The Limited Information Maximum Likelihood Estimator as an Angle *

T. W. Anderson †

Naoto Kunitomo ‡

and

Yukitoshi Matsushita §

March 27, 2009

Abstract

When an econometric structural equation includes two endogenous variables and their coefficients are normalized so that their sum of squares is 1, it is natural to express them as the sine and cosine of an angle. The Limited Information Maximum Likelihood (LIML) estimator of this angle when the error covariance matrix is known has constant variance. Of all estimators with constant variance the LIML estimator minimizes the variance. Competing estimators, such as the Two-Stage Least Squares estimator, has much larger variance for some values of the parameter. The effect of weak instruments is studied.

Key Words

Structural equation, linear functional relationship, natural normalization, angle LIML, finite sample properties.

*This is the version AKM09-3-27-2.

†Department of Statistics and Department of Economics, Stanford University

‡Graduate School of Economics, University of Tokyo

§Graduate School of Economics, University of Tokyo

1. Introduction

Anderson and Rubin (1949) developed the Limited Information Maximum Likelihood (LIML) estimator of the coefficients of a single structural equation in a simultaneous equations model. The "reduced form" of the model may be written

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V} \ , \tag{1.1}$$

where $\mathbf{Y}$ $(T \times G)$ consists of endogenous or dependent variables, $\mathbf{Z}$ $(T \times K)$ consists of exogenous or independent variables, $\boldsymbol{\Pi}$ $(K \times G)$ is a matrix of parameters and $\mathbf{V}$ $(T \times G)$ consists of unobserved disturbances. The rows of $\mathbf{V}$ are independently normally distributed with mean $\mathbf{0}$ and covariance $\boldsymbol{\Omega}$. A "structural equation" is

$$\mathbf{Y}\boldsymbol{\beta} = \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{u} \ , \tag{1.2}$$

where $\mathbf{Z}_1$ $(T \times K_1)$ consists of some K_1 columns of $\mathbf{Z}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_1$ are G and K_1 vectors, and $\mathbf{u} = \mathbf{V}\boldsymbol{\beta}$. In the simple case that $\boldsymbol{\gamma}_1 = \mathbf{0}$, the vector $\boldsymbol{\beta}$ $(G \times 1)$ satisfies

$$\boldsymbol{\Pi}\boldsymbol{\beta} = \mathbf{0} \ . \tag{1.3}$$

If $\boldsymbol{\Pi}$ is known and its rank is $G-1$, (1.3) determines $\boldsymbol{\beta}$ except for a multiplication by a constant. That constant can be determined by setting a specified coefficient of $\boldsymbol{\beta}$ equal to 1 (the *conventional normalization*). Alternatively, the *natural normalization* is

$$\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} = 1 \ . \tag{1.4}$$

To distinguish between solutions $\boldsymbol{\beta}$ and $-\boldsymbol{\beta}$, we can require one specified coordinate to be positive. Both normalizations were treated by Anderson and Rubin (1949). The natural normalization has the important advantage that the second-order moments are finite. [Anderson (2008).] Another advantage of this normalization is that parameters and maximum likelihood estimators transform simply under linear transformations of the observed variables.

If $G = 2$ and $\mathbf{\Omega} = \mathbf{I}_2$, the normalization (1.4) is

$$\boldsymbol{\beta}' \mathbf{\Omega} \boldsymbol{\beta} = \beta_1^2 + \beta_2^2 = 1 \ , \tag{1.5}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Thus $\boldsymbol{\beta}$ can be represented as

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \ , \tag{1.6}$$

where

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{\beta_2}{\beta_1} \ .$$

Then $\mathbf{\Pi}$ $(K \times 2)$ satisfying (1.3) can be represented as

$$\mathbf{\Pi} = \boldsymbol{\pi} \ (-\sin\theta, \cos\theta) \ , \tag{1.7}$$

where $\boldsymbol{\pi}$ is a $K \times 1$ vector. We shall call (1.6) the *angle form of the structural coefficient*. Note that $\mathrm{tr}\mathbf{\Pi}'\mathbf{\Pi} = \boldsymbol{\pi}'\boldsymbol{\pi}$.

If $G = 3$ and $\mathbf{\Omega} = \mathbf{I}_3$, the parameter $\boldsymbol{\beta}$ is a point on the unit sphere $\beta_1^2 + \beta_2^2 + \beta_3^2 = 1$ and can be represented in terms of latitude and longitude. [Anderson (2003), Section 2.7.]

The model consisting of (1.1), (1.3) and (1.4) with $\mathbf{\Omega} = \mathbf{I}_G$ is invariant with respect to orthogonal transformations. Let $\mathbf{\Phi}$ be an orthogonal matrix,

$$\mathbf{Y}^* = \mathbf{Y}\mathbf{\Phi} \ , \ \ \mathbf{V}^* = \mathbf{V}\mathbf{\Phi} \ , \ \ \mathbf{\Pi}^* = \mathbf{\Pi}\mathbf{\Phi} \ , \ \ \boldsymbol{\beta}^* = \mathbf{\Phi}'\boldsymbol{\beta} \ \ . \tag{1.8}$$

Then

$$\mathbf{Y}^* = \mathbf{Z}\mathbf{\Pi}^* + \mathbf{V}^* \ , \ \ \mathbf{\Pi}^*\boldsymbol{\beta}^* = \mathbf{0} \ , \ \ \boldsymbol{\beta}^{*'}\boldsymbol{\beta}^* = 1 \ \ , \tag{1.9}$$

which correspond to (1.1), (1.3) and (1.4). A particular orthogonal matrix is $\mathbf{\Phi} = -\mathbf{I}_G$; the model is invariant with respect to reflections.

When $G = 2$, $\mathbf{\Phi}$ can be written

$$\mathbf{\Phi} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \ . \tag{1.10}$$

Then

$$\boldsymbol{\beta}^* = \boldsymbol{\Phi}^{'}\boldsymbol{\beta} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} \cos(\theta-\phi) \\ \sin(\theta-\phi) \end{bmatrix} . \tag{1.11}$$

A rotation of $\boldsymbol{\beta}$ in the angular form corresponds to a translation of the angle.

If $\boldsymbol{\gamma}_1 = \mathbf{0}$ and $\beta_1 = 1$, the structural equation (1.2) is

$$\mathbf{y}_1 = -\beta_2\mathbf{y}_2 + \mathbf{u} , \tag{1.12}$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2)$ and the angle representation is written as

$$\beta_2 = \frac{\sin\theta}{\cos\theta} = \tan\theta . \tag{1.13}$$

On the other hand, if $\boldsymbol{\gamma}_1 = \mathbf{0}$ and $\beta_2 = 1$, the structural equation (1.2) is written as

$$\mathbf{y}_2 = -\beta_1\mathbf{y}_1 + \mathbf{u}^{'} , \tag{1.14}$$

and the angle representation is

$$\beta_1 = \cot\theta = \frac{1}{\tan\theta} . \tag{1.15}$$

In this paper we consider the distribution of the angle LIML estimator, which is invariant to orthogonal transformations, in a systematic way. On the other hand, the Two-Stage Least Squares (TSLS) estimator and the Generalized Method of Moments (GMM) estimator for coefficients, which have been often used in recent econometric studies, depend crucially on the particular normalization chosen in advance. (See Anderson and Sawa (1977).) This issue is related to the properties of alternative estimators and the optimality of estimation in structural equations. Some aspects of the related problems have been discussed by Anderson, Stein and Zaman (1985), Hiller (1990) and Chamberlain (2007). In particular we shall investigate the behavior of the adaptation of the LIML estimator to the case of $\boldsymbol{\Omega}$ known, that is, the LIMLK estimator, and compare it to the TSLS estimator in some detail. This simple case makes clearer some properties of these estimators.

A typical and simple application of simultaneous equations has been the econometric analysis of the demand-supply relation in markets. In such case we could take the log-quantity as $\mathbf{y}_1$ and the log-price as $\mathbf{y}_2$, and then we can estimate the price elasticity of the demand, for instance. The LIML estimate of the price elasticity is *invariant* whether we estimate it by using the demand function or *the inverse-demand function* as economists often do their empirical analysis interchangeably in real applications.

2. Inference when $G = 2$ and $\mathbf{\Omega} = \mathbf{I}_2$

The maximum likelihood estimator of $\mathbf{\Pi}$ when $\mathbf{\Pi}$ is unrestricted is

$$\mathbf{P} = \mathbf{A}^{-1}\mathbf{Z}'\mathbf{Y} \ , \quad \mathbf{A} = \mathbf{Z}'\mathbf{Z} \ . \tag{2.1}$$

Let

$$\mathbf{G} = \mathbf{P}'\mathbf{A}\mathbf{P} \ , \quad \bar{\mathbf{G}} = \frac{1}{T}\mathbf{G} = \mathbf{P}'\bar{\mathbf{A}}\mathbf{P} \ , \quad \bar{\mathbf{A}} = \frac{1}{T}\mathbf{A} \ . \tag{2.2}$$

Note that $\mathbf{P}$ is a sufficient statistic for $\mathbf{\Pi}$, which is the only parameter when $\mathbf{\Omega}$ is known.

Let d_1 and d_2 $(d_1 < d_2)$ be the roots of

$$0 = |\bar{\mathbf{G}} - d\mathbf{I}_2| = d^2 - (\bar{g}_{11} + \bar{g}_{22})d + \bar{g}_{11}\bar{g}_{22} - \bar{g}_{12}^2 \ , \tag{2.3}$$

where $\bar{\mathbf{G}} = (\bar{g}_{ij})$. The smaller root of (2.3) is

$$d_1 = \frac{1}{2}\left[\bar{g}_{11} + \bar{g}_{22} - \sqrt{(\bar{g}_{11} - \bar{g}_{22})^2 + 4\bar{g}_{12}^2}\right] \ . \tag{2.4}$$

The LIMLK estimator $\mathbf{b}$ of $\boldsymbol{\beta}$ with the natural normalization satisfies $1 = \mathbf{b}'\mathbf{b} = b_1^2 + b_2^2$ and

$$\begin{aligned} \mathbf{0} &= \left(\bar{\mathbf{G}} - d_1\mathbf{I}_2\right)\mathbf{b} \\ &= \begin{bmatrix} (\bar{g}_{11} - d_1)b_1 + \bar{g}_{12}b_2 \\ \bar{g}_{21}b_1 + (\bar{g}_{22} - d_1)b_2 \end{bmatrix} \ . \end{aligned} \tag{2.5}$$

The second component of (2.5) leads to the LIMLK estimator of θ as the solution of

$$\tan\hat{\theta} = \frac{b_2}{b_1} = -\frac{2\bar{g}_{21}}{\bar{g}_{22} - \bar{g}_{11} + \sqrt{(\bar{g}_{22} - \bar{g}_{11})^2 + 4\bar{g}_{12}^2}} \ . \tag{2.6}$$

Note that the denominator of (2.6) is positive.

The Two-Stage Least Squares (TSLS) estimator is defined by the second component of (2.5) with d_1 replaced by 0. The equation is

$$\bar{g}_{21}b_1 + \bar{g}_{22}b_2 = 0 \; . \tag{2.7}$$

The TSLS estimator of θ is defined by

$$\tan\hat{\theta}_{TS} = \frac{b_2}{b_1} = -\frac{\bar{g}_{21}}{\bar{g}_{22}} \; . \tag{2.8}$$

Note that the sign of $\tan\hat{\theta}_{TS}$ is the same as the sign of $\tan\hat{\theta}$ and

$$|\tan\hat{\theta}_{TS}| < |\tan\hat{\theta}| \; . \tag{2.9}$$

In a sense $\tan\hat{\theta}_{TS}$ is biased towards 0.

A transformation (1.8) effects a transformation of $\mathbf{P}$ to

$$\mathbf{P}^* = \mathbf{P}\boldsymbol{\Phi} \tag{2.10}$$

and $\mathbf{b}$, the solution to (2.5), to

$$\boldsymbol{\Phi}'\mathbf{b} = \mathbf{b}^* = \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} \; . \tag{2.11}$$

Then $\hat{\theta}$ defined by (2.5) is transformed to $\hat{\theta}^*$ defined by

$$\frac{b_2^*}{b_1^*} = \tan\hat{\theta}^* \; . \tag{2.12}$$

Thus $\hat{\theta}^* = \hat{\theta} - \phi$.

Theorem 1 : Let $\theta^* = \theta - \phi$, d_1 be the smaller root of (2.3), $\mathbf{b}$ the vector satisfying (2.5) and $\hat{\theta}$ the solution to (2.6). Let d_1^*, $\mathbf{b}^*$ and $\hat{\theta}^*$ be the root, vector and the angle for the model transformed by (1.8). Then $\hat{\theta} - \theta = \hat{\theta}^* - \theta^*$.

Note that the distribution of $\hat{\theta} - \theta$ is symmetric. In this sense $\hat{\theta}$ is an unbiased estimator of θ. Theorem 1 implies that the distribution of $\hat{\theta} - \theta$ is independent of θ.

The variance of the maximum likelihood estimator $\hat{\theta}$ is *invariant* under orthogonal transformations.

Consider the transformation

$$\mathbf{Z}^{+} = \mathbf{Z}\mathbf{M}\ ,\ \mathbf{\Pi}^{+} = \mathbf{M}^{-1}\mathbf{\Pi}\ , \tag{2.13}$$

where $\mathbf{M}$ is nonsingular, then (1.3) is transformed to

$$\mathbf{\Pi}^{+}\boldsymbol{\beta} = \mathbf{0}\ . \tag{2.14}$$

Furthermore, $\mathbf{A}$ transforms to $\mathbf{A}^{+} = \mathbf{M}'\mathbf{A}\mathbf{M}$, $\mathbf{P}$ transforms to $\mathbf{P}^{+} = \mathbf{M}^{-1}\mathbf{P}$ and $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{P}^{+'}\mathbf{A}^{+}\mathbf{P}^{+}$.

Corollary 1 : The LIMLK estimator with the natural normalization is invariant with respect to the transformation (2.13).

Anderson (2008) has shown that the normalization (1.4) implies that $\mathcal{E}(\mathbf{b}'\mathbf{b}) < \infty$.

Note that the LIMLK estimator of θ or of β_2/β_1 is unaffected by a scale trnsformation; that is, the LIMLK estimator of β_2/β_1 based on $\mathbf{G}$ is the same as the LIMLK estimator based on $c\mathbf{G}$, where $c > 0$. Similarly the TSLS estimator is unaffected by a scale transformation.

3. A Canonical Form

Make the transformation (2.13) so that $\bar{\mathbf{A}}^{+} = \mathbf{I}_K$, that is, $\mathbf{M}'\bar{\mathbf{A}}\mathbf{M} = \mathbf{I}_K$. Let $\mathbf{P}^{+} = \mathbf{M}^{-1}\mathbf{P} = \mathbf{Q}$ and

$$\mathbf{M}^{-1}\mathbf{P} = \mathbf{Q}\ ,\quad \mathbf{M}^{-1}\boldsymbol{\pi} = \boldsymbol{\eta}\ , \mathbf{M}^{-1}\mathbf{\Pi} = \boldsymbol{\eta}(-\sin\theta, \cos\theta)\ . \tag{3.1}$$

In the case of $G = 2$, the model is

$$\mathbf{Q} = \boldsymbol{\eta}\ (-\sin\theta, \cos\theta) + \mathbf{W}\ , \tag{3.2}$$

where

$$\mathbf{W} = \mathbf{M}^{-1}\mathbf{A}^{-1}\mathbf{Z}'\mathbf{V} = \mathbf{M}'\mathbf{Z}'\mathbf{V} \ , \tag{3.3}$$

and the rows of $\mathbf{W}$ are independently distributed according to $N(\mathbf{0}, \mathbf{I}_2)$. Note that $\bar{\mathbf{G}} = \mathbf{P}'\bar{\mathbf{A}}\mathbf{P} = \mathbf{Q}'\mathbf{Q}$ and

$$\mathbf{\Pi}'\bar{\mathbf{A}}\mathbf{\Pi} = \boldsymbol{\eta}'\boldsymbol{\eta} \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} (-\sin\theta, \cos\theta) \ . \tag{3.4}$$

The "canonical form" is closely related to the classical "errors-in-variables" model, also known as a "linear functional relationship" in statistics. Suppose $\mathbf{Q} = (q_{ki}), k = 1, \cdots, K; i = 1, 2$, $\mathcal{E}(q_{ki} - \mu_{ki}) = 0$, $\mathcal{E}(q_{ki} - \mu_{ki})^2 = \sigma^2$, $\mathcal{E}(q_{k1} - \mu_{k1})(q_{k2} - \mu_{k2}) = 0$. The pairs (μ_{k1}, μ_{k2}) are considered to lie on a straight line

$$\mu_{k2} = \alpha + \rho\mu_{k1} \ , \ k = 1, \cdots, K. \tag{3.5}$$

When $\alpha = 0$ and $\rho = \tan\theta$, this model corresponds to (3.2). Adcock (1878) recommended estimating α and ρ by minimizing

$$\sum_{k=1}^{K} \left[(q_{k1} - \mu_{k1})^2 + (q_{k2} - \mu_{k2})^2 \right] \tag{3.6}$$

subject to (3.5). This solution agrees with the LIMLK estimator. Anderson (1976) pointed out this relationship between the LIMLK estimator and the maximum likelihood estimator of the slope in the linear functional relationships. See also Anderson (1982) and Anderson and Sawa (1982).

In the linear functional relationship in terms of (3.5) it is natural to describe the line by the angle it makes with the first coordinate axis and the intercept on one of the coordinate axis. The estimator of this angle was studied by Anderson (1976).

4. Admissibility of the LIMLK estimator

Anderson, Stein and Zaman (1985) have considered the best invariant estimator of a direction. The model corresponds to (3.2) with $\boldsymbol{\eta}'\boldsymbol{\eta} = \lambda^2$. The loss function is

$$
\begin{aligned}
(4.1) \qquad L(\boldsymbol{\beta}, \lambda^2; \hat{\boldsymbol{\beta}}) &= 1 - (\boldsymbol{\beta}'\hat{\boldsymbol{\beta}})^2 \\
&= 1 - \left[(\cos\theta, \sin\theta) \begin{pmatrix} \cos\hat{\theta} \\ \sin\hat{\theta} \end{pmatrix} \right]^2 \\
&= \sin^2(\hat{\theta} - \theta) \ .
\end{aligned}
$$

Lemma 1 : Let $\boldsymbol{\beta}^* = \boldsymbol{\Phi}'\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}^* = \boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}$ as in (1.11). Then

$$
(4.2) \qquad L(\boldsymbol{\beta}^*, \lambda^2; \hat{\boldsymbol{\beta}}^*) = L(\boldsymbol{\beta}, \lambda^2; \hat{\boldsymbol{\beta}}) \ .
$$

Let $\boldsymbol{\beta}^- = -\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}^- = -\hat{\boldsymbol{\beta}}$. Then

$$
(4.3) \qquad L(\boldsymbol{\beta}^-, \lambda^2; \hat{\boldsymbol{\beta}}^-) = L(\boldsymbol{\beta}, \lambda^2; \hat{\boldsymbol{\beta}}) \ .
$$

Proof : The left-hand side of (4.3) is 1 minus

$$
(4.4) \qquad \boldsymbol{\beta}^{*\prime}\hat{\boldsymbol{\beta}}^* = \boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}'\hat{\boldsymbol{\beta}} \ .
$$

Q.E.D.

The loss function $L(\boldsymbol{\beta}, \lambda^2; \hat{\boldsymbol{\beta}})$ is invariant under rotations and reflections.

Lemma 2 : Under the transformation (1.8), $\mathcal{E}(\hat{\theta} - \theta)^2$ is invariant.

Anderson, Stein and Zaman (1985) showed that of all invariant procedures the LIMLK estimator has minimum loss.

By using the Taylor series expansion of $\sin x = x - x^3/6 + x^5/120 - \cdots$, we obtain an approximation

$$
(4.5) \qquad \mathcal{E}\sin^2(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \mathcal{E}(\hat{\theta} - \theta)^2 - \frac{1}{3}\mathcal{E}(\hat{\theta} - \theta)^4 + \frac{2}{45}\mathcal{E}(\hat{\theta} - \theta)^6 + \cdots \ .
$$

5. Empirical Distributions

We write (3.2) as

$$q_{k1} = -\eta_k \sin\theta + w_{k1}\ , \quad q_{k2} = \eta_k \cos\theta + w_{k2}\ , \tag{5.1}$$

where the random variables $\{w_{kj}\}\ ,\ k = 1, \cdots, K,\ j = 1, 2$, are $N(0,1)$. Define

$$\lambda^2 = \boldsymbol{\eta}^{'}\boldsymbol{\eta}\ , \quad \boldsymbol{\alpha}^{'} = (-\sin\theta, \cos\theta)\ . \tag{5.2}$$

The matrix $\mathbf{Q}^{'}\mathbf{Q}$ has a noncentral Wishart distribution with covariance matrix $\mathbf{I}_2$, noncentrality matrix $\boldsymbol{\eta}^{'}\boldsymbol{\eta}\boldsymbol{\alpha}\boldsymbol{\alpha}^{'}$ and K degrees of freedom [Anderson and Girshick (1944)]. The exact distributions of $\mathbf{b}$ and $\hat{\theta}$ have been given by Anderson and Sawa (1982), but they are very complicated.

The asymptotic expansions of the distributions of the angle LIMLK and TSLS estimators ($\lambda \to \infty$, K fixed) are

$$P(\lambda(\hat{\theta}_{LIMLK} - \theta) \le \xi) = \Phi(\xi) - \frac{1}{2\lambda^2}[(K-1)\xi + \frac{1}{3}\xi^3]\phi(\xi) + O(\lambda^{-4}), \tag{5.3}$$

and

$$\begin{aligned} P(\lambda(\hat{\theta}_{TS} - \theta) \le \xi) &= \Phi(\xi) + \frac{K-1}{\lambda}\tan\theta\phi(\xi) \\ &- \frac{1}{2\lambda^2}[\{(K-1)^2\tan^2\theta - (K-1)\}\xi + \frac{1}{3}\xi^3]\phi(\xi) + O(\lambda^{-3}), \end{aligned} \tag{5.4}$$

respectively (Anderson (1976)).

For symmetric intervals, the expansions are

$$\begin{aligned} &P(\lambda(\hat{\theta}_{LIMLK} - \theta) \le \xi) - P(\lambda(\hat{\theta}_{LIMLK} - \theta) \le -\xi) \\ &= \Phi(\xi) - \Phi(-\xi) - \frac{1}{\lambda^2}\left[(K-1)\xi + \frac{1}{3}\xi^3\right]\phi(\xi) + O(\lambda^{-4}), \end{aligned} \tag{5.5}$$

and

$$\begin{aligned} &P(\lambda(\hat{\theta}_{TS} - \theta) \le \xi) - P(\lambda(\hat{\theta}_{TS} - \theta) \le -\xi) \\ &= \Phi(\xi) - \Phi(-\xi) - \frac{1}{\lambda^2}\left[\left\{(K-1)^2\tan^2\theta - (K-1)\right\}\xi + \frac{1}{3}\xi^3\right]\phi(\xi) + O(\lambda^{-3})\ . \end{aligned} \tag{5.6}$$

Then

$$\begin{aligned} &P(\lambda|\hat{\theta}_{LIMLK} - \theta| \le \xi) - P(\lambda|\hat{\theta}_{TS} - \theta| \le \xi) \\ &= \frac{1}{\lambda^2}(K-1)\left[(K-1)\tan^2\theta - 2\right]\phi(\xi) + O(\lambda^{-3})\ . \end{aligned} \tag{5.7}$$

The LIMLK estimator dominates the TSLS estimator unless $\tan^2\theta \le 2/(K-1)$.

The expansion (5.3) corresponds to an approximate density

$$\phi(\xi) - \frac{1}{2\lambda^2}\left[K - 1 - (K-2)\xi^2 - \frac{1}{3}\xi^4\right]\phi(\xi) \tag{5.8}$$

and the variance of (5.8) is

$$1 - \frac{1}{2\lambda^2}\left[(K-1) - 3(K-2) - \frac{15}{3}\right] = 1 + \frac{K}{\lambda^2}\,. \tag{5.9}$$

Then the approximate variance of $\hat\theta_{LIMLK} - \theta$ is

$$\frac{1}{\lambda^2} + \frac{K}{\lambda^4}\,. \tag{5.10}$$

The expansion (5.4) corresponds to an approximate density

$$\begin{aligned}(5.11)\quad &\phi(\xi) - \frac{K-1}{\lambda}\tan\theta\xi\phi(\xi)\\ &- \frac{1}{2\lambda^2}\left[(K-1)^2\tan^2\theta - (K-1) + \left\{-(K-1)^2\tan^2\theta + K\right\}\xi^2 - \frac{1}{3}\xi^4\right]\phi(\xi)\,.\end{aligned}$$

The integral of (5.11) times ξ^2, which is an approximate MSE of $\lambda(\hat\theta_{TS} - \theta)$, is

$$1 + \frac{1}{\lambda^2}\left[(K-1)^2\tan^2\theta - K + 2\right]\,. \tag{5.12}$$

Note that the approximate MSE increases with θ from $1 - (K-2)/\lambda^2$ to ∞. The approximate MSE of the TSLS estimator is less than the approximate MSE of the LIMLK estimator if $\tan^2\theta < 2/(K-1)$. For $K = 3$ the inequality is $\tan\theta < 1$ ($\theta < \pi/4$) and for $K = 30$ $\tan\theta < \sqrt{2/29}$ $(= 0.069)$. We give some numerical values of the appromimate MSE in Tables 1 and 2.

The empirical distributions of the (standardized) angle LIMLK and TSLS estimators, $-\frac{\pi}{2} < \hat\theta - \theta < \frac{\pi}{2}$,

$$P(\lambda(\hat\theta - \theta) \le \xi), \tag{5.13}$$

and the asymptotic expansions (5.3) and (5.4) are compared. See Figures 1 to 18 in the appendix.

For each K, λ and θ, 10,000 data sets were obtained. The estimate of the loss $E[\sin^2(\hat\theta - \theta)]$, $\hat E[\sin^2(\hat\theta - \theta)] = \frac{1}{10000}\sum_{j=1}^{10000}\sin^2(\hat\theta_j - \theta)$, and the estimate of the

Table 1: Approximate MSE of $\hat{\theta} - \theta$

	$K = 3$					
	$\lambda^2 = 100$		$\lambda^2 = 50$		$\lambda^2 = 10$	
θ	LIMLK	TSLS	LIMLK	TSLS	LIMLK	TSLS
0.4π	0.0103		0.0212		0.13	
0.2π	0.0103	0.0101	0.0212	0.0204	0.13	0.1111
0	0.0103	0.0099	0.0212	0.0196	0.13	0.0900

Table 2: Approximate MSE of $\hat{\theta} - \theta$

	$K = 30$			
	$\lambda^2 = 100$		$\lambda^2 = 50$	$\lambda^2 = 10$
θ	LIMLK	TSLS	LIMLK	LIMLK
0.4π	0.013		0.032	0.4
0.2π	0.013		0.032	0.4
0	0.013	0.0072	0.032	0.4

Table 3: $\mathcal{E} sin^2(\hat{\theta} - \theta)$

	$K = 3$					
	$\lambda^2 = 100$		$\lambda^2 = 50$		$\lambda^2 = 10$	
θ	LIMLK	TSLS	LIMLK	TSLS	LIMLK	TSLS
0.4π	0.0102	0.0175	0.0206	0.0619	0.1269	0.3580
0.2π	0.0102	0.0100	0.0210	0.0202	0.1276	0.1039
0	0.0102	0.0098	0.0205	0.0187	0.1273	0.0798

Table 4: $\mathcal{E}sin^2(\hat{\theta}-\theta)$

	$K=30$					
	$\lambda^2=100$		$\lambda^2=50$		$\lambda^2=10$	
θ	LIMLK	TSLS	LIMLK	TSLS	LIMLK	TSLS
0.4π	0.0130	0.3503	0.0340	0.5752	0.2760	0.8184
0.2π	0.0130	0.0353	0.0338	0.0806	0.2693	0.2456
0	0.0131	0.0077	0.0335	0.0123	0.2701	0.0244

Table 5: $\mathcal{E}(\hat{\theta}-\theta)^2$

	$K=3$					
	$\lambda^2=100$		$\lambda^2=50$		$\lambda^2=10$	
θ	LIMLK	TSLS	LIMLK	TSLS	LIMLK	TSLS
0.4π	0.0103	0.0191	0.0210	0.0762	0.1561	0.5382
0.2π	0.0103	0.0101	0.0214	0.0206	0.1584	0.1224
0	0.0103	0.0099	0.0209	0.0191	0.1572	0.0879

Table 6: $\mathcal{E}(\hat{\theta}-\theta)^2$

	$K=30$					
	$\lambda^2=100$		$\lambda^2=50$		$\lambda^2=10$	
θ	LIMLK	TSLS	LIMLK	TSLS	LIMLK	TSLS
0.4π	0.0131	0.4214	0.0356	0.8024	0.4055	1.3909
0.2π	0.0136	0.0361	0.0355	0.0844	0.3902	0.2797
0	0.0133	0.0077	0.0354	0.0125	0.3927	0.0250

MSE, $\frac{1}{10000}\sum_{j=1}^{10000}(\hat{\theta}_j - \theta)^2$ were calculated. (See Tables 3 to 6.) The empirical cdf of an estimator is within 0.02 of the true cdf everywhere with probability more than 0.99 on the basis of 10,000 replications by using the Kolmogorov-Smirnov statistic. (See Anderson, Kunitomo and Sawa (1982).)

Comments on the empirical distributions of the LIMLK and TSLS estimators when $\mathbf{\Omega} = \mathbf{I}_2$

1. The empirical values of $E[\sin^2(\hat{\theta} - \theta)]$ and $E[(\hat{\theta} - \theta)^2]$ for the LIMLK estimator are invariant with respect to θ for each pair of value of λ and K.

2. The empirical values of $E[\sin^2(\hat{\theta} - \theta)]$ and $E[(\hat{\theta} - \theta)^2]$ for the TSLS estimator at each pair of values of λ and K increase with θ from a value less than the measure at $\theta = 0$ to a value considerably greater than the measure at $\theta = \pi/2$.

3. Since $\lambda(\hat{\theta}_{LIMLK} - \theta)$ has a limiting distribution $N(0,1)$, the variance of $\hat{\theta}_{LIMLK}$ is approximately $1/\lambda^2$. In Tables 5 and 6 the empirical variance of the LIMLK estimator is approximately .01 for $\lambda^2 = 100$ and .02 for $\lambda^2 = 50$, but is approximately .16 for $\lambda^2 = 10$ and $K = 3$ and .39 for $\lambda^2 = 10$ and $K = 30$. For smaller values of λ^2 the approximation $1/\lambda^2$ underestimates the variance.

4. The estimate of $E[\sin^2(\hat{\theta} - \theta)]$ is less than the estimate of $E[(\hat{\theta} - \theta)^2]$ for $\lambda^2 = 10$ for every θ and K.

5. At $\lambda^2 = 100$ and $\lambda^2 = 50$, the distributions of $\lambda(\hat{\theta} - \theta)$ for the LIMLK estimator are almost exactly $N(0,1)$.

6. More General Models

6.1 Arbitrary Variance

If the covariance matrix $\mathbf{\Omega}$ of the disturbances is $\sigma^2\mathbf{I}_2$, the noncentrality parameter is

$$\lambda^2 = \frac{\boldsymbol{\eta}'\boldsymbol{\eta}}{\sigma^2} \,. \tag{6.1}$$

6.2 Arbitrary Intercept

If $\boldsymbol{\gamma}_1 \neq \mathbf{0}$, we partition $\mathbf{Z}$ and $\boldsymbol{\Pi}$ as

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) \ , \quad \boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \end{bmatrix} \ , \tag{6.2}$$

where $\mathbf{Z}_1$ has K_1 columns, $\mathbf{Z}_2$ has K_2 columns $(K_1 + K_2 = K)$, $\boldsymbol{\Pi}_1$ has K_1 rows, and $\boldsymbol{\Pi}_2$ has K_2 rows. Then (1.3) is replaced by

$$\begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} \ . \tag{6.3}$$

The second part of (6.3) determines $\boldsymbol{\beta}$. Let

$$\mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \ , \quad \mathbf{Z}_{2.1}' = \mathbf{Z}_2' - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{Z}_1' \ . \tag{6.4}$$

Then

$$\mathbf{P}_2 = \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y} \ , \quad \mathbf{G} = \mathbf{P}_2'\mathbf{A}_{22.1}\mathbf{P}_2 \ . \tag{6.5}$$

Then the proceeding analysis applies with $\boldsymbol{\Pi}$ replaced by $\boldsymbol{\Pi}_2$.

6.3 General $\boldsymbol{\Omega}$

When $\boldsymbol{\Omega}$ is known, the LIMLK estimator of $\boldsymbol{\beta}$ in the natural parameterization is defined by

$$\left(\mathbf{P}'\bar{\mathbf{A}}\mathbf{P} - d_1\,\boldsymbol{\Omega}\right)\mathbf{b} = \mathbf{0} \ , \quad \mathbf{b}'\boldsymbol{\Omega}\mathbf{b} = 1 \ , \tag{6.6}$$

where d_1 is the smallest root of

$$|\mathbf{P}'\bar{\mathbf{A}}\mathbf{P} - d\,\boldsymbol{\Omega}| = 0 \ . \tag{6.7}$$

This case can be reduced to the special case of $\mathbf{\Omega} = \mathbf{I}_G$. Write $\mathbf{\Omega} = \mathbf{\Phi}^{'}\mathbf{\Delta}\mathbf{\Phi}$, where $\mathbf{\Delta}$ is diagonal and $\mathbf{\Phi}$ is orthogonal and

$$(6.8) \qquad \mathbf{\Omega}^{1/2} = \mathbf{\Phi}^{'}\mathbf{\Delta}^{1/2}\mathbf{\Phi}$$

is the *symmetric square root* of $\mathbf{\Omega}$. Then (6.6) can be written

$$(6.9) \qquad \left(\mathbf{P}^{'}\bar{\mathbf{A}}\mathbf{P} - d_1\,(\mathbf{\Omega}^{1/2})^2\right)\mathbf{b} = \mathbf{0}\ ,\ \ \mathbf{b}^{'}(\mathbf{\Omega}^{1/2})^2\mathbf{b} = 1\ ,$$

which leads to

$$(6.10) \qquad \left(\mathbf{P}^{*'}\bar{\mathbf{A}}\mathbf{P}^* - d_1\,\mathbf{I}_G\right)\mathbf{b}^* = \mathbf{0}\ ,\ \ \mathbf{b}^{*'}\mathbf{b}^* = 1\ ,$$

where

$$(6.11) \qquad \mathbf{P}^* = \mathbf{P}\mathbf{\Omega}^{-1/2}\ ,\ \ \mathbf{\Omega}^{1/2}\mathbf{b}^* = \mathbf{b}\ .$$

Let

$$(6.12) \qquad \mathbf{\Pi}^* = \mathbf{\Pi}\mathbf{\Omega}^{-1/2}\ ,\ \ \boldsymbol{\beta}^* = \mathbf{\Omega}^{1/2}\boldsymbol{\beta}\ .$$

Then

$$(6.13) \qquad \mathbf{\Pi}^*\boldsymbol{\beta}^* = \mathbf{0}\ ,\ \ \boldsymbol{\beta}^{*'}\boldsymbol{\beta}^* = 1\ .$$

When $G = 2$,

$$(6.14) \qquad \boldsymbol{\beta}^* = \begin{bmatrix} \cos\theta^* \\ \sin\theta^* \end{bmatrix}\ ,\ \ \mathbf{b}^* = \begin{bmatrix} \cos\hat{\theta}^* \\ \sin\hat{\theta}^* \end{bmatrix}\ .$$

The case of $\mathbf{\Omega}$ known, not necessarily $\mathbf{I}_G$, can be reduced to the case of $\mathbf{\Omega} = \mathbf{I}_G$.

7. LIML when $\mathbf{\Omega}$ is unknown

This model consisting of (1.1), (1.3) and (1.4) is invariant with respect to nonsingular transformations

$$(7.1) \qquad \mathbf{Y}^* = \mathbf{Y}\mathbf{C}\ ,\ \ \mathbf{V}^* = \mathbf{V}\mathbf{C}\ ,\ \ \mathbf{\Omega}^* = \mathbf{C}^{'}\mathbf{\Omega}\mathbf{C},\ \ \mathbf{\Pi}^* = \mathbf{\Pi}\mathbf{C}\ ,\ \ \boldsymbol{\beta}^* = \mathbf{C}^{-1}\boldsymbol{\beta}\ ;$$

that is

$$\mathbf{\Pi}^*\boldsymbol{\beta}^* = \mathbf{0}, \;\; \boldsymbol{\beta}^{*'}\mathbf{\Omega}^*\boldsymbol{\beta}^* = 1 \; . \tag{7.2}$$

When $\mathbf{\Omega}$ is unknown, let $\mathbf{b}$ be the solution of

$$(\bar{\mathbf{G}} - d_1\, \bar{\mathbf{H}})\mathbf{b} = \mathbf{0}, \;\; \mathbf{b}'\bar{\mathbf{H}}\mathbf{b} = 1, \tag{7.3}$$

and d_1 is the smallest root of

$$|\bar{\mathbf{G}} - d\, \bar{\mathbf{H}}| = 0, \tag{7.4}$$

$\bar{\mathbf{G}} = (1/T)\mathbf{G}$ and

$$\mathbf{H} = \mathbf{Y}'\mathbf{Y} - \mathbf{G}, \;\; \bar{\mathbf{H}} = \frac{1}{T}\mathbf{H} \; . \tag{7.5}$$

Then $\hat{\mathbf{\Omega}} = (1/T)\mathbf{H} + d_1(1/T)\mathbf{H}\mathbf{b}\mathbf{b}'(1/T)\mathbf{H}$ and the LIML estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \frac{1}{\sqrt{1+d_1}}\mathbf{b} \; . \tag{7.6}$$

The transformation (7.1) effects the transformation

$$\mathbf{P}^* = \mathbf{PC} \; , \;\; \mathbf{G}^* = \mathbf{C}'\mathbf{GC}, \;\; \mathbf{H}^* = \mathbf{C}'\mathbf{HC}, \;\; \mathbf{b}^* = \mathbf{C}^{-1}\mathbf{b} \; . \tag{7.7}$$

The transformed estimator of $\boldsymbol{\beta}$ satisfies

$$(\bar{\mathbf{G}}^* - d_1\bar{\mathbf{H}}^*)\mathbf{b}^* = \mathbf{0}, \;\; \mathbf{b}^{*'}\bar{\mathbf{H}}\mathbf{b}^* = 1, \tag{7.8}$$

In this sense the LIML estimator is invariant with respect to nonsingular linear transformations.

Let $\mathbf{\Omega}^{1/2}$ be the symmetric square root of $\mathbf{\Omega}$ defined by (6.8). Then $\mathbf{\Pi}^* = \mathbf{\Pi}\mathbf{\Omega}^{-1/2}$ and $\boldsymbol{\beta}^* = \mathbf{\Omega}^{1/2}\boldsymbol{\beta}$ satisfy (7.2). When $G = 2$, we can define

$$\boldsymbol{\beta}^* = \begin{bmatrix} \cos\theta^* \\ \sin\theta^* \end{bmatrix} \; . \tag{7.9}$$

Let $\bar{\mathbf{H}}^{1/2}$ be the symmetric square root of $\bar{\mathbf{H}}$. Then $\hat{\boldsymbol{\beta}}^* = \bar{\mathbf{H}}^{1/2}\mathbf{b}$ satisfies

$$(\bar{\mathbf{H}}^{-1/2}\bar{\mathbf{G}}\bar{\mathbf{H}}^{-1/2} - d_1\, \mathbf{I}_2)\hat{\boldsymbol{\beta}}^* = \mathbf{0}, \;\; \hat{\boldsymbol{\beta}}^{*'}\hat{\boldsymbol{\beta}}^* = 1, \tag{7.10}$$

Define $\hat{\theta}^*$ by

$$\hat{\boldsymbol{\beta}}^* = \begin{bmatrix} \cos\hat{\theta}^* \\ \sin\hat{\theta}^* \end{bmatrix} . \tag{7.11}$$

Then $\hat{\theta}^*$ is the maximum likelihood estimator of θ^*.

As $T \to \infty$, $\bar{\mathbf{G}} \xrightarrow{p} \mathbf{\Pi}'(\lim_{T\to\infty} \bar{\mathbf{A}})\mathbf{\Pi}$, $\bar{\mathbf{H}} \xrightarrow{p} \mathbf{\Omega}$, $d_1 \xrightarrow{p} 0$, $\bar{\mathbf{b}} \xrightarrow{p} \boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. Hence $\hat{\boldsymbol{\theta}}^* \xrightarrow{p} \boldsymbol{\theta}^*$. However, $\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*$ does not have an invariant distribution for fixed T because the transformation $\mathbf{C}$ that carries $\boldsymbol{\beta}$ to $\boldsymbol{\beta}^*$ is not the same as the transformation $\mathbf{C}$ that carries $\hat{\boldsymbol{\beta}}$ to $\hat{\boldsymbol{\beta}}^*$.

8. The Noncentral Wishart Distribution

The (central) Wishart distribution of $\mathbf{G}$ with the covariance matrix $\mathbf{\Omega} = \mathbf{I}_2$, $\mathbf{\Pi} = \mathbf{O}$, and K degrees of freedom is

$$w_2(\mathbf{G}|\mathbf{I}_2, K) = \frac{|\mathbf{G}|^{(K-3)/2} e^{-\frac{1}{2}\mathrm{tr}\,\mathbf{G}}}{2^K \pi^{1/2}\Gamma[K/2]\Gamma[(K-1)/2)]} . \tag{8.1}$$

The matrix $\mathbf{G}$ can be represented as

$$\mathbf{G} = \mathbf{ORO}' , \tag{8.2}$$

where $\mathbf{R}$ is diagonal with diagonal elements r_1 and r_2 ($0 \le r_1 \le r_2 < \infty$) and $\mathbf{O}$ is orthogonal. The orthogonal matrix $\mathbf{O}$ can be written as

$$\mathbf{O} = \begin{bmatrix} \cos\, t & -\sin\, t \\ \sin\, t & \cos\, t \end{bmatrix} . \tag{8.3}$$

The diagonal elements of $\mathbf{R}$ are the roots of $|\mathbf{G} - r\mathbf{I}_2| = 0$. The Jacobian of the transformation (8.2) of (g_{11}, g_{12}, g_{22}) to (r_1, r_2, t) is $r_2 - r_1$. (See Chapter 13 of Anderson (2003).) Also $\mathrm{tr}(\mathbf{G}) = r_1 + r_2$ and $|\mathbf{G}| = r_1 r_2$. The density of r_1, r_2 and t is

$$\frac{(r_1 r_2)^{(K-3)/2} e^{-\frac{1}{2}(r_1+r_2)}(r_2 - r_1)}{2^K \pi^{1/2}\Gamma[K/2]\Gamma[(K-1)/2)]} . \tag{8.4}$$

Note that (r_1, r_2) and t are independent. The distribution of $\mathbf{O}$ is the same as the distribution of $\mathbf{O\Phi}$, where $\mathbf{\Phi}$ is the orthogonal matrix defined by (1.10). It follows that the distribution of t is the same as the distribution of $t - \phi$; the distribution of t is uniform on $(0, 2\pi)$. Since we identify t with $-t$ and hence with $2\pi - t$, the density of t is $1/\pi$ on the interval $[-\pi/2, \pi/2]$.

Let

$$\mathbf{G} = \mathbf{Q}'\mathbf{Q} = (\boldsymbol{\eta}\boldsymbol{\alpha}' + \mathbf{W})'(\boldsymbol{\eta}\boldsymbol{\alpha}' + \mathbf{W}) \ . \tag{8.5}$$

The density of $\mathbf{G}$ [Anderson and Girshick (1944)] is

$$w_2(\mathbf{G}|\boldsymbol{\eta}'\boldsymbol{\eta}\boldsymbol{\alpha}\boldsymbol{\alpha}', \mathbf{I}_2, K) = \frac{e^{-\frac{1}{2}\boldsymbol{\eta}'\boldsymbol{\eta}-\frac{1}{2}\mathrm{tr}\,\mathbf{G}}|\mathbf{G}|^{\frac{K-3}{2}}}{2^K\pi^{1/2}\Gamma(\frac{K-1}{2})}\sum_{j=0}^{\infty}\frac{(\boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha}\ \boldsymbol{\eta}'\boldsymbol{\eta})^j}{2^{2j}j!\ \Gamma(\frac{K}{2}+j)} \ , \tag{8.6}$$

where $\boldsymbol{\alpha}' = (-\sin\theta, \cos\theta)$ and

$$\boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha} = g_{11}\sin^2\theta - 2g_{12}\sin\theta\cos\theta + g_{22}\cos^2\theta \ . \tag{8.7}$$

Note that for $\boldsymbol{\eta}'\boldsymbol{\eta} = 0$, (8.6) is the same as (8.1).

Let $\mathbf{G} = \mathbf{ORO}'$ as in (8.2). Then

$$\boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha} = \boldsymbol{\alpha}'\mathbf{ORO}'\boldsymbol{\alpha} \tag{8.8}$$

and

$$\mathbf{O}'\boldsymbol{\alpha} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} = \begin{bmatrix} \sin(t-\theta) \\ \cos(t-\theta) \end{bmatrix} \ , \tag{8.9}$$

which is $\boldsymbol{\alpha}$ with t replaced by $t - \theta$. Thus

$$\begin{aligned} \boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha} &= [\sin(t-\theta), \cos(t-\theta)] \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} \sin(t-\theta) \\ \cos(t-\theta) \end{bmatrix} \\ &= r_1\sin^2(t-\theta) + r_2\cos^2(t-\theta) \\ &= r_1 + (r_2 - r_1)\cos^2(t-\theta) \ . \end{aligned} \tag{8.10}$$

The density of r_1, r_2 and t is

$$(8.11)\quad f(r_1, r_2, t|\boldsymbol{\eta}'\boldsymbol{\eta}, \theta) = \frac{1}{2^K \pi^{1/2}\Gamma(\frac{K-1}{2})} e^{-\frac{1}{2}\boldsymbol{\eta}'\boldsymbol{\eta}-\frac{1}{2}(r_1+r_2)}(r_1 r_2)^{\frac{1}{2}(K-3)}(r_2 - r_1) \times \sum_{j=0}^{\infty} \frac{[r_1 \sin^2(t-\theta) + r_2 \cos^2(t-\theta)]^j (\boldsymbol{\eta}'\boldsymbol{\eta})^j}{2^{2j} j!\, \Gamma(j + \frac{1}{2}K)}$$
$$= \frac{1}{2^K \pi^{1/2}\Gamma(\frac{K-1}{2})} e^{-\frac{1}{2}\lambda^2-\frac{1}{2}(r_1+r_2)}(r_1 r_2)^{\frac{1}{2}(K-3)}(r_2 - r_1) \times \sum_{j=0}^{\infty} \frac{[r_1 + (r_2 - r_1)\cos^2(t-\theta)]^j (\lambda^2)^j}{2^{2j} j!\, \Gamma(j + \frac{1}{2}K)} .$$

Then (8.4) in r_1, r_2, t is $f(r_1, r_2, t|0, 0)$. The marginal density of t is obtained by integrating (8.11) over $0 \le r_1 \le r_2 < \infty$. Note that (8.11) is a convergent series.

For each r_1 and r_2 ($r_1 \le r_2$) (8.10) is a decreasing function of $t - \theta$ for $0 \le t - \theta \le \pi/2$; its maximum occurs at $t - \theta = 0$. If (8.11) is considered as the likelihood function of the parameter θ, it is maximized at $\theta = t$ (defined in (8.4)). Thus $\hat{\theta} = t$ is the maximum likelihood estimator of θ.

In the general case with $G > 2$, it has been known that the distribution of the LIML estimator for coefficients has a complicated form. The finite sample properties of the LIML estimator for coefficients have been explored by Phillips (1984, 1985), for instance.

9. Weak Instruments

Consider the case of LIML when $\boldsymbol{\Omega} = \mathbf{I}_G$. The LIML estimator of $\boldsymbol{\beta}$ when normalized by $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ is defined by

$$(9.1)\quad (\bar{\mathbf{G}} - d_1\,\bar{\mathbf{H}})\hat{\boldsymbol{\beta}} = \mathbf{0}, \quad \hat{\boldsymbol{\beta}}'\bar{\mathbf{H}}\hat{\boldsymbol{\beta}} = 1 .$$

Let $\boldsymbol{\Pi} = \boldsymbol{\Pi}_T$, $\bar{\mathbf{A}}_T = \bar{\mathbf{A}}$, and

$$(9.2)\quad \lambda_T^2 = \mathrm{tr}\ \boldsymbol{\Pi}_T'\bar{\mathbf{A}}_T\boldsymbol{\Pi}_T .$$

Suppose $\bar{\mathbf{A}}_T \to \mathbf{A}_\infty$ and $\lambda_T^2 \to 0$. Then the instruments (or exogenous variables) are called *weak*. Suppose that as $T \to \infty$, $\lambda_T^2 \to 0$. Then $\bar{\mathbf{H}} \xrightarrow{p} \boldsymbol{\Omega} = \mathbf{I}_G$, $\boldsymbol{\Pi}_T \to \mathbf{O}$, and the distribution of $\mathbf{G}$ approaches the Wishart distribution with covariance matrix $\mathbf{\Omega} = \mathbf{I}_G$ and K degrees of freedom. When $G = 2$, the density of the limiting distribution of the matrix $\mathbf{G}$ is (8.1).

Theorem 2 : When $\mathbf{\Omega} = \mathbf{I}_2$, $T \to \infty$ and $\lambda_T^2 \to 0$, the limiting distribution of $\hat{\theta}_{LIML} - \theta$ is the uniform distribution on $[-\pi/2 - \theta, \pi/2 - \theta]$.

We can write (8.11) as

$$(9.3)\ f(r_1, r_2, t|\lambda^2, \theta)$$
$$= \quad f(r_1, r_2, t|0, 0)e^{-\lambda^2/2}\left[1 + \frac{\Gamma(\frac{1}{2}K)}{4\Gamma(\frac{1}{2}K + 1)}\left[r_1 + (r_2 - r_1)\cos^2(t - \theta)\right]\lambda^2\right] + O(\lambda^4)\ .$$

Then by integrating out d_1 and d_2

$$(9.4) \qquad f(t|\lambda^2, \theta) = e^{-\lambda^2/2}\left\{\frac{1}{\pi} + \left[C_1(K) + C_2(K)\cos^2(t - \theta)\right]\lambda^2\right\} + O(\lambda^4)\ ,$$

where

$$\begin{aligned} C_1(K) &= \int_{0<r_1<r_2} f(r_1, r_2, t|0, 0) r_1 \frac{1}{2K} dr_1 dr_2\ , \\ C_2(K) &= \int_{0<r_1<r_2} f(r_1, r_2, t|0, 0)(r_2 - r_1)\frac{1}{2K} dr_1 dr_2\ . \end{aligned}$$

This is an asymptotic expansion of the density of the angle $\hat{\theta}$ as $\lambda^2 \to 0$. If more terms in the density (8.11) were used, the asymptotic expansion would have a smaller error.

10. Conclusions

1. The fact that $\mathcal{E}(\hat{\theta}_{LIMLK} - \theta)^2$ is independent of θ makes conparison with other estimators clearer.
2. The results of Anderson, Stein and Zaman (1985) show that the LIMLK estimator is the best invariant estimator when $\mathbf{\Omega} = \mathbf{I}$ or $\mathbf{\Omega} = \sigma^2\mathbf{I}$.
3. The asymptotic expansions of the LIMLK and TSLS estimators show that the

measn square error of the LIMLK estimator is smaller than that of the TSLS estimator unless θ is very small; $\tan^2\theta \leq 2/(K-1)$. [For $K=3$, the inequality is $\theta \leq \pi/4$.]

4. For given K and λ^2 the quantity $\mathcal{E}(\hat{\theta}-\theta)^2$ for TSLS increases with θ in terms of the asymptotic expansions and simulations. At $\theta = 0$ ($\beta_2 = 0$) $\mathcal{E}(\hat{\theta}-\theta)^2$ is less for TSLS than for LIMLK, but for larger θ $\mathcal{E}(\hat{\theta}-\theta)^2$ is much larger for TSLS than for LIMLK.

References

[1] Adcock, R. J. (1878), "A problem in least squares," *The Analyst*, 5, 53-54.

[2] Anderson, T.W. (1976), "Estimation of linear functional relationships : Approximate distributions and connections with simultanoeous equations in econometrics," *Journal of Royal Statistical Society*, Vol. 36, 1-36.

[3] Anderson, T.W. (1982), "Some recent developments on the distributions of single equation estimators," *Advances in Econometrics: Invited Paper for the Fourth World Congress of the Econometric Society at Aix-en-Provance*, September 1980, Werner Hildenbrand ed, Cambridge University Press, 109-122.

[4] Anderson, T.W. (2008), "The LIML estimator has finite moments!," Unpublished Manuscript.

[5] Anderson, T.W., and M.A. Girshick (1944), "Some extensions of the Wishart distribution," *Annals of Mathematical Statistics*, Vol. 15, 345-357. (Corrections : Vol.35 (1964),923-925).

[6] Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, Vol. 50, 1009-1027.

[7] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2008), "On Finite Sample Properties of Alternative Estimators of Coefficients in a Struc-

tural Equation with Many Instruments," Discussion Paper CIRJE-F-577, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2008), forthcoming in *Journal of Econometrics.*

[8] Anderson, T.W., and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

[9] Anderson, T.W., and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[10] Anderson, T.W., and Takamitsu Sawa (1977), "Two-Stage Least Squares : In Which Direction Should the Residuals Be Minimized?," *Journal of the Americam Statistical Association*, Vol. 72, 187-191.

[11] Anderson, T.W., and Takamitsu Sawa (1982), "Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient," *Journal of Royal Statistical Society*, **B**, Vol. 44, 52-62.

[12] Anderson, T.W., C. Stein and A. Zaman (1985), "Best Invariant Estimation of a Direction Parameter," *Annals of Statistics*, Vol. 13, 526-533.

[13] Chamberlain, G. (2007), "Decision Theory Applied to an Instrumental Variables Model," *Econometrica*, Vol. 75-3, 609-652.

[14] Hillier, Grant H. (1990), "On the normalization of structural equations : properties of direction estimators," *Econometrica*, Vol. 58, 1181-1194.

[15] Phillips, P.C.B. (1984), "The exact distribution of LIML: I," *International Economic Review*, Vol.25-1, 249-261.

[16] Phillips, P.C.B. (1985), "The exact distribution of LIML: II," *International Economic Review*, Vol.26-1, 21-36.

Appendix

This appendix gives the exact and approximate distributions of the LIMLK and TSLS estimators based on simulations in the normalized form of $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$. The empirical values of $\hat{E}[\lambda(\hat{\theta}-\theta)]^2$ are calculated in the normalized form for $\lambda(\hat{\theta}-\theta)$.

The method of simulations are similar to the one used by Anderson, Kunitomo and Matsushita (2008) and we have enough accuracy. (i) We first generate 10,000 data sets by using the two-equations system of $y_{1t} = -\mathbf{z}_t'\boldsymbol{\pi} sin\theta + v_{1t}$ and $y_{2t} = \mathbf{z}_t'\boldsymbol{\pi} cos\theta + v_{2t}$ for $t = 1, \ldots, T$ where $\mathbf{z}_t \sim N(0, I_K)$, $(v_{1t}, v_{2t}) \sim N(\mathbf{0}, \mathbf{I}_2)$, and (v_{1t}, v_{2t}) are independent of $\mathbf{z}_t$. We set $K \times 1$ vector $\boldsymbol{\pi} = c(1, \ldots, 1)'$ so that $\lambda^2 = \boldsymbol{\pi}'\mathbf{A}\boldsymbol{\pi}$ and $\mathbf{A} = \sum_{t=1}^{T} \mathbf{z}_t\mathbf{z}_t'$. (ii) Then the distributions of the LIMLK and TSLS estimators of $tan\theta$ are computed by (2.6) and (2.8), respectively. (iii) Finally, we have the distributions of the angle LIMLK and TSLS estimators by transforming $tan(\hat{\theta})$ to $\hat{\theta}$ so that $-\frac{\pi}{2} < \hat{\theta} - \theta < \frac{\pi}{2}$ in each case.

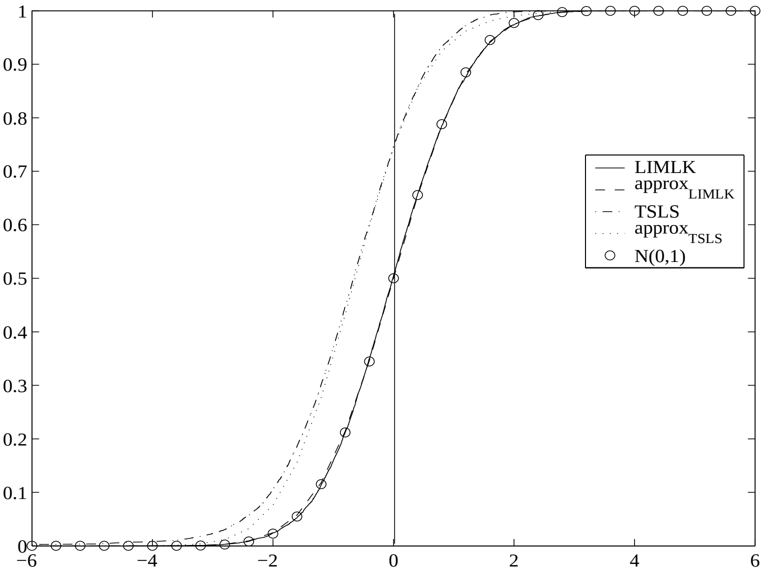

Figure 1: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=3,\ \theta=0.4\pi$ and $\lambda^2=100$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.03, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 1.91$

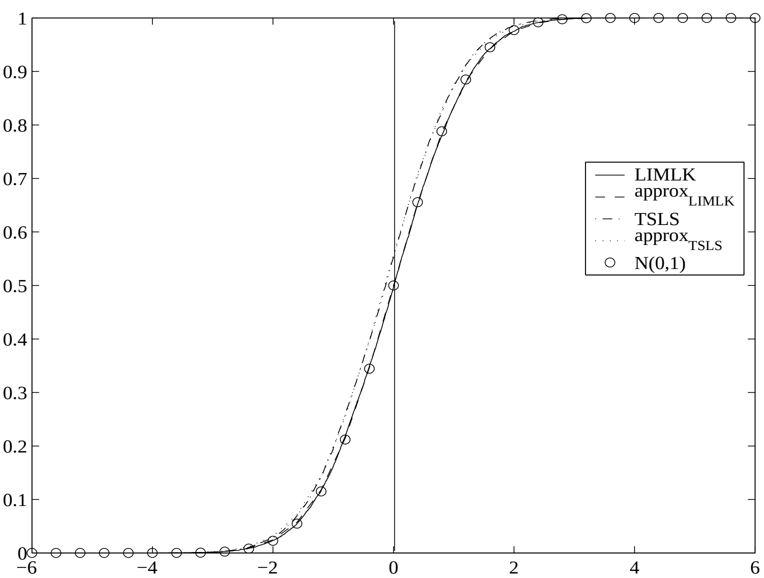

Figure 2: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=3,\ \theta=0.2\pi$ and $\lambda^2=100$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.03, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 1.01$

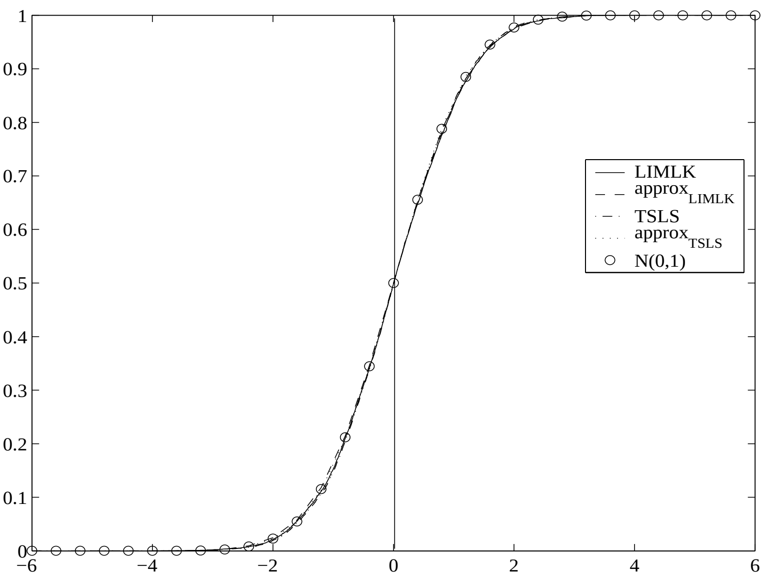

Figure 3: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K = 3$, $\theta = 0$ and $\lambda^2 = 100$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.03$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 0.99$

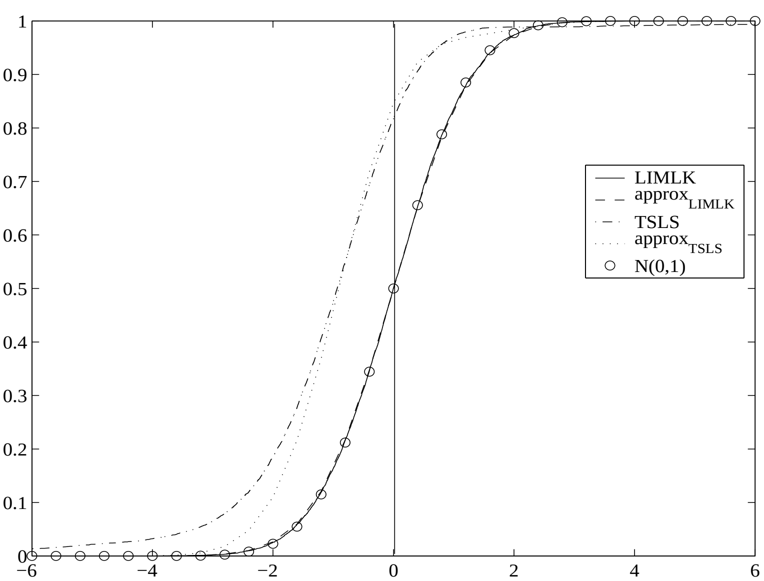

Figure 4: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K = 3$, $\theta = 0.4\pi$ and $\lambda^2 = 50$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.05$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 3.81$

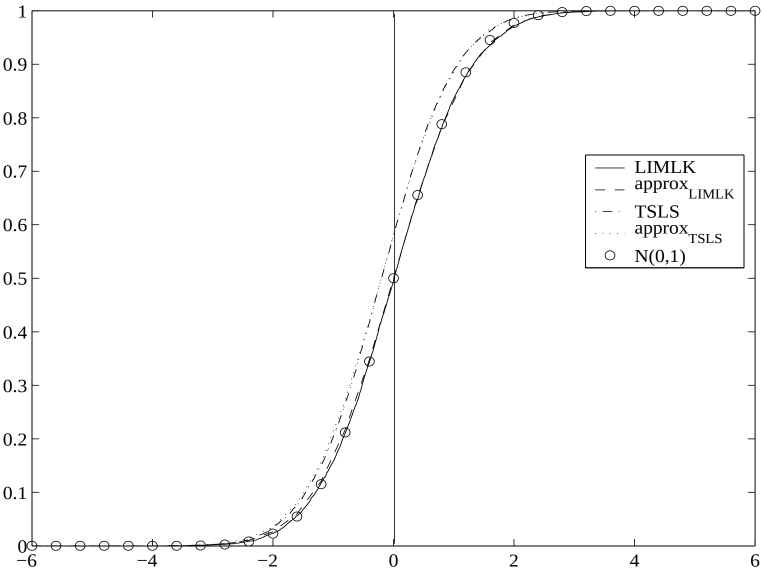

Figure 5: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=3$, $\theta = 0.2\pi$ and $\lambda^2 = 50$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.07$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 1.03$

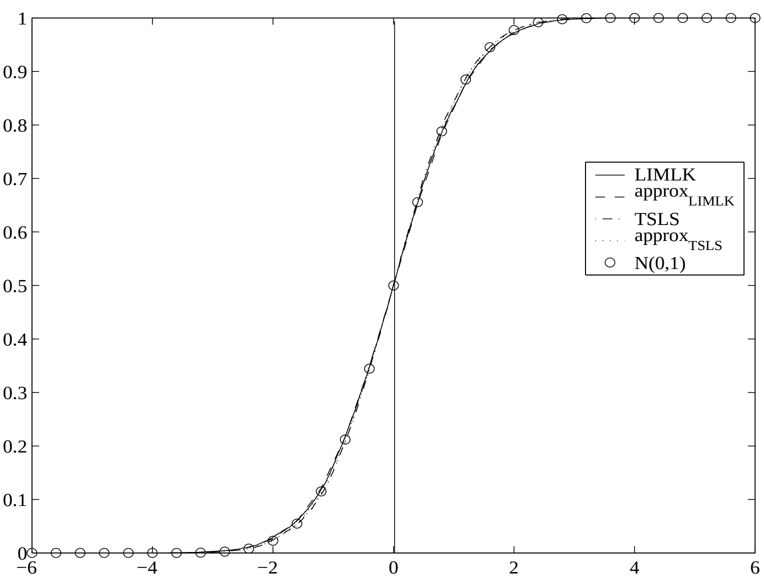

Figure 6: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=3$, $\theta = 0$ and $\lambda^2 = 50$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.045$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 0.955$

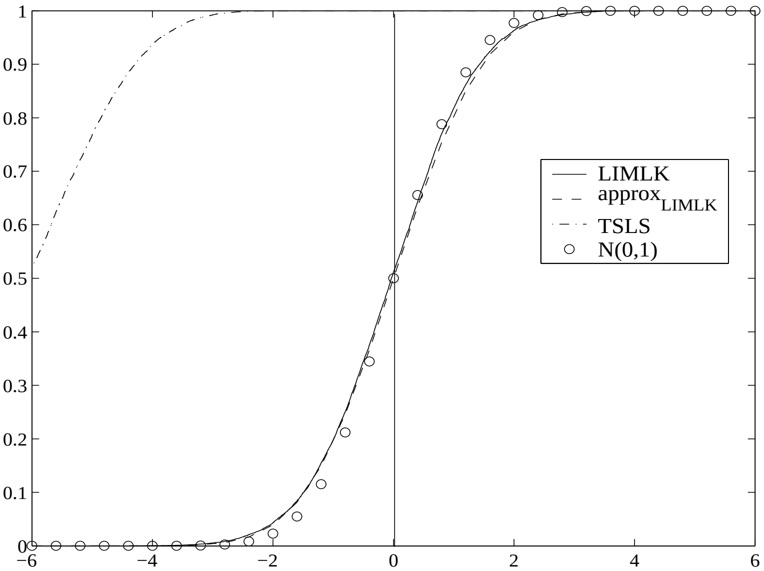

Figure 7: $\Pr\{\lambda(\hat{\theta}-\theta) \le t\}$ for $K = 30$, $\theta = 0.4\pi$ and $\lambda^2 = 100$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.31, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 42.14$

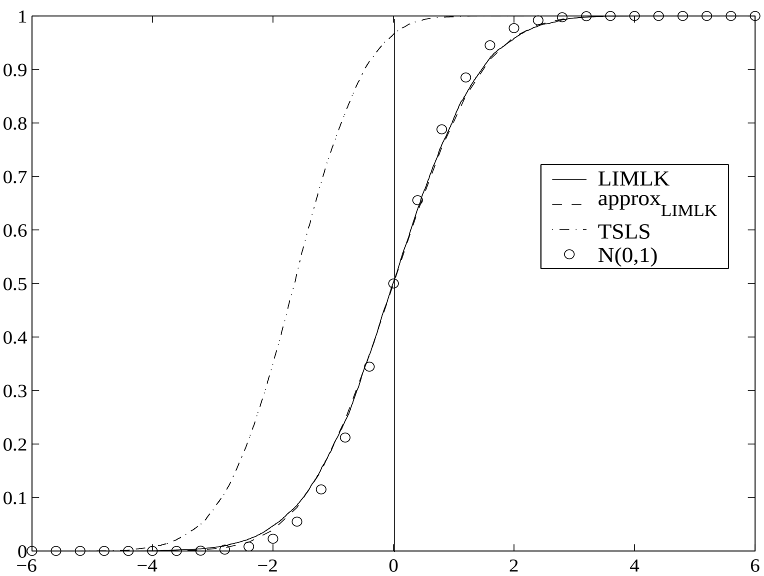

Figure 8: $\Pr\{\lambda(\hat{\theta}-\theta) \le t\}$ for $K = 30$, $\theta = 0.2\pi$ and $\lambda^2 = 100$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2 = 1.36, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2 = 3.61$

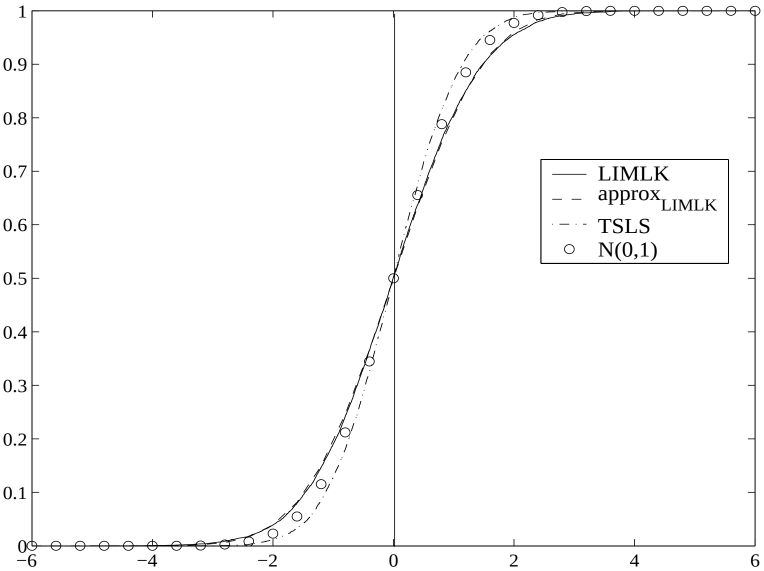

Figure 9: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K = 30$, $\theta = 0$ and $\lambda^2 = 100$

$\hat{E}[\lambda(\hat{\theta}_{LIMLK} - \theta)]^2 = 1.33, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS} - \theta)]^2 = 0.77$

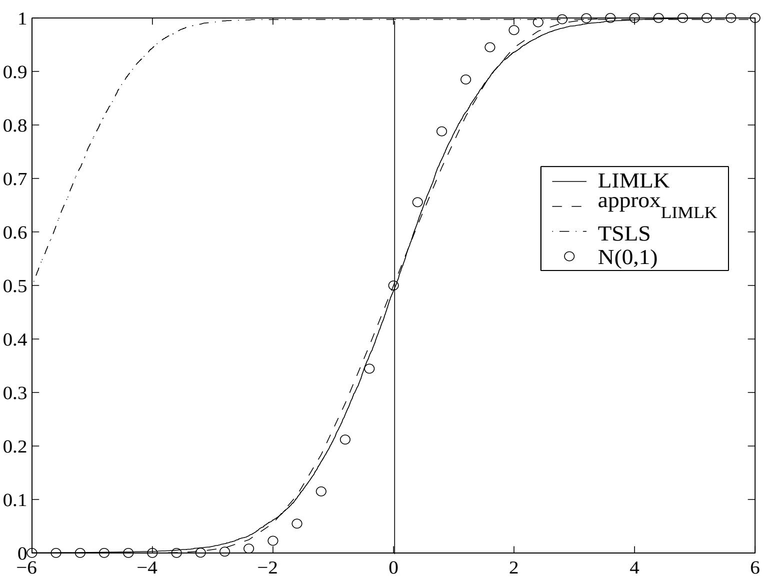

Figure 10: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K = 30$, $\theta = 0.4\pi$ and $\lambda^2 = 50$

$\hat{E}[\lambda(\hat{\theta}_{LIMLK} - \theta)]^2 = 1.78, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS} - \theta)]^2 = 40.12$

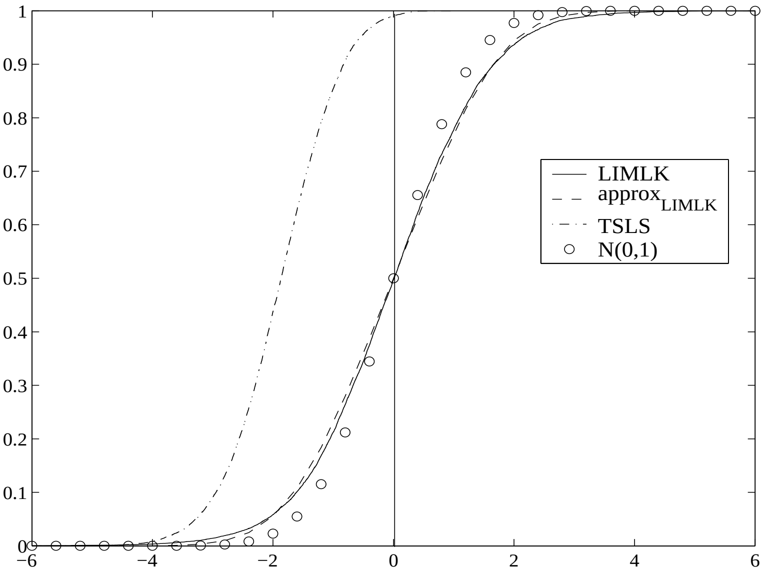

Figure 11: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=30,\ \theta=0.2\pi$ and $\lambda^2=50$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=1.78, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=4.22$

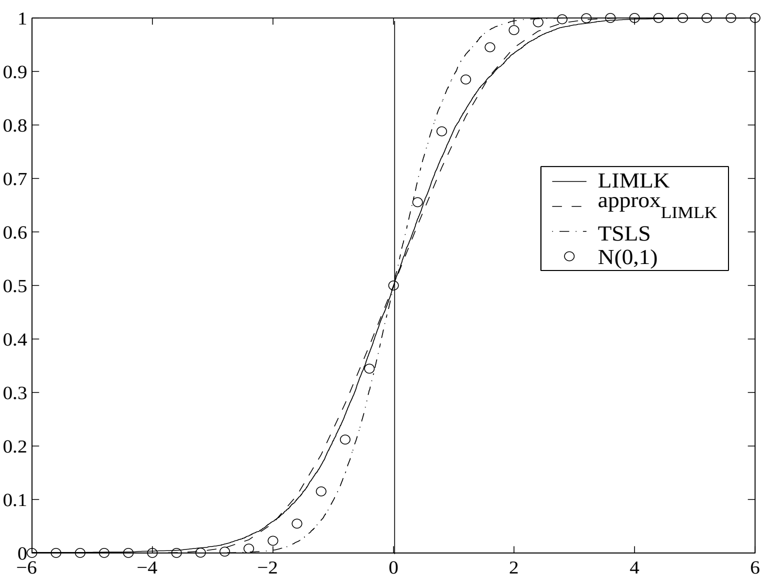

Figure 12: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=30,\ \theta=0$ and $\lambda^2=50$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=1.77, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=0.63$

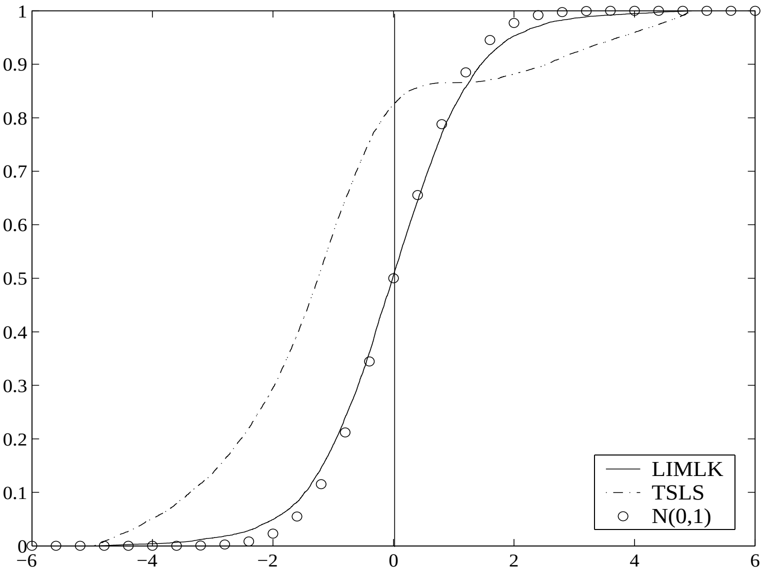

Figure 13: $\Pr\{\lambda(\hat{\theta}-\theta)\leq t\}$ for $K=3$, $\theta=0.4\pi$ and $\lambda^2=10$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=1.56$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=5.38$

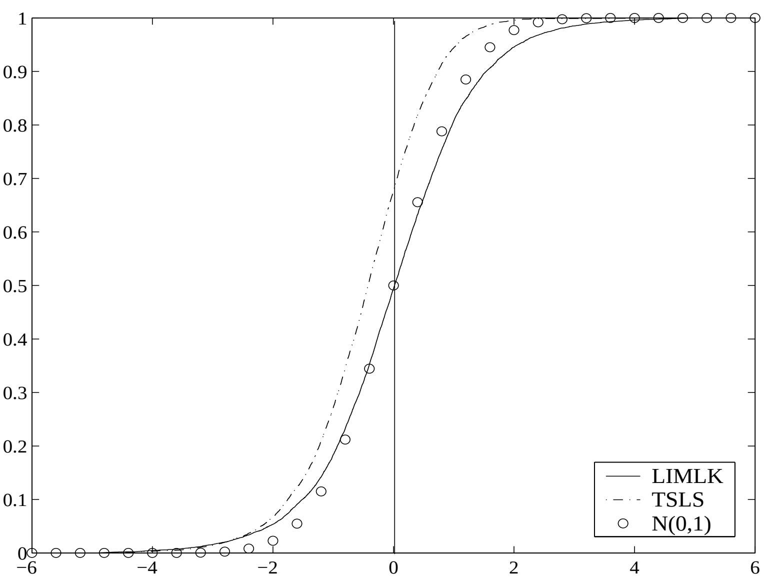

Figure 14: $\Pr\{\lambda(\hat{\theta}-\theta)\leq t\}$ for $K=3$, $\theta=0.2\pi$ and $\lambda^2=10$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=1.58$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=1.22$

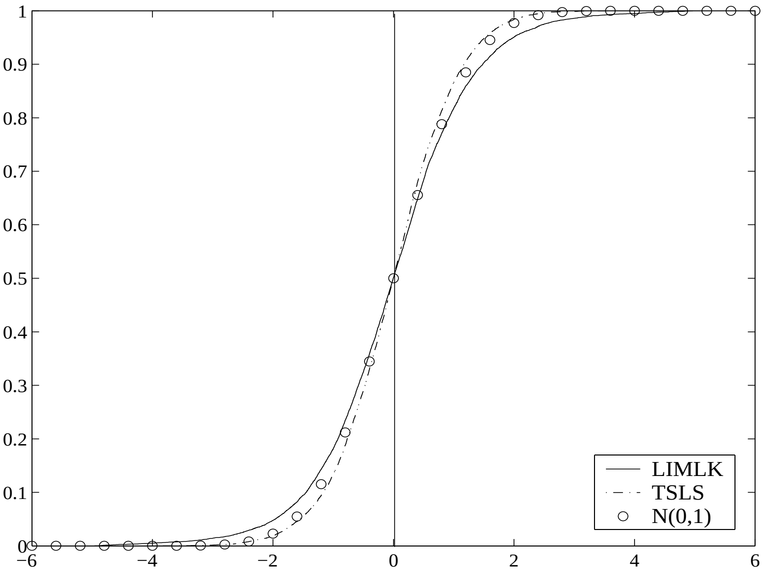

Figure 15: $\Pr\{\lambda(\hat{\theta}-\theta)\leq t\}$ for $K=3$, $\theta=0$ and $\lambda^2=10$

$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=1.57$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=0.88$

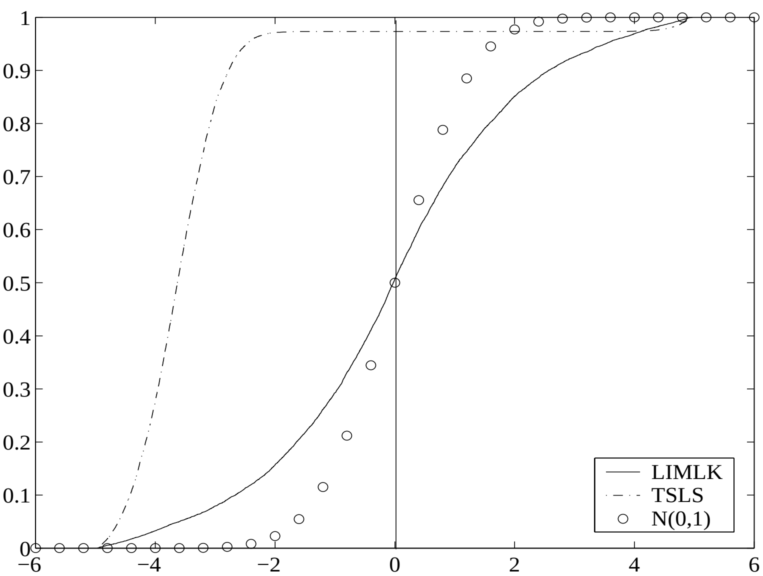

Figure 16: $\Pr\{\lambda(\hat{\theta}-\theta)\leq t\}$ for $K=30$, $\theta=0.4\pi$ and $\lambda^2=10$

$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=4.06$, $\hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=13.9$

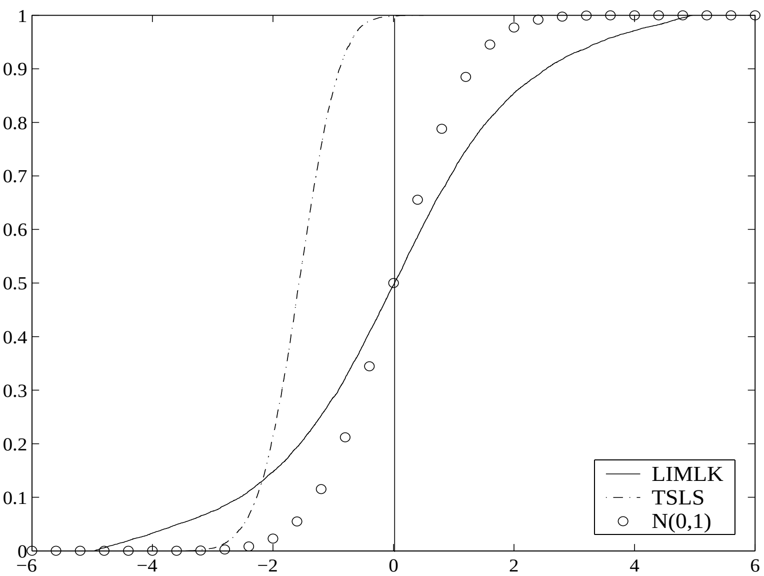

Figure 17: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=30,\ \theta=0.2\pi$ and $\lambda^2=10$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=3.90,\ \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=2.80$

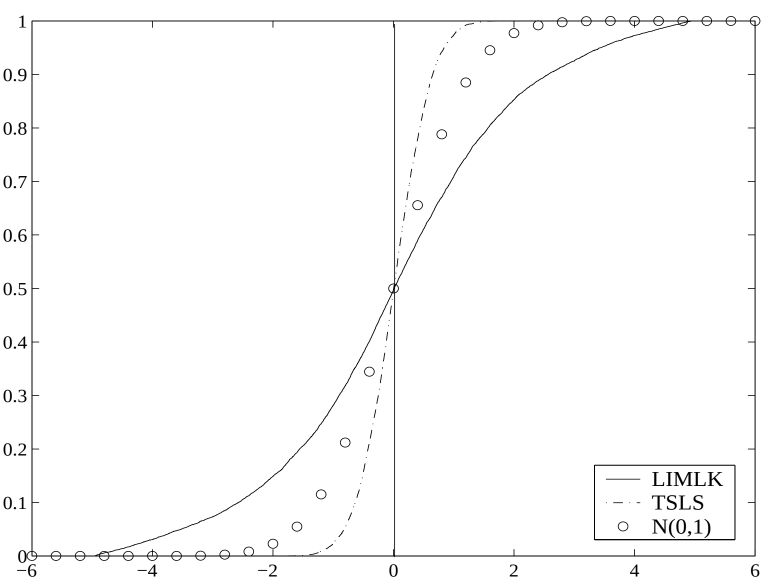

Figure 18: $\Pr\{\lambda(\hat{\theta}-\theta) \leq t\}$ for $K=30,\ \theta=0$ and $\lambda^2=10$
$\hat{E}[\lambda(\hat{\theta}_{LIMLK}-\theta)]^2=3.93, \quad \hat{E}[\lambda(\hat{\theta}_{TSLS}-\theta)]^2=0.25$