

The LIML Estimator Has Finite Moments!

T. W. Anderson*

Department of Economics and Department of Statistics

Stanford University, Stanford, CA 94305

March 25, 2010

Abstract

The Limited Information Maximum Likelihood estimator of the vector of coefficients of a structural equation in a simultaneous equation model is the vector that defines the linear combination maximizing the effect variance relative to the error variance. If this “eigenvector” solution is normalized by setting a designated coefficient equal to 1, the second-order moment of the estimator may be unbounded. However, the second-order moment is finite if the normalization is setting the sample error variance of the linear combination equal to 1.

Key words: Limited Information Maximum Likelihood, bounded moments, normalization.

*Thanks go to Naoto Kunitomo for very helpful suggestions.

1 Introduction.

The Limited Information Maximum Likelihood (LIML) estimator (Anderson and Rubin, 1949) was introduced as the solution of an “eigenvalue” problem, that is, a solution for \mathbf{b} in a homogeneous equation

$$(\mathbf{G} - d_1\mathbf{H})\mathbf{b} = \mathbf{0}, \quad (1)$$

where d_1 is the smallest root of

$$|\mathbf{G} - d\mathbf{H}| = 0, \quad (2)$$

and \mathbf{G} and \mathbf{H} are $p \times p$ positive semi-definite and positive definite matrices, respectively. A solution \mathbf{b} to (1) can be multiplied by a nonzero number to obtain another solution. The *conventional* normalization of \mathbf{b} has been to require

$$\mathbf{b}'\Phi\mathbf{b} = 1, \quad (3)$$

where Φ is a given $p \times p$ constant positive semi-definite matrix. When

$$\Phi = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (4)$$

the normalization is equivalent to setting $b_1 = 1$, where b_1 is the first component of \mathbf{b} .

There is a general perception among econometricians that the LIML estimator does not have finite moments. For example, Fuller (1977) wrote “the limited information estimator does not possess moments.” Mariano and Sawa (1972) have found the exact distribution of b_2 when $\mathbf{b} = (1, b_2)'$ and showed that $\mathcal{E}b_2^2 < \infty$ is not necessarily true.

The purpose of this present paper is to show that the *natural* normalization of

$$\mathbf{b}'\frac{1}{T}\mathbf{H}\mathbf{b} = 1 \quad (5)$$

yields an estimator with finite moments (if $T \geq p + q + 2$). While the conventional normalization has some advantages in interpretation, the natural normalization has the advantage of invariance with respect to linear transformations, for example, changes of the units of measurement. Anderson and Rubin considered both normalizations.

2 The model, normalizations, and invariances.

Let \mathbf{Y} be a $T \times p$ matrix of observed endogenous or dependent variables, \mathbf{Z} a $T \times q$ matrix of exogenous or independent variables, and \mathbf{V} a $T \times p$ matrix of unobserved random disturbances satisfying

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{V}, \tag{6}$$

the “reduced form.” The rows of

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}'_1 \\ \vdots \\ \mathbf{v}'_T \end{bmatrix} \tag{7}$$

are assumed to be independently normally distributed with mean $\mathcal{E}\mathbf{v}_t = \mathbf{0}$ and covariance $\mathcal{E}\mathbf{v}_t\mathbf{v}'_t = \mathbf{\Omega}$ (assumed to be positive definite). A “structural equation” may be defined by

$$\mathbf{Y}\boldsymbol{\beta} = \mathbf{u}, \tag{8}$$

where $\mathbf{u} = \mathbf{V}\boldsymbol{\beta}$ and

$$\mathbf{\Pi}\boldsymbol{\beta} = \mathbf{0}. \tag{9}$$

If $\text{rank}(\mathbf{\Pi}) = p - 1$, (9) determines $\boldsymbol{\beta}$ uniquely except for a multiplicative constant. The natural normalization of the parameter vector $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta} = 1. \tag{10}$$

The conventional normalization is

$$\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta} = 1, \quad (11)$$

where $\boldsymbol{\Phi}$ is a given $p \times p$ constant positive semi-definite matrix. It is conventional to take $\boldsymbol{\Phi}$ as (4), which is equivalent to setting $\beta_1 = 1$, where β_1 is the first component of $\boldsymbol{\beta}$. (Sometimes it is convenient to write $\boldsymbol{\beta}' = (1, -\boldsymbol{\beta}'_2)$.)

An important advantage of the natural normalization is that the model is invariant with respect to linear transformations of the dependent variable. If

$$\mathbf{Y}^* = \mathbf{Y}\mathbf{C}, \quad \boldsymbol{\Omega}^* = \mathbf{C}'\boldsymbol{\Omega}\mathbf{C}, \quad \boldsymbol{\Pi}^* = \boldsymbol{\Pi}\mathbf{C}, \quad \boldsymbol{\beta}^* = \mathbf{C}^{-1}\boldsymbol{\beta}, \quad \mathbf{V}^* = \mathbf{V}\mathbf{C}, \quad (12)$$

where \mathbf{C} is nonsingular, then

$$\boldsymbol{\Pi}^*\boldsymbol{\beta}^* = \mathbf{0}, \quad \boldsymbol{\beta}^{*\prime}\boldsymbol{\Omega}^*\boldsymbol{\beta}^* = 1, \quad (13)$$

corresponding to (9) and (10). Regardless of normalization, the model is invariant with respect to linear transformations of the exogenous variables. If

$$\mathbf{Z}^+ = \mathbf{Z}\mathbf{D}, \quad \boldsymbol{\Pi}^+ = \mathbf{D}^{-1}\boldsymbol{\Pi}, \quad (14)$$

where \mathbf{D} is nonsingular, then

$$\boldsymbol{\Pi}^+\boldsymbol{\beta} = \mathbf{0} \quad (15)$$

corresponding to (9). The model (6), (9), and (10) is *invariant* with respect to nonsingular linear transformations of \mathbf{Y} and \mathbf{Z} .

3 Estimation.

Theorem 1. The LIML estimator of $\boldsymbol{\beta}$ normalized by (10) is invariant with respect to transformations (12) and (14).

A sufficient set of statistics is

$$\mathbf{P} = \mathbf{A}^{-1}\mathbf{Z}'\mathbf{Y}, \quad \mathbf{H} = \mathbf{Y}'\mathbf{Y} - \mathbf{P}'\mathbf{A}\mathbf{P}, \quad (16)$$

where $\mathbf{A} = \mathbf{Z}'\mathbf{Z}$. Under the transformations (12) and (14) \mathbf{A} transforms to $\mathbf{D}'\mathbf{A}\mathbf{D}$, \mathbf{P} transforms to $\mathbf{D}^{-1}\mathbf{P}\mathbf{C}$, and \mathbf{H} transforms to $\mathbf{C}'\mathbf{H}\mathbf{C}$.

Anderson and Rubin derived the maximum likelihood estimates of $\boldsymbol{\Pi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\beta}$ under the conditions (9) and (10). Let \mathbf{b} satisfy (1) where

$$\mathbf{G} = \mathbf{P}'\mathbf{A}\mathbf{P}, \quad \mathbf{H} = \mathbf{Y}'\mathbf{Y} - \mathbf{P}'\mathbf{A}\mathbf{P}, \quad (17)$$

and

$$\mathbf{b}'\frac{1}{T}\mathbf{H}\mathbf{b} = 1. \quad (18)$$

Then the maximum likelihood estimator of $\boldsymbol{\beta}$ satisfying (9) and (10) is

$$\hat{\boldsymbol{\beta}} = \frac{1}{\sqrt{\mathbf{b}'\hat{\boldsymbol{\Omega}}\mathbf{b}}} \mathbf{b} = \frac{1}{\sqrt{1+d_1}} \mathbf{b}, \quad (19)$$

where

$$\hat{\boldsymbol{\Omega}} = \frac{1}{T}\mathbf{H} + d_1 \left(\frac{1}{T}\mathbf{H} \right) \mathbf{b}\mathbf{b}' \left(\frac{1}{T}\mathbf{H} \right). \quad (20)$$

Note that (18) and (20) imply

$$\mathbf{b}'\hat{\boldsymbol{\Omega}}\mathbf{b} = 1 + d_1 \quad (21)$$

and

$$\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\beta}} = 1. \quad (22)$$

Theorem 2. The LIML estimator $\hat{\beta}$ normalized by (10) satisfies

$$\mathcal{E}\hat{\beta}'\hat{\beta} < \infty \quad (23)$$

if $T \geq p + q + 2$.

Proof. By virtue of (18), (19), and (21)

$$\begin{aligned} \mathcal{E}\hat{\beta}'\hat{\beta} &= \mathcal{E} \frac{\mathbf{b}'\mathbf{b}}{1 + d_1} = \mathcal{E} \frac{\mathbf{b}'\mathbf{b}}{(1 + d_1) \mathbf{b}'\frac{1}{T}\mathbf{H}\mathbf{b}} \\ &\leq \mathcal{E} \frac{\mathbf{b}'\mathbf{b}}{\mathbf{b}'\frac{1}{T}\mathbf{H}\mathbf{b}} \leq \mathcal{E} \frac{T}{\min_{\mathbf{c}} \frac{\mathbf{c}'\mathbf{H}\mathbf{c}}{\mathbf{c}'\mathbf{c}}}. \end{aligned} \quad (24)$$

Note that

$$\min_{\mathbf{c}} \frac{\mathbf{c}'\mathbf{H}\mathbf{c}}{\mathbf{c}'\mathbf{c}} = \ell_1, \quad (25)$$

where ℓ_1 is the smallest characteristic root of \mathbf{H} , which has a Wishart distribution with $T - q$ degrees of freedom and $\mathcal{E}\mathbf{H} = (T - q)\mathbf{\Omega}$. By the invariance (12) $\mathcal{E}(1/\ell_1) < \infty$ if and only if $\mathcal{E}(1/m_1) < \infty$, where m_1 is the smallest characteristic root of $\mathbf{H}^* = \mathbf{C}'\mathbf{H}\mathbf{C}$ for $\mathbf{\Omega}^* = \mathbf{I}_p$.

The density of $m_1 \leq m_2 \leq \dots \leq m_p$ (the roots of $|\mathbf{H}^* - m\mathbf{I}_p| = 0$) is

$$\begin{aligned} &f(m_1, \dots, m_p; n, p) \\ &= C \prod_{i=1}^p m_i^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\sum_{i=1}^p m_i} \prod_{i < j} (m_j - m_i), \quad p \leq n, \end{aligned} \quad (26)$$

where $n = T - q$ and

$$C = \frac{\pi^{\frac{1}{2}p^2}}{2^{\frac{1}{2}pn} \Gamma_p(n) \Gamma_p(p)} \quad (27)$$

[Anderson (2003), Theorem 13.3.2]. Then

$$\mathcal{E} \frac{1}{m_1} = \int_0^\infty \int_0^{m_p} \dots \int_0^{m_2} \frac{1}{m_1} f(m_1, \dots, m_p; n, p) dm_1 \dots dm_{p-1} dm_p$$

$$\begin{aligned}
&\leq C \int_0^\infty \int_0^\infty \dots \int_0^\infty m_1^{\frac{1}{2}(n-p-3)} m_2^{\frac{1}{2}(n-p+1)} \dots m_p^{\frac{1}{2}(n+p-3)} e^{-\frac{1}{2} \sum_{i=1}^p m_i} dm_1 \dots dm_p \\
&= C 2^{\frac{1}{2}(p^2-1)} \Gamma\left(\frac{n-p-1}{2}\right) \Gamma\left(\frac{n-p+3}{2}\right) \dots \Gamma\left(\frac{n+p-1}{2}\right). \tag{28}
\end{aligned}$$

Thus $\mathcal{E}(1/m_1) < \infty$ if $n \geq p + 2$, that is, if $T - q \geq p + 2$. Q.E.D.

Mariano and Sawa (1972) pointed out that the LIML estimator for $p = 2$ with the conventional normalization does not necessarily have finite moments. Their analysis is based on a doubly-infinite series for the density of $\hat{\beta}_2$ when the conventional normalization is used and $\hat{\beta}_2$ is a scalar.

Theorem 2 shows that the lack of finite moments under conventional normalization is a feature of the normalization, not of the LIML estimator itself.

4 Generalizations.

In a more general model the structural equation is

$$\mathbf{Y}_1 \boldsymbol{\beta}_1 = \mathbf{Z}_1 \boldsymbol{\gamma}_1 + \mathbf{u}, \tag{29}$$

where $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$, $\mathbf{u} = \mathbf{V}\boldsymbol{\beta} = \mathbf{V}_1 \boldsymbol{\beta}_1$,

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix}. \tag{30}$$

Then $\boldsymbol{\Pi}\boldsymbol{\beta} = \mathbf{0}$ is

$$\begin{bmatrix} \boldsymbol{\Pi}_{11} \boldsymbol{\beta}_1 \\ \boldsymbol{\Pi}_{21} \boldsymbol{\beta}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix}. \tag{31}$$

Define

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1 \mathbf{Z}_1 & \mathbf{Z}'_1 \mathbf{Z}_2 \\ \mathbf{Z}'_2 \mathbf{Z}_1 & \mathbf{Z}'_2 \mathbf{Z}_2 \end{bmatrix}, \quad (32)$$

$$\mathbf{A}_{22 \cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}, \quad (33)$$

$$\mathbf{Z}_{2 \cdot 1} = \mathbf{Z}_2 - \mathbf{Z}_1 \mathbf{A}_{11}^{-1} \mathbf{A}_{12}, \quad \mathbf{P}_{21} = \mathbf{A}_{22 \cdot 1}^{-1} \mathbf{Z}'_{2 \cdot 1} \mathbf{Y}_1, \quad (34)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}, \quad \mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{bmatrix}. \quad (35)$$

Then \mathbf{G} and \mathbf{H} in (1) and (2) are replaced by $\mathbf{P}'_{21} \mathbf{A}_{22 \cdot 1} \mathbf{P}_{21}$ and

$$\mathbf{H}_{11} = \mathbf{Y}'_1 \mathbf{Y}_1 - \mathbf{P}'_{11} \mathbf{A}_{11} \mathbf{P}_{11} - \mathbf{P}'_{2 \cdot 1} \mathbf{A}_{22 \cdot 1} \mathbf{P}_{21}. \quad (36)$$

\mathbf{H}_{11} ($p_1 \times p_1$) has the Wishart distribution with $T - q$ degrees of freedom and $\mathcal{E} \mathbf{H}_{11} = (T - q) \mathbf{\Omega}_{11}$. The condition $\boldsymbol{\beta}' \mathbf{\Omega} \boldsymbol{\beta} = 1$ is

$$\boldsymbol{\beta}'_1 \mathbf{\Omega}_{11} \boldsymbol{\beta}_1 = 1. \quad (37)$$

Then $\mathcal{E} \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} = \mathcal{E} \hat{\boldsymbol{\beta}}'_1 \hat{\boldsymbol{\beta}}_1 < \infty$ if $T - q \geq p_1 + 2$.

It might be pointed out that a LIML estimator can be calculated with one normalization and converted to another. For example, the natural normalization may yield $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1^*, \hat{\beta}_2^*)'$ as the estimator; then $(1, \hat{\beta}_2^*/\hat{\beta}_1^*)$ is an estimator with a conventional normalization; if $\hat{\beta}_1^*$ is very small, one can expect that $\hat{\beta}_2^*/\hat{\beta}_1^*$ or $-\hat{\beta}_2^*/\hat{\beta}_1^*$ will have a large variance.

The argument in (24) and (25) holds whether or not \mathbf{H} has a Wishart distribution. Thus

$$\mathcal{E} \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} \leq \mathcal{E} \frac{T}{m_1} \quad (38)$$

if $T - q \geq p$ when $\hat{\beta}$ has the natural normalization.

Suppose \mathbf{H} has a density $g(m_1, \dots, m_p)$, where $m_1 < \dots < m_p$ are the characteristic roots of \mathbf{H} . Then the density of m_1, \dots, m_p is

$$\frac{\pi^{\frac{1}{2}p^2} g(m_1, \dots, m_p)}{\Gamma_p\left(\frac{1}{2}p\right)} \prod_{i>j} (m_i - m_j) \quad (39)$$

[Anderson (2003), Theorem 13.3.1]. Then

$$\mathcal{E} \frac{1}{m_1} \leq \frac{\pi^{\frac{1}{2}p^2}}{\Gamma_p\left(\frac{1}{2}p\right)} \int_0^\infty \dots \int_0^\infty m_1^{-1} m_2 \dots m_p^{p-1} g(m_1, \dots, m_p) dm_1 \dots dm_p. \quad (40)$$

Fuller (1977) suggested a modification of the LIML estimator with the conventional normalization to avoid the features of unbounded moments. Roughly speaking, the modification is a compromise between the two-stage least squares and the LIML estimator, replacing d_1 in (1) by a quantity smaller than d_1 and using the last $p - 1$ coordinates of (1). With the natural normalization there is no need for the Fuller modification.

5 Known covariance matrix.

When Ω is known, the estimator is based on

$$|\mathbf{G} - d\Omega| = 0, \quad (\mathbf{G} - d_1\Omega) \mathbf{b} = \mathbf{0} \quad (41)$$

and is known as the LIMLK estimator. The natural normalization is $\mathbf{b}'\Omega\mathbf{b} = 1$. Let $\Omega = \Gamma'\Delta\Gamma$, where $\Gamma'\Gamma = \mathbf{I}$ and $\Delta = \text{diag}(\delta_1, \dots, \delta_p)$. Then

$$\begin{aligned} 1 &= \mathbf{b}'\Gamma'\Delta\Gamma\mathbf{b} \leq \mathbf{b}'\Gamma'\Gamma\mathbf{b} \min(\delta_1, \dots, \delta_p) \\ &= \mathbf{b}'\mathbf{b} \min(\delta_1, \dots, \delta_p). \end{aligned} \quad (42)$$

Hence

$$\mathcal{E}\mathbf{b}'\mathbf{b} \leq \frac{1}{\min(\delta_1, \dots, \delta_p)} < \infty. \quad (43)$$

When $\mathbf{\Omega} = \mathbf{\Omega}^* = \mathbf{I}$ in (12), the transformation matrix \mathbf{C} is orthogonal $\mathbf{C}'\mathbf{C} = \mathbf{I}$. Anderson, Stein, and Zaman (1985) have proved an admissibility theorem that demonstrates an advantage of the natural normalization. Let $\mathbf{\Omega} = \mathbf{I}_2$, $\boldsymbol{\beta} = (\cos \phi, \sin \phi)'$, $\boldsymbol{\alpha} = (-\sin \phi, \cos \phi)'$, and $\mathbf{\Pi} = \boldsymbol{\eta}\boldsymbol{\alpha}'$. Consider estimation of ϕ with the loss function

$$\sin^2(\hat{\phi} - \phi) = 1 - (\hat{\boldsymbol{\alpha}}'\boldsymbol{\alpha})^2 = 1 - (\hat{\boldsymbol{\beta}}'\boldsymbol{\beta})^2, \quad (44)$$

where $\hat{\boldsymbol{\alpha}}$ is an estimator of $\boldsymbol{\alpha}$ and $\hat{\phi}$ is the corresponding estimator of ϕ . Then the solution of

$$|\mathbf{G} - d\mathbf{I}_2| = 0, \quad (\mathbf{G} - d_1\mathbf{I}_2)\mathbf{b}, \quad (45)$$

and $\mathbf{b}'\mathbf{b} = 1$ is the best invariant estimator for specified value of $\boldsymbol{\eta}'\boldsymbol{\eta}$ (Theorem 1 of Anderson, Stein, and Zaman). Note that the risk, $\mathcal{E} \sin^2(\hat{\phi} - \phi)$, of an invariant procedure depends only on $\boldsymbol{\eta}'\boldsymbol{\eta}$. It follows (Corollary 1 of Anderson, Stein and Zaman) that the LIMLK estimator with normalization $\mathbf{b}'\mathbf{b} = 1$ and loss function (44) is *admissible*. Roughly speaking, for every $\boldsymbol{\eta}'\boldsymbol{\eta}$ this LIMLK estimator minimizes $\mathcal{E} \sin^2(\hat{\phi} - \phi)$.

An intuitive understanding of the possible difficulty with the conventional normalization is to consider the case of $G = 2$ and $\mathbf{\Omega} = \mathbf{I}_2$. Then in the natural normalization

$$\boldsymbol{\beta} = \begin{bmatrix} \cos \phi \\ \sin \phi \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \cos \hat{\phi} \\ \sin \hat{\phi} \end{bmatrix}, \quad (46)$$

while in the conventional normalization

$$\boldsymbol{\beta}^* = \begin{bmatrix} 1 \\ \frac{\sin \phi}{\cos \phi} \end{bmatrix} = \begin{bmatrix} 1 \\ \tan \phi \end{bmatrix}, \quad \hat{\boldsymbol{\beta}}^* = \begin{bmatrix} 1 \\ \tan \hat{\phi} \end{bmatrix}. \quad (47)$$

If ϕ is close to $\pi/2$, $\tan \phi$ will be large. ($\tan \phi \rightarrow \infty$ as $\phi \rightarrow \pi/2$.) Similarly $\hat{\phi}$ will be close to $\pi/2$, $\tan \hat{\phi}$ will be large, and $\hat{\phi} - \phi$ can be expected to have a large variance.

References

- Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, Third Edition, John Wiley and Sons, Inc.
- Anderson, T.W., and Herman Rubin (1949), Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics*, **20**, 46–63.
- Anderson, T.W., Charles Stein, and Asad Zaman (1985), Best invariant estimation of a direction parameter, *Annals of Statistics*, **13**, 526–533.
- Mariano, Roberto S., and Takamitsu Sawa (1972), The exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables, *Journal of the American Statistical Association*, **67**, 159–163.
- Fuller, Wayne A. (1977), Some properties of a modification of the limited information estimator, *Econometrica* **65**, 939–953.