ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

On the Bimodality of the Exact Distribution of the TSLS Estimator

G. Forchini

May 2005

Working Paper 14/05

On the Bimodality of the Exact Distribution of the TSLS Estimator

G. Forchini *†

Monash University

21 March 2005

Abstract

Nelson and Startz (*Econometrica*, 58, 1990), Maddala and Jong (*Econometrica*, 60, 1992) and Wolgrom (*Econometrica*, 69, 2001) have shown that the density of the two-stage least squares estimator may be bimodal in a just identified structural equation. This paper further investigates the conditions under which bimodality may arise in a just/over-identified model.

JEL Classification C30
Key Words: Bimodality, Identification, Structural equation, Two Stage Least Squares

*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

†I thank Grant Hillier, Patrick Marsh and Don Poskitt for useful and encouraging comments and discussions.

1 Introduction

Although the exact density of the two stage least squares (TSLS) estimator has been known for a few decades (see for example the review by Phillips (1983)) some of its properties are still surprising for econometricians. Bimodality is one of these *unexpected* properties: 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.

Woglom (2001), who made the most recent contribution to this literature, has raised the following three points:

(1) the exact finite sample distribution of the TSLS estimator cannot be easily interpreted (p. 1381);

(2) two conflicting results are available in the literature (p. 1381 and p. 1388): as the structural equation becomes unidentified, the distribution of the TSLS estimator approaches a Cauchy distribution (e.g. Phillips (1983)); however, if the correlation between the right-hand-side endogenous variables and the instruments is one (in absolute value) then the density of the TSLS estimator is bimodal (Nelson and Startz (1990));

(3) when the degree of endogeneity is large and the structural equation is weakly identified, the distribution of the TSLS estimator may have two relevant modes (p. 1388).

This paper addresses these three issues in the context of a just/over-identified structural equation. Firstly, we show that the possible bimodality of the density of the TSLS estimator can be easily understood using the exact results reviewed by Phillips (1983). In fact, it is the product of the interaction of two components of the exact density: one is unimodal and symmetric, and the other has the shape of a pulse wave. Secondly, we explain the apparent conflict between the totally unidentified case considered by Phillips (1983) and the bimodal density derived by Nelson and Startz (1990). We argue that there is 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. Finally, we study the relationship between the degree of overidentification and bimodality of the TSLS estimator, and show that bimodality cannot exist if the degree of overidentification is large enough.

Hillier (2004) has recently investigated the properties of the TSLS estimator in the just identified model considered by Woglom (2001) and offered further insights into the properties of its density by relating them to the normalization used for the structural equations. He has also discussed conditional measures of precision for the TSLS estimator in this context (see also Forchini and Hillier (2003)), and has given a very simple derivation of the density of the TSLS estimator in a just identified model.

The rest of the paper is organized as follows. Section 2 presents the model under consideration. 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 discusses the case where the degree of overidentification is large, and Section 6 concludes. All technical results are proved in the appendix.

2 The two endogenous variables model

Consider the simple instrumental variables model:

$$y_{1t} = y_{2t}\beta + u_t \tag{1}$$

$$y_{2t} = z_{2t}'\pi_2 + v_{2t}, \quad t = 1, 2, ..., T \tag{2}$$

where y_{1t} and y_{2t} are endogenous variables, z_{2t} is a $k_2 \times 1$ vector of exogenous variables, β and π_2 are unknown parameters of dimension 1×1 and $k_2 \times 1$, respectively, and u_t and v_{2t} are random errors. The reduced form for (y_{1t}, y_{2t}) is

$$(y_{1t}, y_{2t}) = z_{2t}'(\pi_1, \pi_2) + (v_{1t}, v_{2t}) \tag{3}$$

where

$$\pi_1 = \pi_2\beta \tag{4}$$

and $v_{1t} = u_t + \beta v_{2t}$. If $\pi_2 \neq 0$, equation (4) uniquely defines the parameter β in terms of the reduced form parameters (π_1, π_2), and if $k_2 = 1$ then $\beta = \pi_1/\pi_2$. If $\pi_2 = 0$, i.e. if the structural equation is not identified, then β can take on any finite value provided $\pi_1 = 0$ (see Forchini and Hillier (2003) for further discussion). It is assumed that the model is identified (i.e. $\pi_2 \neq 0$ and (4) holds), but π_2 can be arbitrarily close to 0. Woglom (2001) and the work cited therein have looked at the special case where $k_2 = 1$.

Without loss of generality we can assume that the standardizing transformations described in Theorem 3.3.1 of Phillips (1983) have been applied, and that v_1 and v_2 are independent vectors of independent standard normal random variables. There are two reasons to do so. First, these transformations allow us to simplify the technical results, but do not affect the nature of the problem. If we denote the variables and the parameters in the unstandardized structural equation and reduced form with an asterisk, so that, for example, β^* is the coefficient of the endogenous variable in the *unstandardized* model equivalent to (1) and (2), and if the 2×2 matrix

$$\Omega^* = \begin{pmatrix} \omega_{11}^* & \omega_{21}^* \\ \omega_{21}^* & \omega_{22}^* \end{pmatrix}$$

denotes the common covariance matrix of the *unstandardized* reduced form errors (v_{1t}^*, v_{2t}^*), then the standardized and the unstandardized parameters β and β^* are related by

$$\beta = \left(\omega_{22}^* / \omega_{11.2}^*\right)^{1/2} \left(\beta^* - \omega_{21}^* / \omega_{22}^*\right),$$

where $\omega_{11.2}^* = \omega_{11}^* - \left(\omega_{21}^*\right)^2 / \omega_{22}^*$. Thus, for a fixed covariance matrix Ω^*, β is a simple linear transformation of β^*. One may note that if the structural equation is unidentified then the ordinary least squares estimator $\widehat{\beta}_{OLS}$ of β^* converges in probability to $\text{plim}\, \widehat{\beta}_{OLS} = \omega_{21}^* / \omega_{22}^*$ as the sample size increases. The quantity $\omega_{22}^* / \omega_{11.2}^*$ is the ratio between the variance of y_{2t}^* and the conditional variance of y_{1t}^* given y_{2t}^*.

A second reason to look at the standardized model is that β is a bijective function of the *degree of endogeneity* (e.g. equations (3.32) and (3.33) of Phillips (1983))

$$\text{corr}\left(u_t, y_{2t}\right) = \text{corr}\left(u_t^*, y_{2t}^*\right) = \rho = -\frac{\beta}{\sqrt{1 + \beta^2}}, \tag{5}$$

a parameter which seems to affect the presence of bimodality in the density of the TSLS estimator (e.g. Maddala and Jeong (1992) and Woglom (2001)). By focusing on β we can better take into accout the influence of the degree of endogeneity on the shape of the distribution. The absolute value of the correlation is close to one for large values of $|\beta|$. For example, β must be at close to 7.02 to produce ρ equal to .99. However, ρ reaches ± 1 only when β tends to $\pm\infty$. Equation (5) seems to have been ignored by Nelson and Startz (1990), Maddala and Jeong (1992), and Woglom (2001).

In this simple model, the TSLS estimator of β is

$$\hat{\beta} = \frac{y_2' P_{Z_2} y_1}{y_2' P_{Z_2} y_2},$$

where y_1 and y_2 are $t \times 1$ having components y_{1t} and y_{2t} respectively, $P_{Z_2} = Z_2 \left(Z_2' Z_2\right)^{-1} Z_2'$ and Z_2 is a $T \times k_2$ matrix having the variables z_{2t}' as rows. Theorem 3.3.2 of Phillips (1983) shows that the TSLS estimator of the standardized coefficient of the endogenous variable ($\hat{\beta}$) is related to the estimator of the unstandardized coefficient ($\hat{\beta}^*$) by the same relationship defining the coefficient, i.e.

$$\hat{\beta} = \left(\omega_{22}^* / \omega_{11.2}^*\right)^{1/2} \left(\hat{\beta}^* - \omega_{21}^* / \omega_{22}^*\right).$$

Woglom (2001) has considered the distribution of

$$w = \hat{\beta}^* - \beta^* = \left(\omega_{22}^* / \omega_{11.2}^*\right)^{-1/2} \left(\hat{\beta} - \beta\right), \tag{6}$$

and has studied its dependence on the concentration parameter μ^2 ($= T\pi_2'\pi_2$ in the standardized model) and the degree of endogeneity ρ. However, we prefer to work with the standardized TSLS estimator $\hat{\beta}$ directly. Its density is given by equation (3.45) of Phillips (1983) as

$$\begin{aligned} \text{pdf}\left(\hat{\beta}\right) &= \frac{\Gamma\left(\frac{k_2+1}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{k_2}{2}\right) \left(1 + \hat{\beta}^2\right)^{\frac{k_2+1}{2}}} \times \\ & \exp\left\{-\frac{\mu^2}{2}\left(1+\beta^2\right)\right\} \sum_{j=0}^{\infty} \frac{\left(\frac{k_2-1}{2}\right)_j}{j! \left(\frac{k_2}{2}\right)_j} \left(\frac{\mu^2 \beta^2}{2}\right)^j \times \\ & {}_1F_1\left(\frac{k_2+1}{2}; \frac{k_2}{2} + j; \mathrm{a}\left(\hat{\beta}\right)\right) \end{aligned} \tag{7}$$

where

$$\mu^2 = T\pi_2'\pi_2$$

is the concentration parameter,

$$\mathrm{a}\left(\hat{\beta}\right) = \frac{\mu^2}{2}\frac{\left(1+\beta\hat{\beta}\right)^2}{1+\hat{\beta}^2}, \tag{8}$$

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 display 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 is a property that will be very usefull later on.

In the just identified model considered by Woglom (2001) (i.e. $k_2 = 1$) the density of the TSLS estimator simplifies to

$$\mathrm{pdf}\left(\hat{\beta}\right) = \frac{\exp\left\{-\frac{\mu^2}{2}\left(1+\beta^2\right)\right\}}{\pi\left(1+\hat{\beta}^2\right)}\,{}_1F_1\left(1;\frac{1}{2};\mathrm{a}\left(\hat{\beta}\right)\right) \tag{9}$$

(e.g. equation (14) of Phillips (1980) or equation (3.35) of Phillips (1983)). Hillier (2004) gives a simple derivation of equation (9), and discusses (conditional) measures of precision of the TSLS estimator.

Equations (7) and (9) depend on two parameters, β and μ^2, only, and although they look complicated, we will be show in the next section that the shape of the density of the TSLS depends on some simple properties of the confluent hypergeometric function.

3 Properties of the exact densities

The density pdf $\left(\hat{\beta}\right)$ can be written as the product of two terms

$$\mathrm{pdf}\left(\hat{\beta}\right) = \mathrm{lt}\left(\hat{\beta}\right) \mathrm{w}\left(\hat{\beta}\right).$$

The first term lt $\left(\hat{\beta}\right)$, usually called the "leading term" (e.g. Phillips (1983)), is obtained by replacing $\mu^2 = 0$ in pdf $\left(\hat{\beta}\right)$ and corresponds to the first line of equation (7). The second term w $\left(\hat{\beta}\right)$ is given by the second and third lines of equation (7).

The function w $\left(\hat{\beta}\right)$ depends on $\hat{\beta}$ only through a $\left(\hat{\beta}\right)$, i.e. w $\left(\hat{\beta}\right)$ $= \mathrm{w}_1\left(\mathrm{a}\left(\hat{\beta}\right)\right)$. The function $\mathrm{w}_1(.)$ is an infinite linear combination (with positive coefficients) of confluent hypergeometric functions. We have already observed that confluent hypergeometric functions like those appearing in $\mathrm{w}_1(.)$ are monotonically increasing, so that $\mathrm{w}_1(.)$ is itself a monotonically increasing function. This implies that w $\left(\hat{\beta}\right)$ being the composition of $\mathrm{w}_1(.)$ and a $\left(\hat{\beta}\right)$ has its shape mainly determined by a $\left(\hat{\beta}\right)$. The function a $\left(\hat{\beta}\right)$ has the form of a pulse wave, and as β increases it tends to become v-shaped since the crest becomes less noticeable.

[FIGURE 1 APPROXIMATELY HERE]

Therefore, we can conclude that

Proposition 1 *(1)* lt $\left(\hat{\beta}\right)$ *is symmetric around the origin* $\hat{\beta} = 0$*;*
(2) w $\left(\hat{\beta}\right)$ *has the form of a pulse wave; its undisturbed level is at*

$$\mathrm{w}^U = \exp\left\{-\frac{\mu^2}{2}\right\} {}_1F_1\left(-\frac{k_2}{2}; \frac{k_2}{2}; -\frac{\mu^2\beta^2}{2}\right), \tag{10}$$

the crest is at $\hat{\beta} = \beta$ *where* $\mathrm{w}\left(\hat{\beta}\right)$ *equals*

$$\mathrm{w}^C = \sum_{j=0}^{\infty} \frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}\right)_j} \left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(j - \frac{1}{2}; \frac{k_2}{2} + j; -\frac{\mu^2\left(1+\beta^2\right)}{2}\right), \tag{11}$$

and the trough is at $\hat{\beta} = -1/\beta$ *where* $\mathrm{w}\left(\hat{\beta}\right)$ *takes on the value*

$$\mathrm{w}^T = \exp\left\{-\frac{\mu^2}{2}\right\} {}_1F_1\left(\frac{1}{2}; \frac{k_2}{2}; -\frac{\mu^2\beta^2}{2}\right). \tag{12}$$

It follows from Proposition 1 and the discussion above that

Proposition 2 *(1) If either* $\beta = 0$ *or* $\mu^2 = 0$*, then* $\mathrm{w}\left(\hat{\beta}\right)$ *is constant, and* $\mathrm{pdf}\left(\hat{\beta}\right)$ *is bell-shaped.*
(2) If $\beta \neq 0$ *and* μ^2 *is large, then* $\mathrm{w}\left(\hat{\beta}\right)$ *has a high crest (*$\mathrm{w}^C - \mathrm{w}^U = O\left(|\mu|\right)$*) and a shallow trough (*$\mathrm{w}^U - \mathrm{w}^T = O\left(\mu^{-2}\right)$*). There could be two modes in the density of* $\hat{\beta}$ *but one of them would be very small and, certainly, undetectable for large values of the concentration parameter* μ^2*.*
(3) If $\mu^2 \neq 0$ *and* $|\beta|$ *is large, then* $\mathrm{w}\left(\hat{\beta}\right)$ *has a high crest (*$\mathrm{w}^C - \mathrm{w}^U = O\left(|\beta|^{k_2}\right)$*) and a deep trough (*$\mathrm{w}^U - \mathrm{w}^T = O\left(|\beta|^{k_2}\right)$*), so that* $\mathrm{pdf}\left(\hat{\beta}\right)$ *could present two relevant modes (one on each side of* $\hat{\beta} = 0$*).*

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). We will focus on such situations from now onwards.

4 The unidentified model

We consider the limit situation where the model is unidentified, and the degree of endogeneity measured by ρ^2 is 1. To do so we define the density of the TSLS estimator in the unidentified case as the limit density when μ^2 tends to zero and ρ^2 tends to 1 along a path of the form $\mu^2 = a(1-\rho^2) + o(1-\rho^2)$, $a \geq 0$. This will also clarify the conflicting results of Phillips (1983) and Nelson and Startz (1990).

The following theorems give some insights about the shape of the limit density.

Theorem 1 *(i) Suppose $\mu^2 \to 0$ and $\rho^2 \to 1$ on the path $\mu^2 = a(1-\rho^2) + o(1-\rho^2)$, $a \geq 0$, then the density of the TSLS estimator is*

$$\begin{aligned}\mathrm{pdf}\left(\hat{\beta}\right) &= \frac{\Gamma\left(\frac{k_2+1}{2}\right)\exp\left\{-\frac{a}{2}\right\}}{\pi^{\frac{1}{2}}\Gamma\left(\frac{k_2}{2}\right)\left(1+\hat{\beta}^2\right)^{\frac{k_2+1}{2}}} \\ &\quad \sum_{j=0}^{\infty}\frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}\right)_j}\left(\frac{a}{2}\right)^j {}_1F_1\left(\frac{k_2+1}{2};\frac{k_2}{2}+j;\frac{a}{2}\frac{\hat{\beta}^2}{1+\hat{\beta}^2}\right).\end{aligned}$$

(ii) If the model is just identified ($k_2 = 1$) then the limit of the density simplifies to

$$\mathrm{pdf}\left(\hat{\beta}\right) = \frac{\exp\left\{-\frac{a}{2}\right\}}{\pi\left(1+\hat{\beta}^2\right)} {}_1F_1\left(1;\frac{1}{2};\frac{a}{2}\frac{\hat{\beta}^2}{1+\hat{\beta}^2}\right).$$

Theorem 2 *The limit density in Theorem 2 has the following properties:*

(i) if $k_2 = 1$ bimodality occurs for $a > 1$;

(iii) if $k_2 = 2$ then bimodality occurs for $a > 3.15991$;

(iv) if $k_2 \geq 3$ the density is always unimodal.

For the just identified case, the different results of Phillips (1983) and Nelson and Startz (1990) are due to the different paths chosen in the calculation of the limit as μ^2 tends to zero and ρ^2 tends to one. Phillips (1983) does not make any assumption on ρ^2 and thus implicitly assumes that $a = 0$. Depending on the value of a, the limit density can be bimodal or unimodal for $k_2 \leq 2$. However, no bimodality arises for $k_2 \geq 3$.

Figure 2 shows the limit density in the just identified case for some values of parameter a.

[FIGURE 2 APPROXIMATELY HERE]

One may note that the ordinary least squares estimator of β would converge in probability to zero in our setup where the standardizing transformations described in Theorem 3.3.1 of Phillips (1983) have been applied. Theorem 2 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 0. 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.

5 The existence of modes when k_2 is large

Section 3 has shown that relevant bimodality requires a relatively small μ^2. The limit densities of the TSLS estimator obtained in Section 4 (where μ^2 tends to zero) are necessarily unimodal when $k_2 \geq 3$. This section investigates the possible bimodality of the density of the TSLS estimator when μ^2 is finite (but not zero) and k_2 is large.

The following theorem and its corollary give formal conditions for the existence of modes.

Theorem 3 *The TSLS estimator has one mode if the equation*

$$\frac{\sum_{j=0}^{\infty}\frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}+1\right)_j}\left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(\frac{k_2+3}{2};\frac{k_2+2}{2}+j;\frac{\mu^2}{2}\frac{(1+\beta x)^2}{1+x^2}\right)}{\sum_{j=0}^{\infty}\frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}\right)_j}\left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(\frac{k_2+1}{2};\frac{k_2}{2}+j;\frac{\mu^2}{2}\frac{(1+\beta x)^2}{1+x^2}\right)} = \frac{k_2 x\left(1+x^2\right)}{\mu^2\left(\beta-x\right)\left(1+\beta x\right)} \quad (13)$$

has only one solution in x; it has two modes if the equation above has three solutions in x. There is one solution to equation (13) in the interval $[\min\{\beta,-1/\beta\},\max\{\beta,-1/\beta\}]$. *If there are three solutions, two will be in the range* $(-1/\beta,+\infty)$ *if* $\beta<0$ *and in* $(-\infty,-1/\beta)$ *if* $\beta>0$.

Corollary 1 *If k_2 is large, $\beta\neq 0$ and $\mu^2>0$, then equation (13) has only one solution at $x=0$.*

Therefore, as k_2 becomes large, the density of the TSLS estimator tends to have only one mode in the neighborhood of $\hat{\beta}=0$. Corollary 1 shows that as the number of instruments increases the distribution of the TSLS estimator is concentrated around the probability limit for the OLS estimator in the unidentified case. Intuitively, $\left(1+\hat{\beta}^2\right)^{-(k_2+1)/2}$ in the leading term $\mathrm{lt}\left(\hat{\beta}\right)$ becomes concentrated around zero when k_2 is large.

6 Conclusions

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. It has argued that 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.

The paper has shown that bimodality of the density of the TSLS estimator may appear if μ^2 is large, but one of the modes would be surely undetectable in this case. As in the just identified case bimodality may occur when ρ^2 is close to one and μ^2 is relatively small. However, it becomes less likely when k_2 is large (in this case the density has only one mode in the neighborhood of zero). In situations where identification is weak, the central tendency of the TSLS estimator is biased away from the true value in the direction of the probability limit of the ordinary least squares estimator (Nelson and Startz (1990) p. 967).

Finally we have shown that both the Cauchy density and a bimodal density can be obtained as μ^2 tends to zero and ρ^2 tends to one. In fact, the density of the TSLS may converge to a large variety of (possibly bimodal) densities as μ^2 approaches zero and ρ^2 goes to one. This reconciles the results of Phillips (1983) and Nelson and Startz (1990) for the just identified case. If $k_2 \geq 3$ no bimodality can be present in the limit density.

A Technical appendix

A.1 Proof of Theorem 1

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

A.2 Proof of Theorem 2

It can be easily checked that the limit density can have only two modes, and that if a trough exists it must occur at $\hat{\beta} = 0$. Moreover, one can easily show that

$$\left.\frac{d^2}{d\hat{\beta}^2}\frac{{}_1F_1\left(\frac{k_2+1}{2};\frac{k_2}{2}+j;\frac{a}{2}\frac{\hat{\beta}^2}{1+\hat{\beta}^2}\right)}{\left(1+\hat{\beta}^2\right)^{\frac{k_2+1}{2}}}\right|_{\hat{\beta}=0}$$

$$= \frac{k_2+1}{4}\Gamma\left(\frac{k_2}{2}\right)\left(\frac{k_2}{2}\right)_j\left[-\frac{4}{\Gamma\left(\frac{k_2}{2}\right)\left(\frac{k_2}{2}\right)_j}+\frac{2a}{\Gamma\left(\frac{k_2}{2}+1\right)\left(\frac{k_2}{2}+1\right)_j}\right]$$

so that

$$\left.\frac{d^2\,\mathrm{pdf}\left(\hat{\beta}\right)}{d\hat{\beta}^2}\right|_{\hat{\beta}=0} = \frac{k_2+1}{4}\frac{\Gamma\left(\frac{k_2+1}{2}\right)\exp\left\{-\frac{a}{2}\right\}}{\pi^{\frac{1}{2}}\left(1+\hat{\beta}^2\right)^{\frac{k_2+1}{2}}}$$

$$\sum_{j=0}^{\infty}\frac{\left(\frac{k_2-1}{2}\right)_j}{j!}\left(\frac{a}{2}\right)^j\left[-\frac{4}{\Gamma\left(\frac{k_2}{2}\right)\left(\frac{k_2}{2}\right)_j}+\frac{2a}{\Gamma\left(\frac{k_2}{2}+1\right)\left(\frac{k_2}{2}+1\right)_j}\right].$$

After using (2.2.4) in Slater (1960) and simplifying one obtains

$$\left.\frac{d^2\,\mathrm{pdf}\left(\hat{\beta}\right)}{d\hat{\beta}^2}\right|_{\hat{\beta}=0} = -\frac{(k_2+1)\,\Gamma\left(\frac{k_2+1}{2}\right)\exp\left\{-\frac{a}{2}\right\}}{\Gamma\left(\frac{k_2}{2}+1\right)\pi^{\frac{1}{2}}\left(1+\hat{\beta}^2\right)^{\frac{k_2+1}{2}}}\,{}_1F_1\left(\frac{k_2-3}{2};\frac{k_2}{2};\frac{a}{2}\right)$$

and $d^2\,\mathrm{pdf}\left(\hat{\beta}\right)/d\hat{\beta}^2\Big|_{\hat{\beta}=0} > 0$ if and only if ${}_1F_1\left(\frac{k_2-3}{2};\frac{k_2}{2};\frac{a}{2}\right) < 0$. For $k_2 = 1$ one has ${}_1F_1\left(-1;\frac{1}{2};\frac{a}{2}\right) = 1-a < 0$ which implies $a > 1$. For $k_2 = 2$, ${}_1F_1\left(-\frac{1}{2};1;\frac{a}{2}\right) < 0$, implies $a > 3.15991$. For $k_2 \geq 3$, ${}_1F_1\left(\frac{k_2-3}{2};\frac{k_2}{2};\frac{a}{2}\right) \geq 1$ for all a so that $d^2\,\mathrm{pdf}\left(\hat{\beta}\right)/d\hat{\beta}^2\Big|_{\hat{\beta}=0} \leq 0$ for all a.

A.3 Proof of Theorem 3

To simplify notation let $\hat{\beta} = x$. By deriving equation (7) with respect to x, setting the derivative equal to zero and rearranging one obtains

$$\frac{\sum_{j=0}^{\infty} \frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}+1\right)_j} \left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(\frac{k_2+3}{2}; \frac{k_2+2}{2}+j; \frac{\mu^2}{2}\frac{(1+\beta x)^2}{1+x^2}\right)}{\sum_{j=0}^{\infty} \frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}\right)_j} \left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(\frac{k_2+1}{2}; \frac{k_2}{2}+j; \frac{\mu^2}{2}\frac{(1+\beta x)^2}{1+x^2}\right)} = \frac{k_2 x\left(1+x^2\right)}{\mu^2\left(\beta-x\right)\left(1+\beta x\right)}. \tag{14}$$

The left hand side of (14) is a non-negative, continuous function of x, having a minimum at $x = -1/\beta$, a maximum at $x = \beta$, and tending to

$$\frac{{}_1F_1\left(-\frac{k_2}{2}; \frac{k_2}{2}+1; -\frac{\mu^2\beta^2}{2}\right)}{{}_1F_1\left(-\frac{k_2}{2}; \frac{k_2}{2}; -\frac{\mu^2\beta^2}{2}\right)}$$

as $x \to \infty$. It has the same shape as $a(.)$ defined in equation (8). The right hand side of (14) can take on any real value as x ranges in the interval $(\min\{-1/\beta, \beta\}, \max\{-1/\beta, \beta\})$. As x goes to infinity the right hand side of (14) has an asymptote of the form

$$-\frac{k_2}{\beta\mu^2}x + \frac{k_2\left(1-\beta^2\right)}{\beta^2\mu^2},$$

and

$$\begin{aligned}\lim_{x\to\max\{-1/\beta,\beta\}_+} \frac{k_2 x\left(1+x^2\right)}{\mu^2\left(\beta-x\right)\left(1+\beta x\right)} &= -\operatorname{sign}(\beta)\,\infty \\ \lim_{x\to\min\{-1/\beta,\beta\}_-} \frac{k_2 x\left(1+x^2\right)}{\mu^2\left(\beta-x\right)\left(1+\beta x\right)} &= \operatorname{sign}(\beta)\,\infty.\end{aligned}$$

Moreover, if $\beta > 0$ the right hand side has a unique maximum in the range $(\beta, +\infty)$ (which stays below the horizontal axis) and a unique minimum in the range $(-\infty, -1/\beta)$ (which stays above the horizontal axis). It follows that there is one solution to equation (14) in the interval $[\min\{\beta, -1/\beta\}, \max\{\beta, -1/\beta\}]$, and, if there are three solutions two of them will be in the range $(-1/\beta, +\infty)$ if $\beta < 0$ and in $(-\infty, -1/\beta)$ if $\beta > 0$.

A.4 Proof of Corollary 1

Let $f(x)$ denote the left-hand-side of (13). Note that $f(x)$ has a minimum at $x = -1/\beta$ where $f^m = f(-1/\beta)$ equals

$$\begin{aligned} f^m &= \frac{{}_1F_1\left(\frac{k_2-1}{2}; \frac{k_2}{2}+1; \frac{\mu^2\beta^2}{2}\right)}{{}_1F_1\left(\frac{k_2-1}{2}; \frac{k_2}{2}; \frac{\mu^2\beta^2}{2}\right)} \\ &= 1 + O\left(k_2^{-1}\right) \end{aligned}$$

(where the last line follows from equation (4.3.6) of Slater (1960)) and a maximum at $x = \beta$ where $f^M = f(\beta)$ equals

$$\begin{aligned} f^M &= \frac{\sum_{j=0}^{\infty} \frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}+1\right)_j} \left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(\frac{k_2+3}{2}; \frac{k_2+2}{2}+j; \frac{\mu^2\left(1+\beta^2\right)}{2}\right)}{\sum_{j=0}^{\infty} \frac{\left(\frac{k_2-1}{2}\right)_j}{j!\left(\frac{k_2}{2}\right)_j} \left(\frac{\mu^2\beta^2}{2}\right)^j {}_1F_1\left(\frac{k_2+1}{2}; \frac{k_2}{2}+j; \frac{\mu^2}{2}\frac{\mu^2\left(1+\beta^2\right)}{2}\right)} \\ &= 1 + O\left(k_2^{-1}\right) \end{aligned}$$

(where again equation (4.3.6) of Slater (1960) has been used), so that when k_2 is large

$$1 + O\left(k_2^{-1}\right) = \frac{k_2 x\left(1+x^2\right)}{\mu^2(\beta - x)(1+\beta x)}.$$

Rearraning the expression in the last display one has

$$\frac{1 + O\left(k_2^{-1}\right)}{k_2} = \frac{x\left(1+x^2\right)}{\mu^2(\beta - x)(1+\beta x)},$$

and the theorem follows.

References

Forchini, G., and G. Hillier (2003): "Conditional Inference for Possibly Unidentified Structural Equations," *Econometric Theory*, 19, 707–743.

Hillier, G. H. (2004): "Yet More on The Exact Properties of IV Estimators," Mimeo, University of Southampton.

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.

——— (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. I, chap. 8, pp. 449–516. North-Holland, Amsterdam.

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

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

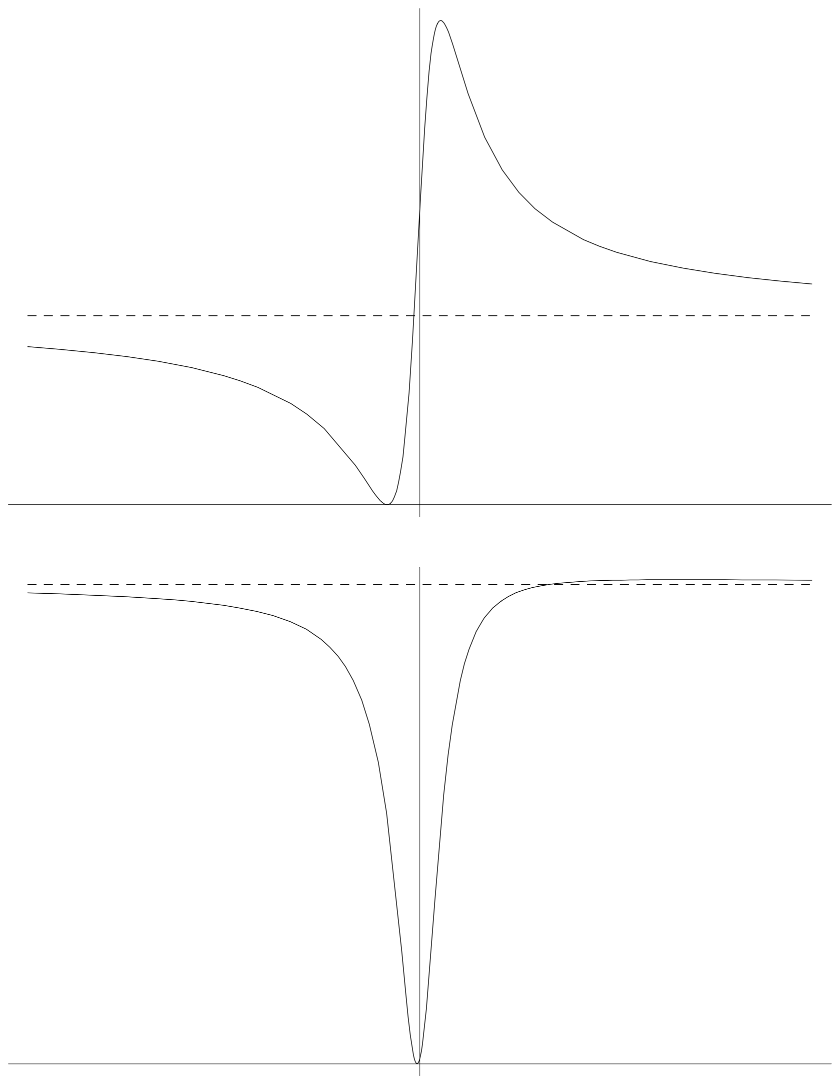

Figure 1: The top graph contains the typical shape of a $\left(\hat{\beta}\right)$ for β small. The bottom one illustrates a $\left(\hat{\beta}\right)$ when β is large.

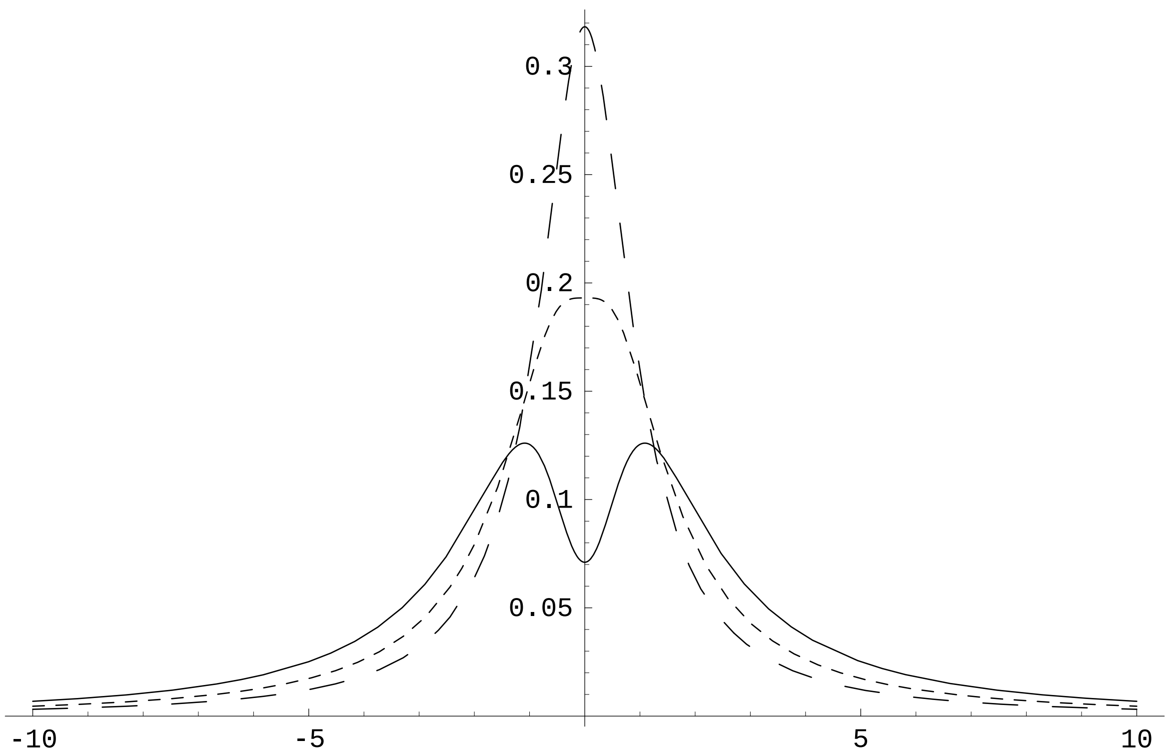

Figure 2: Graph of the limit densities in the just identified case for $a = 0$ (dashed line), $a = 1$ (dotted line) and $a = 3$ (solid line).