\theta<0$ which implies $\theta>1$. For $k=2$, ${}_1F_1\left(-1/2;1;\theta/2\right)<0$, which implies $\theta>3.15991$. For $k\geq 3$, ${}_1F_1\left(\left(k-3\right)/2;k/2;\theta/2\right)\geq 1$ for all θ so that $d^2 pdf\left(w\right)/dw^2\big|_{w=0}\leq 0$ for all θ.

References

Bekker, P. A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62: 657-681.

Chao, J. C. and N. R. Swanson (2005). "Consistent Estimation with a Large Number of Weak Instruments." *Econometrica* 73: 1673-1692.

Han, C. and P. C. B. Phillips (2006). "GMM with Many Moment Conditions." *Econometrica* 74: 147-192.

Hillier, G. H. (1990). "On the Normalization of Structural Equations: Properties of Direction Estimators." *Econometrica* 58: 1181-1194.

Hillier, G. H. (2006). "Yet More on The Exact Properties of IV Estimators." *Econometric Theory* Forthcoming.

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

Maddala, G. S. and J. Jeong (1992). "On the Exact Small Sample Distribution of the Instrumental Variable Estimator." *Econometrica* 60: 181-183.

Nelson, C. R. and 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". *Handbook of Econometrics*. Z. Griliches and M. D. Intriligator. Amsterdam, North-Holland. I**:** 449-516.

Phillips, P. C. B. (2006). "A Remark on Bimodality and Weak Instrumentation in Structural Equation Estimation." *Econometric Theory* Forthcoming.

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

Slater, L. J. (1960). *Confluent Hypergeometric Functions*. London, Cambridge University Press.

Woglom, G. (2001). "More Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica* 69: 1381-1389.