

# MEMORANDUM

No 03/2004

**Exact Small Sample Properties of the Instrumental  
Variable Estimator. A View From a Different Angle.**

*Halvor Mehlum*

ISSN: 0801-1117

---

Department of Economics  
University of Oslo



This series is published by the  
**University of Oslo**  
**Department of Economics**

P. O.Box 1095 Blindern  
N-0317 OSLO Norway  
Telephone: + 47 22855127  
Fax: + 47 22855035  
Internet: <http://www.oekonomi.uio.no/>  
e-mail: [econdep@econ.uio.no](mailto:econdep@econ.uio.no)

In co-operation with  
**The Frisch Centre for Economic  
Research**

Gaustadalleén 21  
N-0371 OSLO Norway  
Telephone: +47 22 95 88 20  
Fax: +47 22 95 88 25  
Internet: <http://www.frisch.uio.no/>  
e-mail: [frisch@frisch.uio.no](mailto:frisch@frisch.uio.no)

List of the last 10 Memoranda:

No 02	<i>Karine Nyborg and Kjetil Telle</i> A dissolving paradox: Firms' compliance to environmental regulation. 30 pp.
No 01	Taryn Ann Galloway To What Extent Is a Transition into Employment Associated with an Exit from Poverty?. 34 pp.
No 39	Atle Seierstad Piecewise deterministic optimal control problems. 86 pp.
No 38	Sheetal K. Chand Risk, Return, and the Optimal Design of a Pension System. 26 pp.
No 37	Rolf Aaberge, Ugo Colombino and Steinar Strøm Do More Equal Slices Shrink the Cake? An Empirical Investigation of Tax-Transfer Reform Proposals in Italy. 31 pp.
No 36	Kari Eika Low Quality-Effective Demand. 43 pp.
No 35	Karl Ove Moene Social Democracy as a Development Strategy. 35 pp.
No 34	Finn Roar Aune, Rolf Golombek and Sverre A. C. Kittelsen Does Increased Extraction of Natural Gas Reduce Carbon Emissions?. 40 pp.
No 33	Christoph Schwierz The Effects of Taxes and Socioeconomic Variables on Market Work and Home Production in Norway in the Years 1970 to 2000. 61 pp.
No 32	John K. Dagsvik, Steinar Strøm and Zhiyang Jia A Stochastic Model for the Utility of Income. 39 pp.

A complete list of this memo-series is available in a PDF® format at:  
<http://www.oekonomi.uio.no/memo/>

# Exact Small Sample Properties of the Instrumental Variable Estimator. A View From a Different Angle.

Halvor Mehlum\*

## Abstract

I derive the exact small sample properties of the instrumental variables estimator using a trigonometric approach. The distribution for the estimation error is decomposed into a product of three components - each with an intuitive interpretation. This approach helps the discussion on what underlies the exact shape of the estimator's distribution and in particular the possibility of a bimodal distribution.

*Keywords:* Instrument variable estimator

*JEL:* C30

## 1 Introduction

One of the core results in econometrics is that the instrumental variable estimator is consistent. For small samples and weak instruments, however, the distribution of the instrumental variable estimator may be both skewed and biased. The many contributions to the exploration of the small sample properties are authoritatively summarized by Phillips (1983). Hence, there should not be more to say about the issue. In 1990 however, Nelson and Startz points to the interesting result that the exact distribution of the instrumental variable estimator not only may be skewed and have mode off the target. It could even

---

\*Department of Economics, University of Oslo P.O. Box 1095, Blindern N-0317 Oslo, Norway. E-mail: halvor.mehlum@econ.uio.no. I want to thank Gunnar Bårdsen, Sheetal Chand, Jo Thori Lind, Ragnar Nymo, and Tore Schweder for useful comments.

be bimodal. Maddala and Jeong (1992), following up on Nelson and Startz, clarified what causes the bimodality. Later Woglom (2001) went in more detail using Marsaglia (1965) to analyze the issue. Several of these discussions supplement their calculations with intuitive reasoning based on the trigonometry of the problem. However, none of them uses trigonometric methods. Hence the discussion is not linked to the analysis and the intuitions for the results are therefore not entirely clear. In this paper I approach the problem head on using trigonometry and geometry and manage to decompose the density of the instrumental variable estimator into three factors, each with intuitive interpretations. The decomposition is the main contribution of the paper<sup>1</sup> and the relevance of the results beyond that of the instrument variable estimator. It can also be used when deriving the exact distribution of any ratio of correlated normal variables. Many such applications appear in econometrics. Examples are the estimation of long run parameters in dynamic models, and the estimation of equilibrium rate of unemployment NAIRU.

## 2 The Analysis

In this paper I follow Woglom's description of the problem. Consider a model with one structural and one reduced form equation:

$$\begin{aligned} Y &= \beta X + u \\ X &= \gamma Z + v \end{aligned} \tag{1}$$

Here  $Y$  and  $X$  are endogenous variables while  $Z$  is an exogenous deterministic variable. The error terms  $u$  and  $v$  are normally distributed with zero mean and a variances  $\sigma_u^2$  and  $\sigma_v^2$  and correlation coefficient  $\rho$ . The degree of endogeneity of  $X$  is measured by  $\rho^2$ . The quality of the instrument  $Z$  is measured by the squared correlation between  $Z$  and  $X$  denoted  $R^2 = \gamma^2 m_{ZZ} / \sigma_X^2$  (where  $m_{kl}$  is the empirical moment between  $k$  and  $l$ ). By using the instrument variable method to estimate the parameter of interest,  $\beta$ , we get the

---

<sup>1</sup>In the statistics literature the results in Nicholson (1941) overlaps part of my analysis. He follows, however, a different route. The decomposition result is to my knowledge entirely new.

estimator  $b$ . It can be shown that the estimation error is

$$b - \beta = w = \frac{m_{Zu}}{m_{ZX}} = \frac{m_{uZ}}{\gamma m_{ZZ} + m_{vZ}} = \frac{r_{uZ}}{\gamma + r_{vZ}}, \quad (2)$$

where  $r_{uZ} = m_{uZ}/m_{ZZ}$  and  $r_{vZ} = m_{vZ}/m_{ZZ}$  are regression coefficients such that  $r_{kZ} \sim N(0, \sigma_k^2/(m_{ZZ}T))$ . The correlation coefficient between  $r_{uZ}$  and  $r_{vZ}$  is  $\rho$ . Hence, the estimation error  $w$  is the fraction of two correlated bivariate normal distributed variables. For a large sample, large  $T$ , the distribution of  $w$  is bell shaped with zero mean. For small  $T$ , however, the probability distribution for  $w$  may have both mode and mean different from zero, it may be highly skewed. It may even have two modes.

In this paper I provide a new method for establishing the exact distribution for  $w$ . The method provides a short and direct link from the distribution of  $r_{vZ}, r_{uZ}$  to the distribution of  $w$ . I start from the observation that  $p \equiv (p_1, p_2) = (\gamma + r_{vZ}, r_{uZ})$  is a point in the plane with a bivariate normal probability distribution. Hence, for any  $p$ ,  $w$  is equal to the tangent to the angle  $\theta$  of the line going through the origin and  $p$ :  $\tan(\theta) = p_2/p_1$ . The first step is to derive the distribution of  $\theta$ . In other words: What is the probability of a  $p$  such that  $n \sin(\theta) = p_2$  and  $n \cos(\theta) = p_1$ ? (where  $n = \sqrt{p_1^2 + p_2^2}$  is the distance from  $p$  to origin.) Since  $p$  is bivariate normal with mean  $(\gamma, 0)$ , it follows from the standard formulas that

$$P_\theta(\theta) = \int_{-\infty}^{\infty} |n| \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{V}{2(1-\rho^2)}} dn \quad (3)$$

$$V \equiv \left[ \frac{(n \cos(\theta) - \gamma)^2}{\sigma_1^2} + \frac{(n \sin(\theta))^2}{\sigma_2^2} - \frac{2\rho(n \cos(\theta) - \gamma)n \sin(\theta)}{\sigma_1\sigma_2} \right]$$

where the variance of  $p_1$  is  $\sigma_1^2 = \sigma_v^2/(m_{ZZ}T)$ , the variance of  $p_2$  is  $\sigma_2^2 = \sigma_u^2/(m_{ZZ}T)$  and the correlation between  $p_1$  and  $p_2$  is  $\rho$ . For a given angle  $\theta \in [-90^\circ, 90^\circ]$ , the two dimensional normal density is integrated by letting the distance from the origin  $n$  run from  $-\infty$  to  $\infty$ . The  $|n|$  term captures that we are integrating over *sectors* of infinitesimal positive width where the width increases in proportion to  $|n|$ . Or to put it differently; as

$n$  increases the distance between the  $\theta$ -line and the  $(\theta + \Delta\theta)$ -line increases in proportion to  $n$ .

Figure 1: Deriving  $P_\theta(\theta)$ . Case 1 ( $\sigma_2^2 = \sigma_1^2 = 0.1$  and  $\rho = 0.7$ )

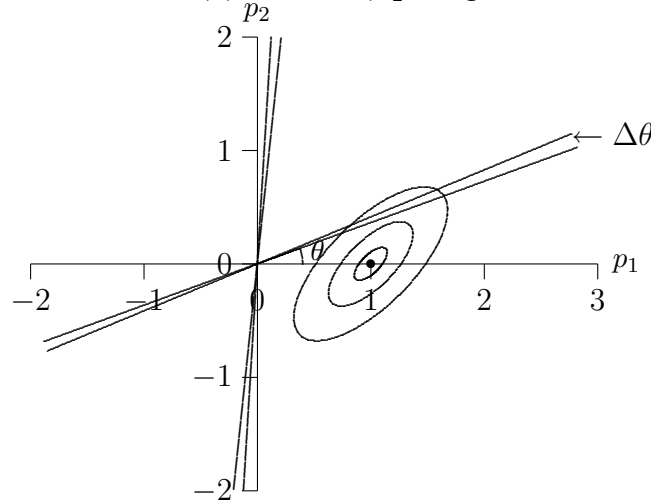


Figure 2: Deriving  $P_\theta(\theta)$ . Case 2 ( $\sigma_2^2 = \sigma_1^2 = 1$  and  $\rho = 0.95$ )

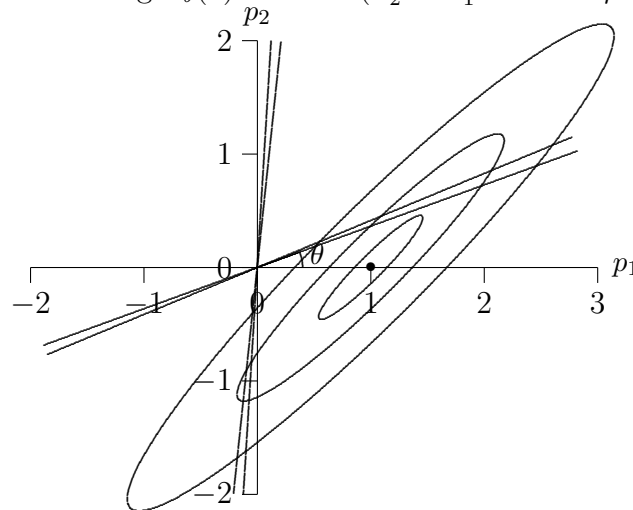


Figure 1 and Figure 2 gives two illustrations of how  $P_\theta(\theta)$  is derived. In both figures it is assumed that  $\gamma = 1$ , hence the distribution of  $p$  is a normal elliptic bell with centre in  $(1, 0)$ . The ellipses in the figures capture from the centre the 10%, the 50%, and the 90% fractiles. In case 1, in Figure 1, the variances and the correlation are modest ( $\sigma_1^2 = \sigma_2^2 = 0.1, \rho = 0.7$ ). In case 2, in Figure 2, the variances and the correlation are high ( $\sigma_1^2 = \sigma_2^2 = 1, \rho = 0.95$ ).  $P_\theta(\theta)$  is the probability for the point  $p$  to lay in the sector between the  $\theta$ -line and the  $(\theta + \Delta\theta)$ -line. The figures contains two different such  $\theta$  sectors

- one for  $\theta = 85^\circ$  and one for  $\theta = 20^\circ$ . A number of essential features can be noted already at this stage:

1. Positive correlation results in the mode of  $P_\theta(\theta)$  to be found for a positive angle  $\theta$ . The reason is that as the density is tilted to the right the parts of the density that lay above the horizontal axis are farther from the origin than the parts of the density that lay below. Hence, as the width of the sectors increases in distance from the origin, it follows that  $P'_\theta(0) > 0$ . (For negative correlation the opposite is true.)
2. When the density is far from the origin ( $\gamma$  high) the mode is close to zero. The reason is that the effect referred to above is less important for densities far to the right.
3. When the density is far from the origin ( $\gamma$  high)  $P_\theta(\theta)$  is peaked. The reason is that for densities far to the right a limited range of  $\theta$ 's covers the density.
4. When the correlation between  $p_1$  and  $p_2$  and their variances are high  $P_\theta(\theta)$  may be bimodal. Consider Figure 2. Most of the density is covered as  $\theta$  goes from  $-90^\circ$  to  $50^\circ$  giving a bell-shape with mode around  $10^\circ$ . However, as  $\theta$  reaches  $70^\circ$  the sector meets the density again - now in the third quadrant. As  $\theta$  goes from  $70^\circ$  to  $90^\circ$ ,  $P_\theta(\theta)$  will grow generation a second mode at  $90^\circ$ .

The density  $P_\theta(\theta)$  in (3) is quite easily calculated using any mathematics software. The main contribution of this paper, however, is to show that (3) can be decomposed into a product of two expressions based on univariate normal densities.

## 2.1 Decomposition

The decomposition builds on the idea that the density covered by a sector, is determined by 1) the density over a *line* in the direction and 2) by the precision in this direction. By precision I simply mean the probability mass' average distance from the origin. Borrowing terminology from the mechanics of the balance, the aim is to decompose the probability

$P_\theta(\theta)$  into a product of 1) weight  $W$  and 2) moment arm  $A$ . Hence

$$P_\theta(\theta) = W(\theta) A(\theta) \quad (4)$$

Such a decomposition is achieved as follows. Let  $\tilde{p} = (\tilde{p}_1, \tilde{p}_2)$ , corresponding to  $p = (p_1, p_2)$ , be a point in the plane that is rotated through the angle  $\theta$  relative to the  $p$ -plane. This linear transformation is described by the following formula

$$\begin{bmatrix} \tilde{p}_1 \\ \tilde{p}_2 \end{bmatrix} = \mathbf{B} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \quad (5)$$

The expectation of  $\tilde{p}$  is then

$$\mathbb{E} \begin{bmatrix} \tilde{p}_1 \\ \tilde{p}_2 \end{bmatrix} = \mathbf{B} \begin{bmatrix} \gamma \\ 0 \end{bmatrix} \equiv \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix} \quad (6)$$

while the covariance matrix is

$$\text{Cov} \begin{bmatrix} \tilde{p}_1 \\ \tilde{p}_2 \end{bmatrix} = \mathbf{B} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \mathbf{B}' \equiv \begin{bmatrix} \tilde{\sigma}_1^2 & \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 \\ \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2 \end{bmatrix} \quad (7)$$

This transformation keeps all angles and distances, hence  $\tilde{p}$  is distributed bivariate normal, with expectation and covariance given above. The following derivations are based on standard formulas for marginal and conditional distributions based on the bivariate normal.

The probability of a line having the angle  $\theta$  in the  $p$ -plane is equal to the probability that  $\tilde{p}_2$  is zero. Hence, the weight  $W$  is simply

$$W(\theta) = f\left(\frac{\tilde{\mu}_2}{\tilde{\sigma}_2}\right) \quad (8)$$

where  $f$  is the standard normal density. The precision, or length of arm, in this direction is



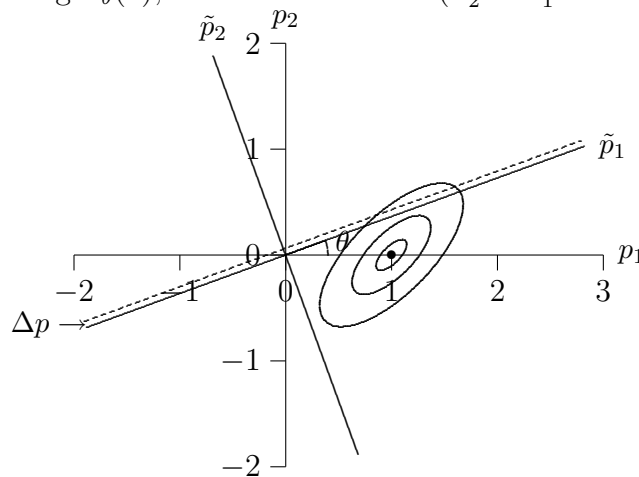
given by the expectation of  $|\tilde{p}_1|$  conditioned on  $\tilde{p}_2 = 0$ .<sup>2</sup> Using the formula for conditional distributions in combination with formulas for truncated distributions it follows that

$$\begin{aligned}
 A(\theta) &= E(|\tilde{p}_1| | \tilde{p}_2 = 0) = \tilde{\mu}_{1,0} \left( 1 - 2F \left( \frac{-\tilde{\mu}_{1,0}}{\tilde{\sigma}_{1,0}} \right) \right) + 2\tilde{\sigma}_{1,0} f \left( \frac{\tilde{\mu}_{1,0}}{\tilde{\sigma}_{1,0}} \right) \quad (9) \\
 \tilde{\mu}_{1,0} &= E(\tilde{p}_1 | \tilde{p}_2 = 0) = \tilde{\mu}_1 - \tilde{\mu}_2 \tilde{\rho} \tilde{\sigma}_1 / \tilde{\sigma}_2 \\
 \tilde{\sigma}_{1,0}^2 &= \text{Var}(\tilde{p}_1 | \tilde{p}_2 = 0) = (1 - \tilde{\rho}^2) \tilde{\mu}_1 \tilde{\sigma}_1^2
 \end{aligned}$$

where  $F$  is the cumulative standard normal distribution function and where the second 0 subscript indicates the conditioning on  $\tilde{p}_2 = 0$ . It can be verified, by using the definitions above, that (8) and (9) inserted in (4) indeed satisfies (3).

Figure 3 illustrates the linear transformation leading to the decomposition. The weight

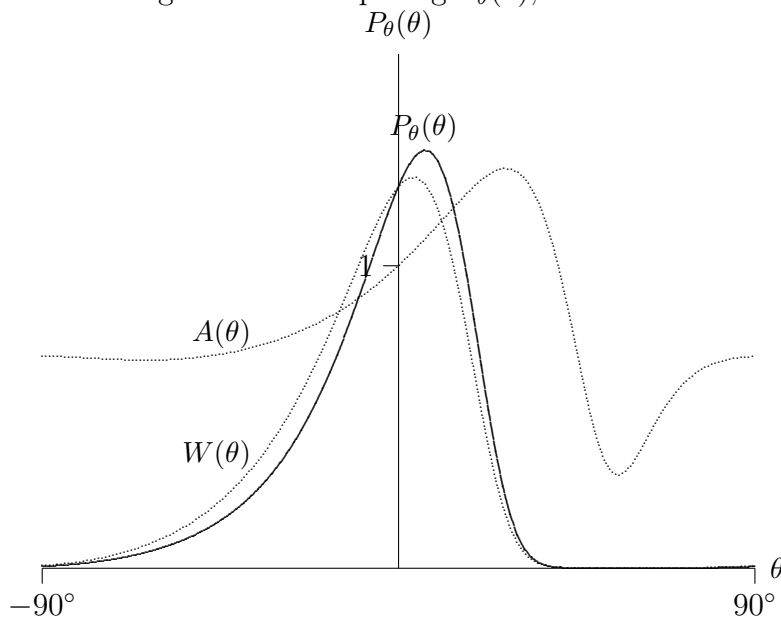
Figure 3: Deriving  $P_\theta(\theta)$ , modest correlation ( $\sigma_2^2 = \sigma_1^2 = 0.1$  and  $\rho = 0.7$ )



$W(\theta)$  is the probability of  $\tilde{p}_2 = 0$ , which is given by the probability mass covered by  $\tilde{p}_2 \in [0, \Delta p]$ . The arm  $A(\theta)$  is the average distance from the origin for the probability density covered by the  $\tilde{p}_2 = 0$  line. Visually, it is clear that for the illustrated  $\theta$  of  $20^\circ$   $A$  is around 1.3 while  $W$  is somewhat lower than its peak.

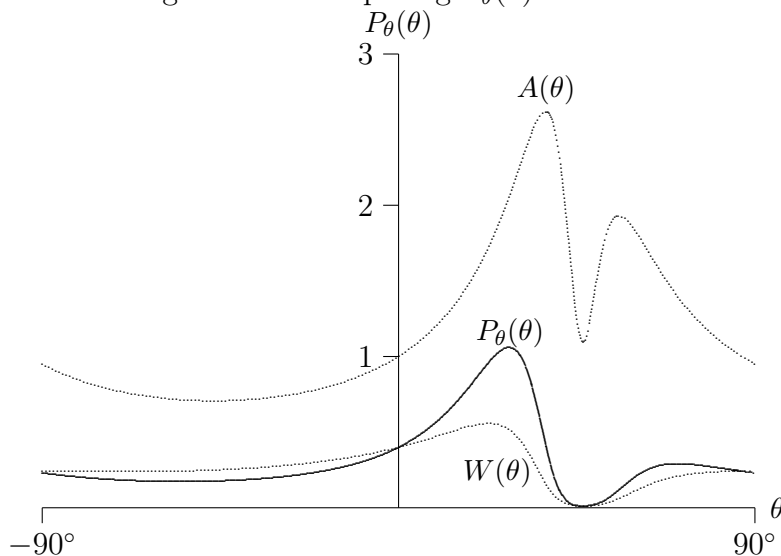
Figure 4 illustrates the decomposition in case 1. The solid line shows  $P_\theta(\theta)$ , while the dotted lines illustrates  $W$  and  $A$ .  $P_\theta(\theta)$  is bell-shaped and quite symmetric. It has

<sup>2</sup>Note that the balance analogy is not perfect. When using absolute value, mass in the negative direction do not balance weight in the positive direction as would be the case if a lever was resting on the origin.

Figure 4: Decomposing  $P_\theta(\theta)$ , case 1.

mode and mean to the right of zero. The shape is more or less determined by the weight function  $W$ . The reason for  $W$  being bell-shaped is the small variance. From Figure 1 it is seen that the reason for the mode of  $W$  being to the right of the origin is that lines that are tilted in the same direction as the density cover a larger part of the distribution than lines with an equal negative tilt. The bimodality of  $A$  follows from the fact that  $A$  has its minimum for the line that is parallel to the main axis of the density, i.e.  $\theta = 45^\circ$ . Note that the bimodality of  $A$  matters only marginally for the shape of  $P_\theta$ . The reason is that  $A$  has its second mode in a region where  $W$  is close to zero.

Case 2, as illustrated by Figure 5, is more complex. Here the weight-function  $W$  is not bell-shaped and it only goes close to zero for angles around  $45^\circ$ . Going back to Figure 2, it is clear that the reason for  $W$  having this shape is that there is significant probability mass in all but the second quadrant. As for  $A$  the minimum for  $W$  is found where the  $\theta$ -line is parallel to the main axis of the density, i.e.  $\theta = 45^\circ$ . For other angles the line covers non-negligible parts of the density. In case 2 both  $A$  and  $W$  have two modes and  $P_\theta(\theta)$  is far from bell-shaped. It also has two modes and does not go to zero. One should note, however, that the distinction between shapes under the present circumstances is somewhat arbitrary. The reason is that when we are analyzing angles going full circle the

Figure 5: Decomposing  $P_\theta(\theta)$  case 2.

distributions with necessity bites their own tail. The shape therefore depends on where you decide to break the circle and make it into a line. In the present case, however,  $\theta = 0$  is the natural centre, as that angle captures zero estimation error,  $w = \tan(\theta) = 0$ .

### 3 The last component

I have now decomposed  $P_\theta(\theta)$  into a product of  $W$  and  $A$ . The main question, however, relates to  $w = \tan(\theta)$ . The density function  $P_\theta(\theta)$  can easily be converted to a density over  $w$ ,  $P_w(w)$ , by multiplying by the Jacobian,  $(\partial \tan(\theta) / \partial \theta)^{-1} = (1 + \tan^2(\theta))^{-1} = \cos^2(\theta)$ . This compensates for the fact that for larger  $\theta$  each sector, between  $\theta$  and  $\theta + \Delta\theta$ , covers a larger range of tangents.<sup>3</sup> It follows that the density for the estimation error is given by

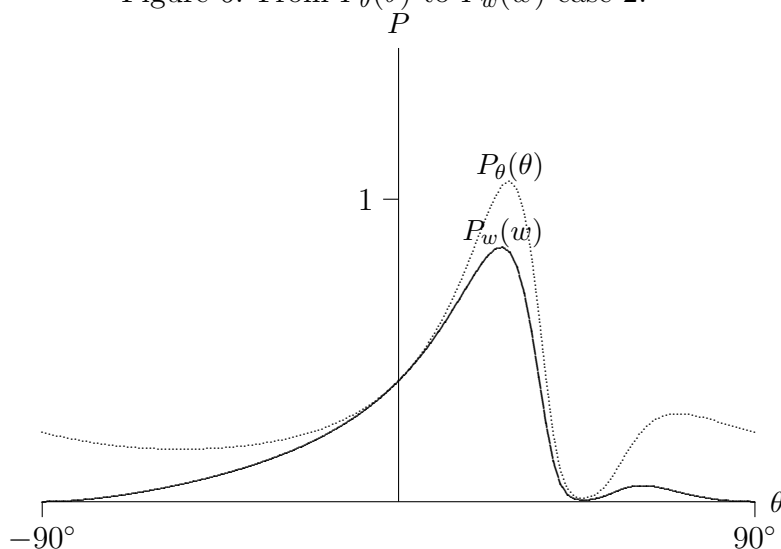
$$P_w(w) = \cos^2(\theta)^2 P_\theta(\theta) = \cos^2(\theta)^2 A(\theta) W(\theta), \text{ where } \theta = \arctan(w)$$

The third factor  $\cos^2(\theta)^2$  is bell-shaped, symmetric around the origin and it approaches zero as  $\theta \rightarrow \pm 90^\circ$ . The distribution for  $P_w(w)$  will always inherit the last feature. The question is under what conditions distinct features of  $P_\theta(\theta)$  - e.g. skewness and

<sup>3</sup>Note that  $\cos(\arctan(w)) = 1/\sqrt{1+w^2}$  is in the Cauchy family of distributions.

bimodality- survives after being multiplied by  $\cos(\theta)^2$ . The answer is simple for  $\theta$  close to zero. When  $\theta = 0$  then  $\cos(\theta)^2 = 1$  and the slope of  $\cos(\theta)^2$  is zero, hence both the level and the shape of  $P_\theta$  is transferred to  $P_w$ . Therefore, if the mode of  $P_\theta$  is to the right of zero so is the mode of  $P_w$ . In other words bias is retained. For  $\theta$  far from zero, however,  $\cos(\theta)^2$  is small and declines sharply with  $|\theta|$ , hence a second mode of  $P_\theta$  need to be peaked and steep and reasonably close to  $\theta = 0$  in order to show up also in  $P_w$ . Figure 6 illustrates<sup>4</sup> the transition from  $P_\theta(\theta)$  to  $P_w(\tan(\theta))$  by multiplying by  $\cos(\theta)^2$ .

Figure 6: From  $P_\theta(\theta)$  to  $P_w(w)$  case 2.



The figure reveals that the bias of  $P_\theta(\theta)$  in case 2 is indeed retained and so is the bimodal shape. The second mode of  $P_w(w)$  is not as high as it is for  $P_\theta(\theta)$ . It is, however, more distinct, in the sense that  $P_w$  goes to zero at each side of the second mode.

Geometrically, the reason for the survival of the bimodality is that  $P_\theta$  has a region where it is close to zero between the two modes.  $P_\theta$  is close to zero between its two modes if there is strong correlation between  $p_1$  and  $p_2$ . Whether the second mode is significant or not depends on whether the variances of  $p$  are high giving a significant probability mass in the third quadrant. To sum up: the condition for a significant second mode is strong endogeneity (giving high correlation) combined with weak instrument (giving high

<sup>4</sup>Note that  $P_w$  in this figure is drawn as a function of  $\theta$ ;  $P_w(\tan(\theta))$ .  $P_w$  is therefore not a probability density proper. For the qualitative discussion this does not matter much as  $w$  is a continuously increasing function of  $\theta$ .

variances).

## 4 Conclusion

I have provided a method to decompose the exact density of the instrument variable estimator. The results show that by trigonometric reasoning one may efficiently derive the distribution of the estimation error. The decomposition following from the trigonometry approach allows for intuitive interpretations of how the shape of the distribution is determined by instrument quality and endogeneity. I have employed the method to instrument variable estimation. It can be used whenever deriving the exact distribution of any ratio of correlated normal variables. Many such applications appear in econometrics. Examples are the estimation of long run parameters in AR models, the estimation of equilibrium rate of unemployment NAIRU, and the estimation of the monetary condition index MCI.<sup>5</sup>

## References

- Ericsson, N. R., N. A. Kerbeshian, R. Nymoen, and E.S. Jansen (2002) "Understanding an MCI", Mimeo University of Oslo.
- Haavelmo, Trygve (1947) "Methods of Measuring the Marginal Propensity to Consume" *Journal of the American Statistical Association*, 42, 105-122.
- Maddala, G.S., and J. Jeong (1992) "On the Exact Small Sample Distribution of the Instrumental Variable Estimator.", *Econometrica*, 60/1, 181-183.
- Marsaglia, G. (1965) "Ratios of Normal Variables and the Roots of Sums of Uniform Variables.", *Journal of the American Statistical Association*, 60, 193-204.
- Nelson, C.R., and R. Startz (1990) "Some Further Results on the Small Sample Properties of the Instrumental Variable Estimator." *Econometrica*, 58/4, 967-976.

---

<sup>5</sup>See for example Staiger, Stock, and Watson (1997) for the exact distribution for the NAIRU estimate and Ericsson et al. (2002) for the exact distribution of the MCI estimate.

Nicholson, C. (1941) “A Geometrical Analysis of the Frequency Distribution of the Ratio Between Two Variables”, *Biometrika*, 32/1, pp. 16-28.

Phillips, P.C.B. (1983) “Exact Small Sample Theory in the Simultaneous Equations Model.” in *Handbook of Econometrics, Vol II*, ed. Z. Griliches and M.D. Intriligator. Amsterdam: North Holland, 881-936.

Staiger, D, J. H. Stock and M. W. Watson (1997) “The NAIRU, Unemployment and Monetary Policy” *The Journal of Economic Perspectives*, 11/1 , pp. 33-49

Woglom, Geoffrey (2001) “More Results on the Exact Small Sample Properties of the Instrumental Variable Estimator.”, *Econometrica*, 69/5, 1381-89