A MINIMUM-DISTANCE INTERPRETATION
OF LIMITED-INFORMATION ESTIMATION

BY

A. S. GOLDBERGER and I. OLKIN

TECHNICAL REPORT NO. 44
JUNE 1, 1970

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP 17172

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A MINIMUM-DISTANCE INTERPRETATION
OF LIMITED-INFORMATION ESTIMATION

by
A. S. GOLDBERGER* and I. OLKIN
University of Wisconsin and Stanford University

TECHNICAL REPORT NO. 44
JUNE 1, 1970

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT 17172
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

*The research of A.S. Goldberger was supported in part by a grant from the Graduate School of the University of Wisconsin. This report is also being issued as a Social Systems Research Institute Workshop Paper, EME 7001, January 1970, University of Wisconsin.
1. **Introduction.**

For estimating the parameters of a single over-identified structural equation in a simultaneous equation model, the limited-information maximum-likelihood method was developed by Anderson and Rubin (1949). As is well-known (e.g., see Goldberger, (1964, p. 338)) this method amounts to estimating a relevant portion of the reduced form by minimizing the determinant of its residual moment matrix subject to the over-identifying restrictions. (This is provided that the underlying disturbances are normally distributed. In the absence of this proviso, the determinant-minimization algorithm produces "quasi-maximum-likelihood" estimates; cf. Chernoff and Rubin (1953). For simplicity, we ignore this qualification in the sequel.)

Focusing on the reduced form, one may view the situation as a multivariate linear regression model with constraints on the coefficients. From that vantage point, a variety of estimation procedures appear plausible. Many such procedures amount to minimizing some scalar function, other than the determinant, of the residual moment matrix $W$ (cf. Goldberger (1969)). In particular a minimum-distance method, suggested by Zellner's (1962) approach to efficient estimation of seemingly unrelated regressions, minimizes $\text{tr}(S^{-1}W)$ where $S$ is the residual moment matrix of unconstrained regression (e.g., Malinvaud (1966, pp. 281-290)).

In this paper, we show that minimization of $|W|$ is equivalent to minimization of $\text{tr}(S^{-1}W)$ in the present setting. We thus
provide a minimum-distance interpretation of single-equation limited-information maximum-likelihood estimation.

2. The Model.

Consider the single structural equation

\[ y_1 + x_1 \beta + u = 0, \]

where \( y \) is the \( T \times G \) matrix of observations on the endogenous variables which appear in the structural equation, \( x_1 \) is the \( T \times K_1 \) matrix of observations on the exogenous variables which appear in the structural equation, \( u \) is a \( T \times 1 \) vector of disturbances (independently and identically distributed with zero expectation and variance \( \sigma^2 \)) in the structural equation, and \( \gamma \) and \( \beta \) are unknown parameter vectors of dimensions \( G \times 1 \) and \( K_1 \times 1 \), respectively. It is understood that \( \gamma \) is normalized by having its first element set at \(-1\).

The structural equation is embedded in a conventional simultaneous equation model. Apart from the \( K_1 \) exogenous variables which appear in our structural equation, the model contains another \( K_2 \) exogenous variables whose observations are given in the \( T \times K_2 \) matrix \( x_2 \). The \( T \times (K_1 + K_2) \) matrix \( X = (x_1, x_2) \) has full column rank.

The only portion of the reduced form which concerns us in this paper is that which refers to the endogenous variables which appear
in our structural equation. We write it as

\begin{equation}
Y = \Pi Y + V = \begin{pmatrix} X_1 X_2 \end{pmatrix} \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix} + V.
\end{equation}

Here \( \Pi \), the \((K_1+K_2) \times G\) matrix of unknown reduced form coefficients, has been partitioned in conformity with the partitioning of \( X \), while \( V \) is the \( T \times G \) matrix of reduced form disturbances (independently and identically distributed with zero means and variance \( \sigma^2 \)).

Post-multiplying (2) by \( \gamma \), subtracting the result from (1), and taking expectations, yields the identity

\begin{equation}
X_1 (\beta + \Pi_1 \gamma) + X_2 \Pi_2 \gamma = 0.
\end{equation}

The familiar relations linking structural- and reduced-form parameters

\begin{equation}
-\Pi_1 \gamma = \beta,
\end{equation}

\begin{equation}
\Pi_2 \gamma = 0,
\end{equation}

follow from (3) and the full-rank assumption on \( X \). We suppose that the structural equation is over-identified so that \( K_2 \geq G \) and (5) constrains the reduced-form coefficient matrix \( \Pi \).

The setting is thus one of multivariate linear regression subject to (nonlinear) constraints. We seek to estimate \( \Pi \) subject to \( \Pi_2 \gamma = 0 \) for \( \gamma \neq 0 \). The estimated \( \Pi \) can then be inserted into (4)-(5) to produce estimates of \( \gamma \) and \( \beta \), upon normalization.
To avoid encumbering the notation we henceforth use the symbol \( \Pi \) to denote an argument or estimate. For an arbitrary choice of \( \Pi \), the reduced form residual moment matrix is

\[
W = (Y - X\Pi)' (Y - X\Pi) ,
\]

One particular choice of \( \Pi \) is

\[
P = (X' X)^{-1} X' Y ,
\]

which is obtained by unconstrained regression of \( Y \) on \( X \). Its residual moment matrix is

\[
S = (Y - XP)' (Y - XP) .
\]

(6)

Clearly,

\[
W = S + (\Pi - P)' (X' X)(\Pi - P) .
\]

(7)

The single-equation limited-information maximum-likelihood method chooses \( \Pi \) to minimize \(|W|\) subject to \( \Pi_2 \gamma = 0 \). Since \(|S^{-1}W| = |S|^{-1}|W|\) and \( |S| \) is a positive constant, we may as well say that the maximum-likelihood method chooses \( \Pi \) to minimize \(|S^{-1}W|\) subject to \( \Pi_2 \gamma = 0 \). Zellner's minimum-distance method chooses \( \Pi \) to minimize \( \text{tr}(S^{-1}W) \) subject to \( \Pi_2 \gamma = 0 \). To show
that the two methods are equivalent, it suffices to show that under the constraint, the product and the sum of the characteristic roots of $S^{-1}W$ are minimized by the same choice of $\Pi$.

3. Reduction to Canonical Form.

We start with a familiar orthogonalization process. Partitioning

$$ P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \quad \text{in conformity with} \quad \Pi = \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix}, $$

let

$$ P_{1 \cdot 2} = P_1 - (X_1' X_1)^{-1} X_1' X_2 P_2, \quad \Pi_{1 \cdot 2} = \Pi_1 - (X_1' X_1)^{-1} X_1' X_2 \Pi_2, $$

so that

$$ XP = X_1 P_1 + X_2 P_2 = X_1 P_{1 \cdot 2} + M_1 X_2 P_2, $$

$$ X\Pi = X_1 \Pi_1 + X_2 \Pi_2 = X_1 \Pi_{1 \cdot 2} + M_1 X_2 \Pi_2, $$

where

$$ M_1 = I - X_1 (X_1' X_1)^{-1} X_1'. $$

Then, since $M_1 = X_1'$ and $X_1' M_1 X_1 = 0$, we have

$$(8) \quad W = S + (\Pi_{1 \cdot 2} - P_{1 \cdot 2})' (X_1' X_1) (\Pi_{1 \cdot 2} - P_{1 \cdot 2}) + (P_2 - P_2)' (X_2' M_1 X_2) (P_2 - P_2).$$

Let $A$ be a $G \times G$ nonsingular matrix such that $AA' = S^{-1}$ and $A'SA = I$, and write $S^{-1}W$ as $AA'W$. The roots of $AA'W$ coincide with the roots of $A'WA = \overline{W}$, say. Pre- and post-multiply (8)
by $A'$ and $A$ respectively, which gives

$$(9) \quad \bar{W} = I + (\bar{\Pi}_2 - \bar{P}_2)' (\bar{\Pi}_2 - \bar{P}_2) + (\bar{\Pi}_1 - \bar{P}_1)' (\bar{\Pi}_1 - \bar{P}_1) ,$$

say, where

$$\bar{\Pi}_1 = B_1 \Pi_1 A \quad \quad \quad \quad \bar{\Pi}_2 = B_2 \Pi_2 A$$

$$\bar{P}_1 = B_1 P_1 A \quad \quad \quad \quad \bar{P}_2 = B_2 P_2 A ,$$

with $B_1$ being a $K_1 \times K_1$ nonsingular matrix such that $B_1^T B_1 = X_1^T X_1$, and $B_2$ being a $K_2 \times K_2$ nonsingular matrix such that $B_2^T B_2 = X_2^T M_1 X_2$.

Since $\bar{\Pi}_2 = B_2 \Pi_2 A$ is a nonsingular transformation of $\Pi_2$, the constraint $\Pi_2 \gamma = 0$ is equivalent to the constraint $\bar{\Pi}_2 \bar{\gamma} = 0$, where $\bar{\gamma} = A^{-1} \gamma$. The latter, in turn, is equivalent to $\bar{\Pi}_2 \bar{\delta} = 0$ where

$$\bar{\delta} = (\bar{\gamma}' \bar{\gamma})^{-1/2} \bar{\gamma} \quad \text{(so } \bar{\delta}' \bar{\delta} = 1).$$

Choosing $\bar{\Pi}$ to minimize a function of the roots of $S^{-1} W$ subject to $\Pi_2 \gamma = 0$ is, in other words, equivalent to choosing $\bar{\Pi} = \begin{pmatrix} \bar{\Pi}_1 \\ \bar{\Pi}_2 \end{pmatrix}$ to minimize the same function of the roots of $\bar{W}$ subject to $\bar{\Pi}_2 \bar{\delta} = 0$, with $\bar{\Pi}_1$ being unrestricted.

The fact that $\bar{\Pi}_1$ is unrestricted is the main point in the reduction to canonical form, for now minimization can be carried out in a direct manner. To see this, let

$$(10) \quad \bar{W}_0 = I + (\bar{\Pi}_2 - \bar{P}_2)' (\bar{\Pi}_2 - \bar{P}_2) ,$$
so that

$$\bar{W} = \bar{W}_0 + (\bar{P}_1 - \bar{P}_1)' (\bar{P}_1 - \bar{P}_1)$$

Since $\bar{W} - \bar{W}_0$ is nonnegative definite, it follows that $|\bar{W}| \geq |\bar{W}_0|^2$ and that $\text{tr}(\bar{W}) \geq \text{tr}(\bar{W}_0)$, with equality holding if and only if $\bar{P}_1 - \bar{P}_1 = 0$. The choice of $\bar{P}_1$ being unrestricted, it follows that to minimize either function of the roots, we must take $\bar{P}_1 = \bar{P}_1$. If this is done, then

$$\bar{W} = \bar{W}_0 = I + (\bar{P}_2 - \bar{P}_2)' (\bar{P}_2 - \bar{P}_2)'$$

the roots of which coincide with the roots of

$$\bar{W} = I + (\bar{P}_2 - \bar{P}_2)' (\bar{P}_2 - \bar{P}_2)'$$

apart from $K_2 - G$ unities. The unit roots do not affect the location of the minimum, so that maximum-likelihood chooses $\bar{P}_2$ to minimize $|\bar{W}|$ and minimum-distance chooses $\bar{P}_2$ to minimize $\text{tr}(\bar{W})$, both subject to $\bar{P}_2 \bar{s} = 0$.

For a given $\bar{s}$, with $\bar{s}' \bar{s} = 1$ we can find a $G \times (G-1)$ matrix $C$ such that $(\bar{s}, C)$ is orthogonal. Then, because $\bar{s}\bar{s}' + CC' = I$,

$$\bar{P}_2 = \bar{P}_2 I = \bar{P}_2 \bar{s}\bar{s}' + \bar{P}_2 CC'$$
If $\overline{\Pi}_2$ satisfies the constraint $\overline{\Pi}_2 \theta = 0$, then

$$\overline{\Pi}_2 = \theta \overline{C}'$$

where $\theta = \overline{\Pi}_2 \overline{C}$ is an unrestricted $K_2 \times (G-1)$ matrix.

Confining our attention to those $\overline{\Pi}$ which satisfy the restriction, we have

$$\overline{W} = I + (\theta \overline{C}' - \overline{P}_2) (\theta \overline{C}' - \overline{P}_2)'$$

as the canonical form.

4. **Solution.**

We continue conditional on $\theta$, and thus $\overline{C}'$, being given. The problem is to choose $\theta$ to minimize the determinant, or the trace, of $\overline{W}$ in (14). Let

$$\theta_0 = \overline{P}_2 \overline{C}'$$

and write

$$\theta \overline{C}' - \overline{P}_2 = \theta \overline{C}' - \overline{P}_2 (\theta \overline{C}' + \overline{C} \theta)$$

$$= -\overline{P}_2 \theta \overline{C}' + \theta \overline{C}' - \overline{P}_2 \overline{C} \theta$$

$$= -\overline{P}_2 \theta \overline{C}' + (\theta - \theta_0) \overline{C}'$$

(15)
Substituting (15) into (14) and using the orthogonality of \((\delta, c)\)
yields

\begin{equation}
\bar{W} = I + \bar{P}_2 \delta \bar{P}_2' + (\theta - \theta_0)(\theta - \theta_0)' \equiv \bar{W}_0 + (\theta - \theta_0)(\theta - \theta_0)'. 
\end{equation}

Since \(\bar{W} - \bar{W}_0\) is nonnegative definite, and \(\theta\) is unrestricted, it
follows that to minimize either function of the roots, we must take
\(\theta = \theta_0\).

Doing so, we have

\begin{equation}
\bar{W} = \bar{W}_0 = I + \bar{P}_2 \delta \bar{P}_2'. 
\end{equation}

Since

\[
\text{tr} \bar{W} = c + \delta' \bar{P}_2' \bar{P}_2 \delta, \quad \bar{W}_0 = 1 + \delta' \bar{P}_2' \bar{P}_2 \delta, 
\]

it is clear that both the maximum-likelihood and minimum distance
criteria are minimized by the same choice of \(\delta\). Retracing our steps,
this means that they choose the same \(\theta\), and hence the same \(\bar{\Pi}\). But
\(\bar{\Pi}\) is uniquely determined from \(\bar{\Pi}\), which completes the proof.

Remark. For \(\delta' \delta = 1\), the quantity \(\delta' \bar{P}_2' \bar{P}_2 \delta\) is minimized by the
normalized characteristic vector corresponding to the smallest root of

\[
\bar{P}_2' \bar{P}_2 = A' P_2' B_2' B_2 P A = A' P_2' X_2 M_2 X_2 P_2 A = A' S A, 
\]

9
where

\[ S_1 = P_2 X_2'M_2X_2'P_2 = Y'M_2X_2'X_2'M_2X_2^{-1}X_2'M_2Y, \]

which by standard multivariate regression theory, is the residual moment matrix which results when \( Y \) is regressed on \( X_1 \) alone. Thus we seek the smallest root \( \lambda \) and the corresponding normalized characteristic vector \( \delta \) of

\[ (A'S_1A^-1 - \lambda I) \delta = 0. \]

Pre- and post-multiplication of \( A'^{-1} \) and \( A^{-1} \) respectively gives

\[ (S_1 - \lambda S) \delta = 0, \]

since \( A'^{-1}A^{-1} = (AA')^{-1} = S \). This will be recognized as the usual characterization of limited-information maximum-likelihood estimation (cf. Goldberger (1964, p. 340)).

**Remark.** We have shown that the trace of \( S^{-1}W \) is, under the constraint, minimized by the \( \Pi \) which minimizes the determinant of \( S^{-1}W \). But many other functions of the roots share this property, as will be evident from our analysis if one bears in mind the following theorem; cf. Bellman (1960, pp. 113-115): Let \( \lambda_1^0 \geq \lambda_2^0 \geq \ldots \geq \lambda_N^0 \) denote the ordered roots of the \( N \times N \) positive definite matrix \( Q_0 \), let \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \) denote the ordered roots of the \( N \times N \) positive definite matrix \( Q \). If \( Q - Q_0 \) is nonnegative definite, then \( \lambda_i \geq \lambda_i^0 \), \( i=1,...,N \).
REFERENCES


