Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 127 (2005) 1–16
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Origins of the limited information maximum likelihood and two-stage least squares estimators

T.W. Anderson*

Departments of Economics and Statistics, Seqnoia Hall, Stanford University, Stanford, CA 94305-4065, USA

Received 9 July 2004
Available online 5 November 2004

Abstract

Theil, Basmann, and Sargan are often credited with the development of the two-stage least squares (TSLS) estimator of the coefficients of one structural equation in a simultaneous equations model. However, Anderson and Rubin had earlier derived the asymptotic distribution of the limited information maximum likelihood (LIML) estimator by finding the asymptotic distribution of what is essentially the TSLS estimator.

JEL classification: C13; C30

Keywords: Limited information maximum likelihood estimator; Two-stage least squares estimator; Asymptotic distributions

1. Introduction

To estimate the coefficients of one equation in a system of simultaneous equations the two-stage least squares (TSLS) estimator has been widely used. It was introduced more or less independently by Theil (1953a, b, 1954, 1961), Basmann (1957), and Sargan (1958). However, Anderson and Rubin (1950) had derived the asymptotic

*Corresponding author. Tel.: +1 650 723 4732; fax: +1 650 725 8977.
E-mail address: twa@stat.stanford.edu (T.W. Anderson).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.09.012

distribution of the limited information maximum likelihood (LIML) estimator (Anderson and Rubin, 1949) by deriving the asymptotic distribution of an approximating estimator. This approximating estimator, which became known later as the two-stage least squares estimator, was defined obliquely. The LIML estimator was expressed explicitly in terms of the cofactors of a certain matrix involving the "effect" matrix of the reduced form and the "error" covariance matrix while the TSLS estimator was implied as the same function of the cofactors of the effect matrix alone. The asymptotic distribution of the cofactors defining the LIML estimator is the same as the asymptotic distribution of the cofactors used for the TSLS estimator.

Since this aspect of econometric theory may be entering into a historical phase, it seems worthwhile to clarify the relations between the LIML and TSLS estimators. It is a good guess that not many econometricians are aware of the contents of Anderson and Rubin (1950) even though it is a basic contribution to statistical inference in simultaneous equations models. Theil, Basmann, and Sargan cite the two Anderson–Rubin papers, but do not point out that their treatment includes indirectly the TSLS estimator and its asymptotic distribution. Admittedly, the notation of that paper is difficult and the exposition somewhat obscure. The purpose of this present paper is to elucidate the part of the earlier papers pertaining to these two single-equation estimators.

2. LIML and TSLS estimators

2.1. The LIML estimator

Let the simultaneous equations model be

$$\mathbf{B}\mathbf{y}_t = \mathbf{\Gamma}\mathbf{z}_t + \mathbf{u}_t, \tag{2.1}$$

where $\mathbf{y}_t$ is a vector of G current endogenous variables, $\mathbf{z}_t$ is a vector of K predetermined variables, $\mathbf{u}_t$ is a vector of G unobserved independent disturbances with $\mathscr{E}\mathbf{u}_t = \mathbf{0}$ and $\mathscr{E}\mathbf{u}_t\mathbf{u}_t' = \mathbf{\Sigma}$, and $\mathbf{B}$ is nonsingular. Let the reduced form of the model be

$$\mathbf{y}_t = \mathbf{\Pi}\mathbf{z}_t + \mathbf{v}_t, \tag{2.2}$$

where $\mathbf{\Pi} = \mathbf{B}^{-1}\mathbf{\Gamma}$ and $\mathbf{v}_t = \mathbf{B}^{-1}\mathbf{u}_t$ with $\mathscr{E}\mathbf{v}_t = \mathbf{0}$ and $\mathscr{E}\mathbf{v}_t\mathbf{v}_t' = \mathbf{\Omega} = \mathbf{B}^{-1}\mathbf{\Sigma}(\mathbf{B}')^{-1}$. Let the structural equation of interest be

$$\boldsymbol{\beta}'\mathbf{y}_t = \boldsymbol{\gamma}'\mathbf{z}_t + u_{1t}. \tag{2.3}$$

Suppose that the equation is identified by assigned zero coefficients. Let the components of the vectors be numbered so that

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{It} \\ \mathbf{y}_{Et} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_I \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{v}_t = \begin{bmatrix} \mathbf{v}_{It} \\ \mathbf{v}_{Et} \end{bmatrix}, \quad \mathbf{z}_t = \begin{bmatrix} \mathbf{z}_{it} \\ \mathbf{z}_{et} \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_i \\ \mathbf{0} \end{bmatrix}. \tag{2.4}$$

I and i indicate the G_I and K_i variables *included* in the equation, and E and e indicate the G_E and K_e variables in the model that are *excluded* from (2.3). The structural equation is written as

$$\boldsymbol{\beta}_I'\mathbf{y}_{It} = \boldsymbol{\gamma}_i'\mathbf{z}_{it} + u_{1t}. \tag{2.5}$$

Partition the matrix of coefficients in the reduced form as

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_{Ii} & \boldsymbol{\Pi}_{Ie} \\ \boldsymbol{\Pi}_{Ei} & \boldsymbol{\Pi}_{Ee} \end{bmatrix}. \tag{2.6}$$

The relation between the structural equation and the reduced form is $\boldsymbol{\beta}'\boldsymbol{\Pi} = \boldsymbol{\gamma}'$ or in partitioned form

$$(\boldsymbol{\beta}_I', \mathbf{0})\begin{bmatrix} \boldsymbol{\Pi}_{Ii} & \boldsymbol{\Pi}_{Ie} \\ \boldsymbol{\Pi}_{Ei} & \boldsymbol{\Pi}_{Ee} \end{bmatrix} = (\boldsymbol{\beta}_I'\boldsymbol{\Pi}_{Ii}, \boldsymbol{\beta}_I'\boldsymbol{\Pi}_{Ie}) = (\boldsymbol{\gamma}_i', \mathbf{0}). \tag{2.7}$$

The relevant part of the reduced form is

$$\mathbf{y}_{It} = \boldsymbol{\Pi}_{Ii}\mathbf{z}_{it} + \boldsymbol{\Pi}_{Ie}\mathbf{z}_{et} + \mathbf{v}_{It} \tag{2.8}$$

with $\mathscr{E}\mathbf{v}_{It}\mathbf{v}_{It}' = \boldsymbol{\Omega}_{II}$. Given $\boldsymbol{\Pi}$, the vector $\boldsymbol{\beta}_I$ can be determined by $\boldsymbol{\beta}_I'\boldsymbol{\Pi}_{Ie} = \mathbf{0}$ and a normalization $\boldsymbol{\beta}_I'\boldsymbol{\Phi}_{II}\boldsymbol{\beta}_I = 1$ when the rank of $\boldsymbol{\Pi}_{Ie}$ is $G_I - 1$. (This implies that $K_e \geqslant G_I - 1$.) If

$$\boldsymbol{\Phi}_{II} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{2.9}$$

the normalization is that the first component of $\boldsymbol{\beta}_I$ is ± 1. Further $\boldsymbol{\gamma}_i' = \boldsymbol{\beta}_I'\boldsymbol{\Pi}_{Ii}$.

A sample $(\mathbf{y}_1, \mathbf{z}_1), \ldots, (\mathbf{y}_T, \mathbf{z}_T)$ is observed. The predetermined variables $\mathbf{z}_1, \ldots, \mathbf{z}_T$ will be assumed exogenous and nonstochastic. The second-order sample moments are written as

$$\mathbf{M}_{yy} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' = \begin{bmatrix} \mathbf{M}_{II} & \mathbf{M}_{IE} \\ \mathbf{M}_{EI} & \mathbf{M}_{EE} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \mathbf{y}_{It}\mathbf{y}_{It}' & \mathbf{y}_{It}\mathbf{y}_{Et}' \\ \mathbf{y}_{Et}\mathbf{y}_{It}' & \mathbf{y}_{Et}\mathbf{y}_{Et}' \end{bmatrix}. \tag{2.10}$$

The matrices $\mathbf{M}_{yz}$ and $\mathbf{M}_{zz}$ are defined and partitioned similarly.

Anderson and Rubin (1949) derived the maximum likelihood estimator of $\boldsymbol{\beta}_I$ first. It is convenient to transform the predetermined variables so that the excluded variables are uncorrelated with the included variables. Let

$$\mathbf{z}_{\perp t} = \mathbf{z}_{et} - \mathbf{M}_{ei}\mathbf{M}_{ii}^{-1}\mathbf{z}_{it} \tag{2.11}$$

(the residuals from the regression of the $\mathbf{z}_{et}$'s on the $\mathbf{z}_{it}$'s). Then the $\mathbf{z}_{\perp t}$'s are uncorrelated with the $\mathbf{z}_{it}$'s:

$$\mathbf{M}_{\perp i} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{\perp t}\mathbf{z}_{it}' = \mathbf{0}, \quad \mathbf{M}_{\perp\perp} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{\perp t}\mathbf{z}_{\perp t}' = \mathbf{M}_{ee} - \mathbf{M}_{ei}\mathbf{M}_{ii}^{-1}\mathbf{M}_{ie}. \tag{2.12}$$

The reduced form (2.8) is written as

$$\begin{aligned} \mathbf{y}_{It} &= \boldsymbol{\Pi}_{Ii}\mathbf{z}_{it} + \boldsymbol{\Pi}_{Ie}(\mathbf{z}_{\perp t} + \mathbf{M}_{ei}\mathbf{M}_{ii}^{-1}\mathbf{z}_{it}) + \mathbf{v}_{It} \\ &= (\boldsymbol{\Pi}_{Ii} + \boldsymbol{\Pi}_{Ie}\mathbf{M}_{ei}\mathbf{M}_{ii}^{-1})\mathbf{z}_{it} + \boldsymbol{\Pi}_{Ie}\mathbf{z}_{\perp t} + \mathbf{v}_{It} \\ &= \overline{\boldsymbol{\Pi}}_{Ii}\mathbf{z}_{it} + \boldsymbol{\Pi}_{Ie}\mathbf{z}_{\perp t} + \mathbf{v}_{It}, \end{aligned} \tag{2.13}$$

where $\overline{\mathbf{\Pi}}_{Ii} = \mathbf{\Pi}_{Ii} + \mathbf{\Pi}_{Ie}\mathbf{M}_{ei}\mathbf{M}_{ii}^{-1}$. The regression of $\mathbf{y}_{It}$ on $(\mathbf{z}'_{it}, \mathbf{z}'_{\perp t})'$ gives the sample regression coefficients

$$(\overline{\mathbf{P}}_{Ii}, \mathbf{P}_{Ie}) = (\mathbf{M}_{Ii}\mathbf{M}_{ii}^{-1}, \mathbf{M}_{I\perp}\mathbf{M}_{\perp\perp}^{-1}), \tag{2.14}$$

where $\mathbf{M}_{I\perp} = (1/T)\sum_{t=1}^{T} \mathbf{y}_{It}\mathbf{z}'_{\perp t}$. An estimator of $\mathscr{E}\mathbf{v}_{It}\mathbf{v}'_{It} = \mathbf{\Omega}_{II}$ is

$$\mathbf{W}_{II} = \mathbf{M}_{II} - \overline{\mathbf{P}}_{Ii}\mathbf{M}_{ii}\overline{\mathbf{P}}'_{Ii} - \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}. \tag{2.15}$$

The logarithm of the normal likelihood function is

$$\begin{aligned} \log L = & -\frac{1}{2}TG_I \log 2\pi - \frac{1}{2}T \log |\mathbf{\Omega}_{II}| \\ & -\frac{1}{2}\mathrm{tr}\mathbf{\Omega}_{II}^{-1}[\mathbf{W}_{II} + (\overline{\mathbf{P}}_{Ii} - \overline{\mathbf{\Pi}}_{Ii})\mathbf{M}_{ii}(\overline{\mathbf{P}}'_{Ii} - \overline{\mathbf{\Pi}}'_{Ii}) \\ & + (\mathbf{P}_{Ie} - \mathbf{\Pi}_{Ie})\mathbf{M}_{\perp\perp}(\mathbf{P}'_{Ie} - \mathbf{\Pi}'_{Ie})]. \end{aligned} \tag{2.16}$$

The matrices $\mathbf{W}_{II}$, $\overline{\mathbf{P}}_{Ii}$ and $\mathbf{P}_{Ie}$ are a set of sufficient statistics for $\mathbf{y}_1, \ldots, \mathbf{y}_T$. The maximum likelihood estimators of $\mathbf{\Omega}_{II}$, $\overline{\mathbf{\Pi}}_{Ii}$, and $\mathbf{\Pi}_{Ie}$ are obtained subject to the restrictions $\boldsymbol{\beta}'_I\mathbf{\Pi}_{Ie} = \mathbf{0}$ and $\boldsymbol{\beta}'_I\mathbf{\Phi}_{II}\boldsymbol{\beta}_I = 1$.

Let v be the smallest root of

$$|\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - \lambda\mathbf{W}_{II}| = 0, \tag{2.17}$$

and let $\mathbf{b}$ be a solution of

$$(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - v\mathbf{W}_{II})\mathbf{b} = \mathbf{0}. \tag{2.18}$$

The limited information maximum likelihood estimator of $\boldsymbol{\beta}_I$ is

$$\widehat{\boldsymbol{\beta}}_I = \frac{1}{\sqrt{\mathbf{b}'\mathbf{\Phi}_{II}\mathbf{b}}}\mathbf{b}. \tag{2.19}$$

This is (5.2) of Anderson and Rubin (1949). The LIML estimator of $\boldsymbol{\gamma}'_i$ is $\widehat{\boldsymbol{\gamma}}'_i = \widehat{\boldsymbol{\beta}}'_I\overline{\mathbf{P}}_{Ii}$. These estimators are maximum likelihood when $\mathbf{v}_{It}$ is normally distributed.

2.2. The TSLS estimator as an approximation to the LIML estimator

In the second paper Anderson and Rubin (1950) developed the asymptotic properties of $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\boldsymbol{\gamma}}_i$ under very general conditions on the model. Before reviewing their development (in Section 4) we present the ideas of the LIML and TSLS estimators in a more straightforward manner.

Let $\mathrm{vec}(\mathbf{a}_1, \ldots, \mathbf{a}_m) = (\mathbf{a}'_1, \ldots, \mathbf{a}'_m)'$, and let $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$ be the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.

A basic step in the development is that under appropriate conditions

$$\sqrt{T}v\mathbf{W}_{II} \xrightarrow{p} \mathbf{0}. \tag{2.20}$$

Theorem 1. *Assume conditions on* $\{\mathbf{z}_t\}$ *and* $\{\boldsymbol{\eta}_t\}$ *so*

$$\mathbf{M}_{zz} \to \mathbf{M}^0_{zz} \text{ nonsingular}, \tag{2.21}$$

$$\mathbf{P} = \mathbf{M}_{yz}\mathbf{M}_{zz}^{-1} \xrightarrow{p} \boldsymbol{\Pi}, \tag{2.22}$$

$$\sqrt{T}\ \mathrm{vec}\,(\mathbf{P} - \boldsymbol{\Pi}) = \ \mathrm{vec}\ \mathbf{P}^* \xrightarrow{d} N[\mathbf{0}, (\mathbf{M}^0_{zz})^{-1} \otimes \boldsymbol{\Omega}], \tag{2.23}$$

$$\mathbf{W}_{II} \xrightarrow{p} \boldsymbol{\Omega}_{II}. \tag{2.24}$$

Then $\sqrt{T}v\mathbf{W}_{II} \xrightarrow{p} \mathbf{0}$.

Proof. We have

$$\begin{aligned} v &= \frac{\mathbf{b}'\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\mathbf{b}}{\mathbf{b}'\mathbf{W}_{II}\mathbf{b}} \\ &= \min_{\mathbf{c}} \frac{\mathbf{c}'\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\mathbf{c}}{\mathbf{c}'\mathbf{W}_{II}\mathbf{c}} \\ &\leqslant \frac{\boldsymbol{\beta}'\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{W}_{II}\boldsymbol{\beta}} \\ &= \frac{\boldsymbol{\beta}'(\boldsymbol{\Pi}_{Ie} + \frac{1}{\sqrt{T}}\mathbf{P}^*_{Ie})\mathbf{M}_{\perp\perp}(\boldsymbol{\Pi}_{Ie} + \frac{1}{\sqrt{T}}\mathbf{P}^*_{Ie})'\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{W}_{II}\boldsymbol{\beta}} \\ &= \frac{\boldsymbol{\beta}'\mathbf{P}^*_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}^{*\prime}_{Ie}\boldsymbol{\beta}}{T\boldsymbol{\beta}'\mathbf{W}_{II}\boldsymbol{\beta}}. \end{aligned} \tag{2.25}$$

Thus $\sqrt{T}v \xrightarrow{p} 0$ and $\sqrt{T}v\mathbf{W}_{II} \xrightarrow{p} \mathbf{0}$. □

Let A^{kj} be the cofactor of a_{kj} in the square matrix $\mathbf{A} = (a_{kj})$ of order K. Then $|\mathbf{A}| = \sum_{k=1}^{K} a_{Jk}A^{Jk}$ for any $J = 1, \ldots, K$. If $\mathbf{A}$ has rank $K - 1$ (and hence $|\mathbf{A}| = 0$), a solution of

$$\mathbf{A}\mathbf{x} = \mathbf{0} \tag{2.26}$$

is

$$\mathbf{x} = \begin{bmatrix} A^{J1} \\ \vdots \\ A^{JK} \end{bmatrix} \tag{2.27}$$

for any J, $J = 1, \ldots K$. A solution $\mathbf{b}$ to (2.18) is given by

$$\mathbf{b}_J = \begin{bmatrix} (\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - v\mathbf{W}_{II})^{J1} \\ \vdots \\ (\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - v\mathbf{W}_{II})^{JG_I} \end{bmatrix} \tag{2.28}$$

for any $J = 1, \ldots, G_I$. Because $\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - v\mathbf{W}_{II}$ has rank $G_I - 1$ (with probability 1), the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{G_I}$ are proportional to each other. Let $\widehat{\rho}_J = 1/\sqrt{\mathbf{b}'_J\boldsymbol{\Phi}_{II}\mathbf{b}_J}$. Then

$$\widehat{\boldsymbol{\beta}}_I = \widehat{\rho}_J\mathbf{b}_J, \quad J = 1, \ldots, G_I. \tag{2.29}$$

The import of $\sqrt{T}v\mathbf{W}_{II} \xrightarrow{p} \mathbf{0}$ is that

$$\sqrt{T}[(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - v\mathbf{W}_{II}) - \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}] \xrightarrow{p} \mathbf{0} \tag{2.30}$$

and hence that

$$\sqrt{T}[(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie} - v\mathbf{W}_{II})^{kj} - (\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})^{kj}] \xrightarrow{p} 0, \quad k,j = 1,\ldots,G_I. \tag{2.31}$$

Since the term $v\mathbf{W}_{II}$ is relatively small, an approximation to $\mathbf{b}_J$ given by (2.28) is

$$\mathbf{b}_J^{\mathrm{T}} = \begin{bmatrix} (\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})^{J1} \\ \vdots \\ (\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})^{JG_I} \end{bmatrix} \tag{2.32}$$

for any J. Because the rank of $\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}$ is G_I (with probability 1), the vectors $\mathbf{b}_1^{\mathrm{T}},\ldots,\mathbf{b}_{G_I}^{\mathrm{T}}$ are not proportional to each other. However, because of (2.31), the limiting distribution of $\sqrt{T}(\widehat{\rho}_J\mathbf{b}_J - \boldsymbol{\beta}_I)$ is the same as the limiting distribution of $\sqrt{T}(\widehat{\rho}_J^{\mathrm{T}}\mathbf{b}_J^{\mathrm{T}} - \boldsymbol{\beta}_I)$, where $\widehat{\rho}_J^{\mathrm{T}} = 1/\sqrt{\mathbf{b}_J^{\mathrm{T}\prime}\boldsymbol{\Phi}_{II}\mathbf{b}_J^{\mathrm{T}}}$.

If the term $v\mathbf{W}_{II}$ is deleted from (2.18), the resulting equation is $\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\mathbf{b} = \mathbf{0}$, which has only the trivial solution $\mathbf{b} = \mathbf{0}$. However, $\mathbf{b}_J^{\mathrm{T}}$ satisfies

$$(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})_{J\cdot}\mathbf{b} = \mathbf{0}, \tag{2.33}$$

where $(\quad)_{J\cdot}$ denotes $(\quad)$ with the J-th row deleted. A form of the TSLS estimator is $(1/\sqrt{\mathbf{b}_J^{\mathrm{T}\prime}\boldsymbol{\Phi}_{II}\mathbf{b}_J^{\mathrm{T}}})\mathbf{b}_J^{\mathrm{T}}$. Usually $J = 1$.

2.3. The asymptotic distribution

The limiting distribution of $\sqrt{T}(\widehat{\rho}_J^{\mathrm{T}}\mathbf{b}_J^{\mathrm{T}} - \boldsymbol{\beta}_I)$ under the conditions of Theorem 1 is normal because the limiting distribution of $\sqrt{T}(\mathbf{P}_{Ie} - \boldsymbol{\Pi}_{Ie})$ is normal and $\mathbf{M}_{\perp\perp} \xrightarrow{p} \mathbf{M}_{\perp\perp}^0 = \mathbf{M}_{ee}^0 - \mathbf{M}_{ei}^0(\mathbf{M}_{ii}^0)^{-1}\mathbf{M}_{ie}^0$. Anderson and Rubin found the covariance matrix of the limiting normal distribution of $\sqrt{T}(\widehat{\rho}_J^{\mathrm{T}}\mathbf{b}_J^{\mathrm{T}} - \boldsymbol{\beta}_I)$ for a general normalization $\boldsymbol{\Phi}_{II}$ and under general conditions on $\{\mathbf{z}_t\}$ and $\{\mathbf{V}_t\}$; some details are recorded in Section 4.

It is straightforward to obtain the covariance matrix when $\boldsymbol{\beta}$ is normalized by requiring the first component to be $\beta_1 = 1$, corresponding to use of (2.9). We can partition $\boldsymbol{\beta}_I$ and $\mathbf{y}_{It}$ as

$$\boldsymbol{\beta}_I = \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{bmatrix}, \quad \mathbf{y}_{It} = \begin{bmatrix} y_{1t} \\ \mathbf{y}_{2t} \end{bmatrix}. \tag{2.34}$$

Then the structural equation (2.5) can be written

$$y_{1t} = \boldsymbol{\beta}'_2\mathbf{y}_{2t} + \boldsymbol{\gamma}'_i\mathbf{z}_{it} + u_{1t}. \tag{2.35}$$

Let a solution $\mathbf{b}$ to (2.18) be similarly normalized so $b_1 = 1$; this $\mathbf{b}$ is partitioned as $\mathbf{b}' = (1, -\mathbf{b}'_2)'$. The LIML estimator of $(1, -\boldsymbol{\beta}'_2)'$ is $(1/b_{1J})\mathbf{b}_J$ for any J. Similarly, the TSLS estimator of $(1, -\boldsymbol{\beta}'_2)'$ is $(1/b_{1J}^{\mathrm{T}})\mathbf{b}_J^{\mathrm{T}}$. Anderson and Rubin (1950), p. 578, recommended selecting $J = 1$. Because of (2.31) the LIML and TSLS estimators have the same asymptotic distribution.

Let $\mathbf{P}_{Ie}$ be partitioned into one row and $G_I - 1$ rows as

$$\mathbf{P}_{Ie} = \begin{bmatrix} \mathbf{p}_{1e} \\ \mathbf{P}_{2e} \end{bmatrix}. \tag{2.36}$$

Then

$$(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})_{1\cdot} = \left[\begin{pmatrix} \mathbf{p}_{1e} \\ \mathbf{P}_{2e} \end{pmatrix}\mathbf{M}_{\perp\perp}(\mathbf{p}'_{1e}, \mathbf{P}'_{2e})\right]_{1\cdot} = \mathbf{P}_{2e}\mathbf{M}_{\perp\perp}(\mathbf{p}'_{1e}, \mathbf{P}'_{2e}), \tag{2.37}$$

and $\mathbf{b}^{\mathrm{T}} = (1, -\mathbf{b}_2^{\mathrm{T}'})'$; (2.33) for $J = 1$ is

$$\begin{aligned} \mathbf{0} = (\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})_{1\cdot}\mathbf{b} &= \mathbf{P}_{2e}\mathbf{M}_{\perp\perp}(\mathbf{p}'_{1e}, \mathbf{P}'_{2e})\begin{bmatrix} 1 \\ -\mathbf{b}_2 \end{bmatrix} \\ &= \mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{p}'_{1e} - \mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e}\mathbf{b}_2. \end{aligned} \tag{2.38}$$

The solution to (2.38) is

$$\mathbf{b}_2^{\mathrm{T}} = (\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e})^{-1}\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{p}'_{1e}. \tag{2.39}$$

We can write

$$\begin{aligned} \mathbf{b}_2^{\mathrm{T}} - \boldsymbol{\beta}_2 &= (\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e})^{-1}\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}(\mathbf{p}'_{1e} + \mathbf{P}'_{2e}\boldsymbol{\beta}_2) \\ &= (\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e})^{-1}\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\boldsymbol{\beta} \\ &= (\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e})^{-1}\mathbf{P}_{2e}\mathbf{M}_{\perp I}\boldsymbol{\beta} \\ &= (\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e})^{-1}\mathbf{P}_{2e}\frac{1}{T}\sum_{t=1}^{\mathrm{T}} \mathbf{z}_{\perp t}(\mathbf{z}'_{it}\overline{\boldsymbol{\Pi}}'_{Ii} + \mathbf{z}'_{\perp t}\boldsymbol{\Pi}'_{Ie} + \mathbf{V}'_{It})\boldsymbol{\beta} \\ &= (\mathbf{P}_{2e}\mathbf{M}_{\perp\perp}\mathbf{P}'_{2e})^{-1}\mathbf{P}_{2e}\frac{1}{T}\sum_{t=1}^{\mathrm{T}} \mathbf{z}_{\perp t}\mathbf{U}_{1t}. \end{aligned} \tag{2.40}$$

We have used (2.14), (2.13), and $\boldsymbol{\Pi}'_{Ie}\boldsymbol{\beta}_I = \mathbf{0}$.

Theorem 2 (*TSLS*). *Under the conditions of Theorem* 1

$$\sqrt{T}(\mathbf{b}_2^{\mathrm{T}} - \boldsymbol{\beta}_2) \xrightarrow{d} N[\mathbf{0}, \sigma_{11}(\boldsymbol{\Pi}_{2e}\mathbf{M}^0_{\perp\perp}\boldsymbol{\Pi}'_{2e})^{-1}], \tag{2.41}$$

where $\sigma_{11} = \boldsymbol{\beta}'_I\boldsymbol{\Omega}_{II}\boldsymbol{\beta}_I$.

Corollary 1 (*LIML*). *Under the conditions of Theorem* 1 *and normalization* $\boldsymbol{\beta}'_I\boldsymbol{\Phi}_{II}\boldsymbol{\beta}_I = 1$, *where* $\boldsymbol{\Phi}_{II}$ *is* (2.9),

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_I - \beta_I) \xrightarrow{d} N[\mathbf{0}, \sigma_{11}(\boldsymbol{\Pi}_{2e}\mathbf{M}^0_{\perp\perp}\boldsymbol{\Pi}'_{2e})^{-1}]. \tag{2.42}$$

The mathematics in Anderson and Rubin (1950) is more complicated than outlined here because a more general normalization $\boldsymbol{\beta}'_I\boldsymbol{\Phi}_{II}\boldsymbol{\beta}_I = 1$ was treated and because the asymptotic distribution of $\mathbf{P}$ was based on weaker conditions than those of Theorem 1. (See Assumption M of Section 4.) Let

$$\boldsymbol{\psi} = \boldsymbol{\Phi}_{II}\boldsymbol{\beta}_I. \tag{2.43}$$

Corollary 1 of Anderson and Rubin (1950) gave as the covariance matrix of the limiting normal distribution of $\sqrt{T}(\widehat{\boldsymbol\beta}-\boldsymbol\beta)$

$$\sigma_{11}(\mathbf{I}_{G_I}-\boldsymbol\beta\boldsymbol\psi')_{\cdot k}[(\boldsymbol\Pi_{Ie}\mathbf{M}_{\perp\perp}^{(0)}\boldsymbol\Pi_{Ie}')_{kk}]^{-1}(\mathbf{I}_{G_I}-\boldsymbol\psi\boldsymbol\beta')_{k\cdot}, \tag{2.44}$$

where $(\ )_{\cdot k}$ is the matrix with the kth column omitted, $(\ )_{k\cdot}$ with the kth row omitted, and $(\ )_{kk}$ is the matrix with the kth row and column omitted (for $\boldsymbol\beta_k \neq \mathbf{0}$). When $\boldsymbol\Phi$ is (2.9), $\boldsymbol\beta=(1,-\boldsymbol\beta_2')'$, and $k=1$, the vector $\boldsymbol\psi$ is $(1,\mathbf{0})'$ and (2.44) is

$$\sigma_{11}\begin{bmatrix}0 & \mathbf{0}\\ \mathbf{0} & (\boldsymbol\Pi_{2e}\mathbf{M}_{\perp\perp}^{0}\boldsymbol\Pi_{2e}')^{-1}\end{bmatrix}. \tag{2.45}$$

The lower right-hand submatrix of (2.45) is the covariance matrix in (2.41).

Another form of Eq. (2.38) and $\mathbf{g}^{\mathrm{T}}=\overline{\mathbf{P}}_{Ii}'\mathbf{b}^{\mathrm{T}}$, the estimator of $\boldsymbol\gamma_i$, is

$$\begin{aligned}
\mathbf{0} &= \begin{bmatrix}\mathbf{M}_{2z}\\ \mathbf{M}_{iz}\end{bmatrix}\mathbf{M}_{zz}^{-1}(\mathbf{M}_{zI},\mathbf{M}_{zi})\begin{bmatrix}\mathbf{b}^{\mathrm{T}}\\ -\mathbf{g}^{\mathrm{T}}\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{P}_{2i} & \mathbf{P}_{2e}\\ \mathbf{I}_{K_i} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{M}_{ii} & \mathbf{M}_{ie}\\ \mathbf{M}_{ei} & \mathbf{M}_{ee}\end{bmatrix}\begin{bmatrix}\mathbf{p}_{1i}' & \mathbf{P}_{2i}' & \mathbf{I}_{K_i}\\ \mathbf{p}_{1e}' & \mathbf{P}_{2e}' & \mathbf{0}\end{bmatrix}\begin{bmatrix}1\\ -\mathbf{b}_2^{\mathrm{T}}\\ -\mathbf{g}^{\mathrm{T}}\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{P}_{2i} & \mathbf{P}_{2e}\\ \mathbf{I}_{K_i} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{M}_{ii} & \mathbf{M}_{ie}\\ \mathbf{M}_{ei} & \mathbf{M}_{ee}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{2i}' & \mathbf{I}_{K_i}\\ \mathbf{P}_{2e}' & \mathbf{0}\end{bmatrix}\begin{bmatrix}-\mathbf{b}_2^{\mathrm{T}}\\ -\mathbf{g}^{\mathrm{T}}\end{bmatrix}\\
&\quad + \begin{bmatrix}\mathbf{P}_{2i} & \mathbf{P}_{2e}\\ \mathbf{I}_{K_i} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{M}_{ii} & \mathbf{M}_{ie}\\ \mathbf{M}_{ei} & \mathbf{M}_{ee}\end{bmatrix}\begin{bmatrix}\mathbf{p}_{1i}'\\ \mathbf{p}_{1e}'\end{bmatrix}.
\end{aligned} \tag{2.46}$$

The Eq. (2.38) for $\mathbf{b}^{\mathrm{T}}$ is found from (2.46) by elimination of $\mathbf{g}^{\mathrm{T}}$. Then $\mathbf{g}^{\mathrm{T}}$ is found by substitution of $\mathbf{b}^{\mathrm{T}}$ in (2.46).

The solution of (2.45) is

$$\begin{bmatrix}\mathbf{b}_2^{\mathrm{T}}\\ \mathbf{g}_{\mathrm{T}}\end{bmatrix} = \left\{\begin{bmatrix}\mathbf{P}_{2i} & \mathbf{P}_{2e}\\ \mathbf{I}_{K_i} & \mathbf{0}\end{bmatrix}\mathbf{M}_{zz}\begin{bmatrix}\mathbf{P}_{2i}' & \mathbf{I}_{K_i}\\ \mathbf{P}_{2e}' & \mathbf{0}\end{bmatrix}\right\}^{-1}\begin{bmatrix}\mathbf{P}_{2i} & \mathbf{P}_{2e}\\ \mathbf{I}_{K_i} & \mathbf{0}\end{bmatrix}\mathbf{M}_{zz}\mathbf{p}_{1z}'. \tag{2.47}$$

Under the conditions of Theorem 1

$$\sqrt{T}\begin{bmatrix}\mathbf{b}_2^{\mathrm{T}}-\boldsymbol\beta_2\\ \mathbf{g}_i^{\mathrm{T}}-\boldsymbol\gamma_i\end{bmatrix} \xrightarrow{d} N\left\{\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix},\ \sigma_{11}\left[\begin{pmatrix}\boldsymbol\Pi_{2i} & \boldsymbol\Pi_{2e}\\ \mathbf{I}_{K_i} & \mathbf{0}\end{pmatrix}\mathbf{M}_{zz}^{0}\begin{pmatrix}\boldsymbol\Pi_{2i}' & \mathbf{I}_{K_i}\\ \boldsymbol\Pi_{2e}' & \mathbf{0}\end{pmatrix}\right]^{-1}\right\}. \tag{2.48}$$

Corollary 1 of Anderson and Rubin (1950) gives the other covariances in the limiting distribution of $\sqrt{T}(\widehat{\boldsymbol\beta}_I-\boldsymbol\beta_I)$, $\sqrt{T}(\widehat{\boldsymbol\gamma}_i-\boldsymbol\gamma_i)$ as

$$\sigma(\widehat{\boldsymbol\beta}_I,\widehat{\boldsymbol\gamma}_i) = -\sigma_{11}(\mathbf{I}-\boldsymbol\beta\boldsymbol\psi')_{\cdot k}[(\boldsymbol\Pi\mathbf{M}_{zz}^{0}\boldsymbol\Pi')_{kk}]^{-1}(\boldsymbol\Pi_{Ii}+\boldsymbol\psi\boldsymbol\gamma')_{k\cdot}, \tag{2.49}$$

$$\sigma(\widehat{\boldsymbol\gamma}_i,\widehat{\boldsymbol\gamma}_i) = \sigma_{11}[(\boldsymbol\Pi_{Ii}'+\boldsymbol\gamma\boldsymbol\psi')_{\cdot k}[(\boldsymbol\Pi\mathbf{M}_{zz}^{0}\boldsymbol\Pi')_{kk}]^{-1}(\boldsymbol\Pi_{Ii}+\boldsymbol\psi\boldsymbol\gamma')_{k\cdot}+\mathbf{M}_{ii}^{-1}]. \tag{2.50}$$

These agree with (2.48).

3. Why was LIML favored over TSLS?

Why did Anderson and Rubin not propose the procedure now known as TSLS as an alternative to LIML, particularly since it is much easier to compute the TSLS estimator? The primary reason was the impression and belief that the maximum likelihood method yielded superior estimators. In fact, the development of methods for estimating all of the parameters of simultaneous equation models was based on using maximum likelihood and assuming normal distributions. Much of this research was carried out in 1945–46 at the Cowles Commission for Research in Economics. The work of members, such as Tjalling Koopmans, Roy Leipnik, Leo Hurwicz, and Trygve Haavelmo, followed this pattern. See the papers in *Statistical Methods in Dynamic Economic Models* (Koopmans, 1950).

The LIML estimation procedure can also be based on an analysis of variance approach as indicated in Section 4 of Anderson and Rubin (1949); this approach is attractive to statisticians. Even though the assumption of normally distributed disturbances may be suspect, the fact that the LIML takes into account the covariances of the errors was considered advantageous.

The TSLS estimator was used by Anderson and Rubin for the purpose of obtaining the asymptotic distribution of the LIML estimators. The fact that the TSLS and LIML estimators had the same asymptotic distribution was not in itself considered a sufficient reason for supporting the TSLS estimator.

There was another factor as well. The main thrust of the econometric modelling begun by Haavelmo was that the observed quantities (endogenous variables) were generated by the interaction of behavior described by several structural equations. As far as finding the solution of a set of such equations is concerned, the normalization was irrelevant and was considered incidental. In fact, writing the equations in the form of (2.35) was frowned on because it could be interpreted as suggesting regression of y_{1t} on $\mathbf{y}_{2t}$ and $\mathbf{z}_{it}$. Use of the TSLS estimator as outlined here treats the components of $\boldsymbol{\beta}_I$ asymmetrically and that was contrary to the point of view of simultaneous equations.

4. Derivation of the asymptotic distributions of the estimators by Anderson and Rubin

4.1. Distribution of $\widehat{\boldsymbol{\beta}}_I$

The asymptotic distribution of $\widehat{\boldsymbol{\beta}}_I$ given by (2.18) and (2.19) was derived in Section 4.2 of Anderson and Rubin (1950) for an arbitrary normalization $\boldsymbol{\beta}_I'\boldsymbol{\Phi}_{II}\boldsymbol{\beta}_I = 1$ for $\boldsymbol{\Phi}_{II}$ a constant symmetric matrix by exploiting (2.31). The first part of the derivation is proving that $\sqrt{T}v\mathbf{W}_{II} \xrightarrow{p} \mathbf{0}$. The basic idea is that of Section 2 of this paper, but the details are different.

Anderson and Rubin used the notation

$$\mathrm{d}\mathbf{y} = \mathbf{A}\,\mathrm{d}\mathbf{x} \tag{4.1}$$

to mean $\text{plim}_{T\to\infty}\mathbf{y}_T = \lim \mathbf{A}_T\,\text{plim}_{T\to\infty}\mathbf{x}_T, \mathbf{A}_T \to \mathbf{A}$

$$\sqrt{T}(\mathbf{y}_T - \underset{T\to\infty}{\text{plim}}\ \mathbf{y}_T) = \lim_{T\to\infty} \mathbf{A}_T\sqrt{T}(\mathbf{x}_T - \underset{T\to\infty}{\text{plim}}\ \mathbf{x}_T) + \text{o}_p(1). \tag{4.2}$$

Anderson and Rubin (1950)gave a general theorem to the effect that if $\mathbf{x}_T$ has an asymptotic normal distribution, then $\mathbf{y}_T$ has an asymptotic normal distribution, and the parameters of the two asymptotic distributions were defined. However, the proof was not given. A proof of a somewhat simpler version was given in Anderson (1963) and is stated in Appendix A.

For a $p \times p$ matrix $\mathbf{A} = (a_{ij})$

$$\text{d}|\mathbf{A}| = \sum_{i,j}^{p} A^{ij}\,\text{d}a_{ij}. \tag{4.3}$$

Let $s(\mathbf{A})$ and $\ell(\mathbf{A})$ denote the smallest and largest characteristic roots of the symmetric matrix $\mathbf{A}$.

Lemma 1. *For* $\mathbf{B}$ *positive definite and* $\mathbf{A}$ *positive semidefinite*

$$s(\mathbf{AB}^{-1}) \leqslant \frac{s(\mathbf{A})}{s(\mathbf{B})}. \tag{4.4}$$

Proof. Let $\mathbf{w}$ be a vector such that $\mathbf{w}'\mathbf{Aw}/\mathbf{w}'\mathbf{w} = s(\mathbf{A})$. Then

$$s(\mathbf{AB}^{-1}) = \min_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{Ax}}{\mathbf{x}'\mathbf{Bx}} \leqslant \frac{\mathbf{w}'\mathbf{Aw}}{\mathbf{w}'\mathbf{w}} \Big/ \frac{\mathbf{w}'\mathbf{Bw}}{\mathbf{w}'\mathbf{w}} \leqslant s(\mathbf{A}) \Big/ \min_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{Bx}}{\mathbf{x}'\mathbf{x}} = \frac{s(\mathbf{A})}{s(\mathbf{B})}. \qquad \square \tag{4.5}$$

Then

$$v = s(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\mathbf{W}_{II}^{-1}) \leqslant \frac{s(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})}{s(\mathbf{W}_{II})} \tag{4.6}$$

with probability approaching 1.

Lemma 2. *For a positive definite matrix* $\mathbf{C} = (c_{jk})$

$$|c_{jk}| \leqslant \ell(\mathbf{C}), \quad s(\mathbf{C}) \leqslant \min(c_{ii}). \tag{4.7}$$

To show $v\mathbf{W}_{II} \xrightarrow{p} 0$ we shall show that $s(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}) \xrightarrow{p} 0$. We have

$$s(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}) = \frac{|\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}|}{\mathbf{\Pi}_{i=2}^{G_I} ch_i(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie})}, \tag{4.8}$$

where $ch_1(\mathbf{A}) \leqslant \cdots \leqslant ch_{G_I}(\mathbf{A})$ denote the characteristic roots of $\mathbf{A}$. Since $\mathbf{\Pi}_{Ie}$ has rank $G_I - 1$, $ch_i(\mathbf{P}_{ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{ie})$ converges in probability to a positive constant, $i = 2, \ldots, p$. Then (4.7) and (4.8) imply that $v\mathbf{W}_{II} \xrightarrow{p} 0$.

Lemma 3.

$$\sqrt{T}(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}' - \mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}^0\mathbf{\Pi}_{Ie}') = \mathbf{P}_{Ie}^*\mathbf{M}_{\perp\perp}^0\mathbf{\Pi}_{Ie}' + \mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}^0\mathbf{P}_{Ie}^{*\prime} + o_p(1), \tag{4.9}$$

$$\sqrt{T}(|\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}'| - |\mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{\Pi}_{Ie}'|) \xrightarrow{p} 0. \tag{4.10}$$

Proof. Since $\mathbf{P} = \mathbf{\Pi} + (\frac{1}{\sqrt{T}})\mathbf{P}^*$,

$$\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}' = \mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}^0\mathbf{\Pi}_{Ie}' + \frac{1}{\sqrt{T}}\mathbf{P}_{Ie}^*\mathbf{M}_{\perp\perp}\mathbf{\Pi}_{Ie}' + \frac{1}{\sqrt{T}}\mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}^{*\prime} + \frac{1}{T}\mathbf{P}_{Ie}^*\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}^{*\prime}. \tag{4.11}$$

This implies (4.9). Use of (4.3), $|\mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{\Pi}_{Ie}'| = 0$,

$$(\mathbf{\Pi}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{\Pi}_{Ie}')^{kj} = \rho^k\beta_j = \rho^j\beta_k, \tag{4.12}$$

and $\boldsymbol{\beta}_I'\mathbf{\Pi}_{Ie} = \mathbf{0}$ yields $\sqrt{T}|\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}'| \xrightarrow{p} 0$. □

Note that (4.9) is a form of (4.15) of Anderson and Rubin (1950). Lemma 3 implies

$$\sqrt{T}|\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}'| \xrightarrow{p} 0. \tag{4.13}$$

In turn this implies $\sqrt{T}s(\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}') \xrightarrow{p} 0$. Lemmas 2 and 3 imply that $\sqrt{T}v \xrightarrow{p} 0$.

The maximum likelihood estimator of $\mathbf{\Pi}_{Ie}$ of rank $G_I - 1$ is

$$\widehat{\mathbf{\Pi}}_{Ie} = \left(\mathbf{I}_{G_I} - \frac{1}{\widehat{\boldsymbol{\beta}}_I'\mathbf{W}\widehat{\boldsymbol{\beta}}_I}\mathbf{W}_{II}\widehat{\boldsymbol{\beta}}_I\widehat{\boldsymbol{\beta}}_I'\right)\mathbf{P}_{Ie} = \left(\mathbf{I}_{G_I} - \frac{1}{\mathbf{b}'\mathbf{W}_{II}\mathbf{b}}\mathbf{W}_{II}\mathbf{b}\mathbf{b}'\right)\mathbf{P}_{Ie}. \tag{4.14}$$

A version of this is (5.17) of Anderson and Rubin (1949). The LIML estimator of $\boldsymbol{\beta}_I$ can be defined by the solution of $\widehat{\boldsymbol{\beta}}_I'\widehat{\mathbf{\Pi}}_{Ie} = \mathbf{0}$ and $\widehat{\boldsymbol{\beta}}_I'\mathbf{\Phi}_{II}\widehat{\boldsymbol{\beta}}_I = 1$.

Instead of (2.30), Anderson and Rubin (1950) used the following:

Theorem 3.

$$\sqrt{T}(\widehat{\mathbf{\Pi}}_{Ie}\mathbf{M}_{\perp\perp}\widehat{\mathbf{\Pi}}_{Ie}' - \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}') \xrightarrow{p} \mathbf{0}. \tag{4.15}$$

Proof. We have

$$\sqrt{T}(\widehat{\mathbf{\Pi}}_{Ie}\mathbf{M}_{\perp\perp}\widehat{\mathbf{\Pi}}_{Ie}' - \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}') = -\frac{\sqrt{T}}{\mathbf{b}'\mathbf{W}_{II}\mathbf{b}}\Bigg[\mathbf{W}_{II}\mathbf{b}\mathbf{b}'\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}' + \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}'\mathbf{b}\mathbf{b}'\mathbf{W}_{II} + \frac{1}{\mathbf{b}'\mathbf{W}_{II}\mathbf{b}}\mathbf{W}_{II}\mathbf{b}\mathbf{b}'\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}'\mathbf{b}\mathbf{b}'\mathbf{W}_{II}\Bigg].$$

From (2.18) we obtain $\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\mathbf{b} = v\mathbf{W}_{II}\mathbf{b}$ and $\mathbf{b}'\mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}\mathbf{b} = v\mathbf{b}'\mathbf{W}_{II}\mathbf{b}$. The right-hand side of (4.16) is

$$-\frac{3\sqrt{T}}{\mathbf{b}'\mathbf{W}_{II}\mathbf{b}}\, v\mathbf{W}_{II}\mathbf{b}\mathbf{b}'\mathbf{W}_{II} \xrightarrow{p} \mathbf{0}. \qquad \square \tag{4.16}$$

The last paragraph of page 577 of Anderson and Rubin (1950) may benefit by further elucidation. We shall write $\widehat{\beta}_i$ instead of $\widehat{\beta}^i$.

Let $\boldsymbol{\Theta} = \boldsymbol{\Pi}_{Ie}\mathbf{M}^0_{\perp\perp}\boldsymbol{\Pi}'_{Ie}$. Anderson and Rubin (1950) treat $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Pi}}_{Ie}\mathbf{M}_{\perp\perp}\widehat{\boldsymbol{\Pi}}'_{Ie}$ interchangeably with $\mathbf{T} = \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}'_{Ie}$. In this notation the parameter, LIML, and TSLS estimators are

$$\beta_i = \rho_J \Theta^{iJ}, \quad \widehat{\beta}_i = \widehat{\rho}_J \widehat{\Theta}^{iJ}, \quad b_i^{\mathrm{T}} = \widehat{\rho}_J^{\mathrm{T}} T^{iJ}, \tag{4.17}$$

$$\rho_J = \left(\sum_{i,j=1}^{G_I} \Theta^{iJ} \phi_{ij} \Theta^{iJ} \right)^{-1/2}, \quad \widehat{\rho}_J = \left(\sum_{i,j=1}^{G_I} \widehat{\Theta}^{iJ} \phi_{ij} \widehat{\Theta}^{iJ} \right)^{-1/2},$$

$$\widehat{\rho}_J^{\mathrm{T}} = \left(\sum_{i,j=1}^{G_I} T^{iJ} \phi_{ij} T^{jJ} \right)^{-1/2}. \tag{4.18}$$

From (4.18) we obtain

$$\beta_i^* = \sqrt{T}(\widehat{\beta}_i - \beta_i) = \rho_J \sqrt{T}(\widehat{\Theta}^{iJ} - \Theta^{iJ}) + \Theta^{iJ}\sqrt{T}(\widehat{\rho}_J - \rho_J) + \mathrm{o}_p(1). \tag{4.19}$$

Calculations will be made for $\sum_{i=1}^{G_I} \theta_{ki}\widehat{\beta}_i$. Note that $\sum_{i=1}^{G_I} \theta_{ki}\Theta^{iJ} = \delta_{kJ}|\boldsymbol{\Theta}| = 0$. Here $\delta_{JJ} = 1$ and $\delta_{kJ} = 0$, $k \neq J$. Let $\theta^*_{ki} = \sqrt{T}(\widehat{\theta}_{ki} - \theta_{ki})$. Then

$$\begin{aligned} 0 = \sum_{i=1}^{G_I} \widehat{\theta}_{ki}\widehat{\beta}_i &= \sum_{i=1}^{G_I} \left(\theta_{ki} + \frac{1}{\sqrt{T}}\theta^*_{ki} \right)\left(\beta_i + \frac{1}{\sqrt{T}}\beta^*_i \right) \\ &= \frac{1}{\sqrt{T}} \sum_{i=1}^{G_I} (\theta_{ki}\beta^*_i + \theta^*_{ki}\beta_i) + \mathrm{o}_p(1/\sqrt{T}). \end{aligned} \tag{4.20}$$

Thus

$$\boldsymbol{\Theta}\boldsymbol{\beta}^* = -\boldsymbol{\Theta}^*\boldsymbol{\beta} + \mathrm{o}_p(1), \tag{4.21}$$

where $\boldsymbol{\Theta}^* = \sqrt{T}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})$. The limiting distribution of $\boldsymbol{\Theta}\boldsymbol{\beta}^*$ is the limiting distribution of $-\boldsymbol{\Theta}^*\boldsymbol{\beta}$.

Let $\mathbf{T}^* = \sqrt{T}(\mathbf{T} - \boldsymbol{\Theta})$. Theorem 3 asserts that

$$\widehat{\boldsymbol{\Theta}}^* - \mathbf{T}^* = \sqrt{T}(\widehat{\boldsymbol{\Theta}} - \mathbf{T}) \xrightarrow{p} \mathbf{0}. \tag{4.22}$$

From (4.9)

$$\boldsymbol{\beta}'\mathbf{T}^* = \boldsymbol{\beta}'\mathbf{P}^*_{Ie}\mathbf{M}^0_{\perp\perp}\boldsymbol{\Pi}'_{Ie} + \mathrm{o}_p(1). \tag{4.23}$$

To find the limiting distribution of $\boldsymbol{\beta}'\mathbf{T}^*$ we use

$$\begin{aligned}\boldsymbol{\beta}'\mathbf{P}^*_{Ie} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\beta'\mathbf{v}_{It}\mathbf{z}'_{\perp t}\mathbf{M}^{-1}_{\perp\perp} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{1t}\mathbf{z}'_{\perp t}\mathbf{M}^{-1}_{\perp\perp} \\ &\overset{d}{\to} N[\mathbf{0},\sigma_{11}(\mathbf{M}^0_{\perp\perp})^{-1}],\end{aligned} \tag{4.24}$$

where $\sigma_{11}=\boldsymbol{\beta}'\boldsymbol{\Omega}_{II}\boldsymbol{\beta}$. Hence

$$\boldsymbol{\beta}'\mathbf{T}^* \overset{d}{\to} N(\mathbf{0},\mathbf{R}), \tag{4.25}$$

where $\mathbf{R}=\sigma_{11}\boldsymbol{\Theta}$. From (4.19) we obtain

$$\begin{aligned}\sqrt{T}(\widehat{\rho}_J-\rho_J) &= -\left(\sum_{k,j=1}^{G_I}\Theta^{kJ}\phi_{kj}\Theta^{iJ}\right)^{-3/2}\sum_{k,j=1}^{G_I}\Theta^{kJ}\phi_{kj}\sqrt{T}(\widehat{\Theta}^{jJ}-\Theta^{jJ})+\mathrm{o}_p(1) \\ &= -\rho_J^3\sum_{k,j=1}^{G_I}\Theta^{kJ}\phi_{kj}\sqrt{T}(\widehat{\Theta}^{jJ}-\Theta^{jJ})+\mathrm{o}_p(1).\end{aligned} \tag{4.26}$$

Then

$$\begin{aligned}\sum_{i=1}^{G_I}\psi_i\beta_i^* &= \sum_{i,j=1}^{G_I}\beta_j\phi_{ji}[\rho_J\sqrt{T}(\widehat{\Theta}^{iJ}-\Theta^{iJ})+\Theta^{iJ}\sqrt{T}(\widehat{\rho}_J-\rho_J)]+\mathrm{o}_p(1) \\ &= \sum_{i,j=1}^{G_I}\beta_j\phi_{ji}\left[\rho_J\sqrt{T}(\widehat{\Theta}^{iJ}-\Theta^{iJ})-\sum_{k,\ell=1}^{G_I}\beta_i\beta_k\phi_{k\ell}\sqrt{T}(\widehat{\Theta}^{\ell J}-\Theta^{\ell J})\right]+\mathrm{o}_p(1) \\ &= \sum_{i,j=1}^{G_I}\beta_j\phi_{ji}\rho_J\sqrt{T}(\widehat{\Theta}^{iJ}-\Theta^{iJ})-\sum_{k,l=1}^{G_I}\beta_k\phi_{k\ell}\sqrt{T}(\widehat{\theta}^{\ell J}-\Theta^{J})+\mathrm{o}_p(1) \\ &= \mathrm{o}_p(1).\end{aligned} \tag{4.27}$$

Let $\mathscr{E}\boldsymbol{\beta}^*\boldsymbol{\beta}^{*'}=\mathbf{Q}$. Then the following equations are to be solved for $\mathbf{Q}$:

$$\boldsymbol{\Theta}\mathbf{Q}\boldsymbol{\Theta}=\mathbf{R}, \tag{4.28}$$

$$\mathbf{Q}\boldsymbol{\psi}=\mathbf{0}. \tag{4.29}$$

We restrict ourselves to $\boldsymbol{\Phi}$ defined by (2.9); that is, $\boldsymbol{\beta}=(1,-\boldsymbol{\beta}_2')'$.

Since $\boldsymbol{\Theta}\boldsymbol{\beta}=\mathbf{0}$,

$$\boldsymbol{\Theta}=\begin{bmatrix}\boldsymbol{\beta}_2' \\ \mathbf{I}_{G_I-1}\end{bmatrix}\boldsymbol{\Theta}_{11}(\boldsymbol{\beta}_2,\mathbf{I}_{G_I-1}), \tag{4.30}$$

where $\boldsymbol{\Theta}_{11}$ is $\boldsymbol{\Theta}$ with the first row and column deleted. Partition $\mathbf{Q}$ as

$$\mathbf{Q}=\begin{bmatrix}q_{11} & \mathbf{q}_1' \\ \mathbf{q}_1 & \mathbf{Q}_{11}\end{bmatrix}. \tag{4.31}$$

Then

$$\mathbf{0} = \mathbf{Q}\boldsymbol{\psi} = \begin{bmatrix} q_{11} & \mathbf{q}_1' \\ \mathbf{q}_1 & \mathbf{Q}_{11} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} q_{11} \\ \mathbf{q}_1 \end{bmatrix}, \tag{4.32}$$

which implies

$$\mathbf{Q} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{11} \end{bmatrix}. \tag{4.33}$$

Then

$$\begin{aligned} \boldsymbol{\Theta}\mathbf{Q}\boldsymbol{\Theta} &= \begin{bmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_I-1} \end{bmatrix} \boldsymbol{\Theta}_{11}\mathbf{Q}_{11}\boldsymbol{\Theta}_{11}(\boldsymbol{\beta}_2, \mathbf{I}_{G_I-1}) \\ &= \mathbf{R} = \sigma_{11}\boldsymbol{\Pi}_{Ie}\mathbf{M}_{\perp\perp}\Pi_{Ie}' \\ &= \sigma_{11} \begin{bmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_I-1} \end{bmatrix} \boldsymbol{\Pi}_{2e}\mathbf{M}_{\perp\perp}\Pi_{2e}(\boldsymbol{\beta}_2, \mathbf{I}_{G_I-1}) \end{aligned} \tag{4.34}$$

because $\mathbf{0} = \boldsymbol{\beta}'\boldsymbol{\Pi} = \boldsymbol{\Pi}_{1e} - \boldsymbol{\beta}_2'\boldsymbol{\Pi}_{2e}$. The lower right-hand corner of (4.35) is

$$\boldsymbol{\Theta}_{11}\mathbf{Q}_{11}\boldsymbol{\Theta}_{11} = \sigma_{11}\boldsymbol{\Pi}_{2e}\mathbf{M}_{\perp\perp}\boldsymbol{\Pi}_{2e}'. \tag{4.35}$$

When $\boldsymbol{\Phi}$ is given by (2.9)

$$\boldsymbol{\psi} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}, \tag{4.36}$$

$$[\mathbf{I}_{G_I} - \boldsymbol{\beta}\boldsymbol{\psi}']_{\cdot 1} = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{\beta}_2 & \mathbf{I}_{G_I-1} \end{bmatrix}_{\cdot 1} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{G_I-1} \end{bmatrix}. \tag{4.37}$$

4.2. LIML and TSLS estimators in Anderson and Rubin (1950)

To summarize: let $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Pi}}_{Ie}\mathbf{M}_{\perp\perp}\widehat{\boldsymbol{\Pi}}_{Ie}'$ and $\mathbf{T} = \mathbf{P}_{Ie}\mathbf{M}_{\perp\perp}\mathbf{P}_{Ie}'$. Then (4.15) of Anderson and Rubin (1950) is

$$\sqrt{T}(\widehat{\boldsymbol{\Theta}} - \mathbf{T}) \xrightarrow{p} \mathbf{0}. \tag{4.38}$$

The LIML estimator of $\boldsymbol{\beta}_I$ is given by

$$\widehat{\boldsymbol{\beta}}_i = \widehat{\rho}_J\widehat{\boldsymbol{\Theta}}^{iJ}, \quad i = 1, \dots, G_I, \tag{4.39}$$

and any $J (J = 1, \dots, G_I)$, in particular for $J = 1$. In this notation the TSLS estimator is

$$\widehat{\beta}_i^{\mathrm{T}} = \widehat{\rho}_1^{\mathrm{T}} T^{i1}, \quad i = 1, \dots, G_I. \tag{4.40}$$

This statement was not made explicitly, only implied; however, the asymptotic distribution of $\sum_{i=1}^{G_I} \theta_{ki}\widehat{\beta}_i = -\sum_{i=1}^{G_I} \theta_{ki}\widehat{\beta}_i^{\mathrm{T}} + o_p$ is referred to (4.9).

4.3. Asymptotic distributions of the full set of coefficients

To derive the limiting distribution of $\sqrt{T}(\mathbf{b}^{\mathrm{T}} - \boldsymbol{\beta}_I)$ and $\sqrt{T}(\mathbf{g}^{\mathrm{T}} - \boldsymbol{\gamma}_i)$ Anderson and Rubin make the following additional assumptions:

A. $K_e \geqslant G_I - 1$.

F. $\mathbf{M}_{zz} \xrightarrow{p} \mathbf{R}$ nonsingular.

H. $\mathbf{M}_{vz} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{v}_t \mathbf{z}_t' \xrightarrow{p} \mathbf{0}$.

K. $\mathscr{E}|v_{it}|^{4+\lambda} < M$ for suitable $\lambda > 0$ and $M > 0$; exogenous variables are bounded; $\mathscr{E}\,\mathbf{v}_t = \mathbf{0}$; $\{\mathbf{v}_t\}$ are independent; and some components of $\mathbf{z}_t$ may be components of $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots$.

L. $\mathbf{\Phi}_{II}$ is constant.

M.

$$\frac{1}{T}\sum_{t=1}^{T} \mathscr{E} v_{it}\, v_{jt}\, z_{kt}\, z_{\ell t} \longrightarrow \kappa_{ijk\ell}. \tag{4.41}$$

Theorem 3 of Anderson and Rubin (1950) gave the covariances of the limiting distribution of $\sqrt{T}(\widehat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I)$ and $\sqrt{T}(\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)$ under conditions A, F, H, I, K, L, and M. Corollary 1 had the additional condition *N*.

N.

$$\frac{1}{T}\mathscr{E}\varepsilon_{1t}^2\, \mathbf{z}_t\, \mathbf{z}_t' \longrightarrow \sigma_{11}\mathbf{R}. \tag{4.42}$$

Conditions H and N are weaker than the usual condition that $\{\mathbf{v}_t\}$ and $\{\mathbf{z}_t\}$ are independent.

The covariances (2.44), (2.49), and (2.50) were based on Conditions A, F, H, I, K, L, M, and N. When Φ_{II} is defined by (2.9), these covariances are identical to the covariance matrix in (2.48).

5. Further history

In a letter to T.C. Koopmans in early 1946 M.A. Girshick estimated the structural coefficients in a just-identified system of two endogenous and two exogenous variables by estimating the reduced form by regression and then solving for the coefficients in each structural equation; it was an example of the instrumental variable method. For the over-identified case Anderson drew on his dissertation (Anderson, 1945) in which he obtained the maximum likelihood estimator of the equivalent of $\mathbf{\Pi}_{Ie}$ of assigned rank in an analysis of variance model.

Theil (1953a, b, 1954, 1961), suggested a two-step procedure, first estimating the coefficients of the reduced form by regression (as in (2.14)), and then estimating the coefficients of a structural equation by weighted least squares (as in (2.39)). This is essentially the TSLS procedure; it is a kind of instrumental variables estimator. Theil derived the asymptotic distribution of his TSLS estimator and showed that the difference between his TSLS estimator and the LIML estimator is $o_p(1/\sqrt{T})$.

Basmann (1957) developed the TSLS estimator by finding the "best" linear unbiased estimator of the coefficients. Sargan (1958) approached the problem in terms of instrumental variables; he also considered the least variance ratio estimator (as did Anderson and Rubin), which is the LIML estimator.

Appendix. Rubin's Theorem (Anderson, 1963)

Theorem on limiting distributions. *Let $F_n(\mathbf{u})$ be the cumulative distribution function of a random matrix $\mathbf{U}_n$. Let $\mathbf{V}_n$ be a (matrix-valued) function of $\mathbf{U}_n$, $\mathbf{V}_n = f_n(\mathbf{U}_n)$, and let $G_n(\mathbf{v})$ be the (induced) distribution of $\mathbf{V}_n$. Suppose $\lim_{n\to\infty} F_n(\mathbf{u}) = F(\mathbf{u})$ (in every continuity point of $f(\mathbf{u})$) and suppose for every continuity point $\mathbf{u}$ of $f(\mathbf{u})$, $\lim_{n\to\infty} f_n(\mathbf{u}) = f(\mathbf{u})$, when $\lim_{n\to\infty} \mathbf{u}_n = \mathbf{u}$. Let $G(\mathbf{v})$ be the distribution of the random matrix $\mathbf{V} = f(\mathbf{U})$, where $\mathbf{U}$ has the distribution $F(\mathbf{u})$. If the probability of the set of discontinuities of $f(\mathbf{u})$ according to $F(\mathbf{u})$ is 0, then*

$$\lim_{n\to\infty} G_n(\mathbf{v}) = G(\mathbf{v}) \tag{A.1}$$

(in every continuity point of $G(\mathbf{v})$).

References

Anderson, T.W., 1945. The non-central Wishart distribution and its application to problems in multivariate statistics. Ph.D. Dissertation, Princeton University.

Anderson, T.W., 1963. Asymptotic theory for principal component analysis. Annals of Mathematical Statistics 34, 122–148.

Anderson, T.W., Rubin, H., 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., Rubin, H., 1950. The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 21, 570–582.

Basmann, R.L., 1957. A generalized classical method of linear estimation of coefficients in a structural equation. Econometrica 25, 77–83.

Koopmans, T.C., 1950. Statistical Methods in Dynamic Economic Models. Wiley, New York.

Sargan, J.D., 1958. Estimation of economic relationships using instrumental variables. Econometrica 67, 557–586.

Theil, H., 1953a. Repeated least-squares applied to complete equation systems. Centraal Planbureau Memorandum.

Theil, H., 1953b. Estimation and simultaneous correlation in complete equation systems. Centraal Planbureau Memorandum. (Reprinted In: Raj, B., Koerts, J. (Eds.), 1992 Henri Theil's Contributions to Economics and Econometrics, Vol. 1, Kluwer, Dordrecht.)

Theil, H., 1954. Estimation of parameters in econometric models. Bulletin of the International Statistical Institute 34, 122–129.

Theil, H. (assisted by Cramer J.S., Moerman H, Russchen A.) 1961. Economic Forecasts and Policy (2nd rev. ed.), (North-Holland, Amsterdam). Contributions to Economic Analysis No. XV, (first published 1958).