SAMPLE SIZE REQUIREMENTS IN FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATION

Bryan W. Brown

Econometric Research Program
Research Memorandum No. 222

February 1978

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

SAMPLE SIZE REQUIREMENTS IN FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATION

Bryan W. Brown

1. INTRODUCTION

The full-information maximum likelihood estimation (FIML) procedure for estimating the parameters of linear equation systems was developed at an early stage in the history of econometrics [5]. Despite the admirable properties that FIML is known to possess, the technique was seldom applied, due to the heavy computational burden involved in the solution of the FIML normal equations, which are nonlinear. The two-stage least squares (TSLS) and three-stage least squares (3SLS) techniques represent an attempt to develop estimators that share many of the desirable properties of FIML but avoid its computational complexity.

In recent years the increasing efficiency of electronic computers has greatly reduced the significance of this problem. This did not result in the widespread adoption of the FIML approach, however, since it was thought that the technique has excessive sample size requirements. Klein [3,4] has suggested that FIML estimation requires a full-rank data matrix and, hence, as many observations as there are variables in the system. In a large econometric model this can present a problem, since many of our economic time series are short relative to the number of variables in the system. As an alternative to FIML in the "under-sized"

sample, instrumental variable estimators for complete systems of equations have been proposed [1].

Some econometricians have disagreed with Klein's result, however, and argue that over-identifying restrictions can reduce the sample size requirements below the condition given by Klein. In the appendix of a recent contribution, Sargan [7] has demonstrated that Klein's condition is indeed correct when the system to be estimated is only subject to exclusion restrictions. This result will hold even if there are sufficient restrictions to overidentify the system. The belief persists, though, that more complicated restrictions might permit FIML estimation when the number of variables in the system exceeds the number of observations.

The purpose of this paper is to establish conditions on the data matrix for FIML estimation when the system is subject to linear homogeneous restrictions on the coefficients of single equations. In the following section, the basic model and terminology will be introduced. The just-identified case will be examined, in the third, and some general conditions which establish upper and lower limits on the rank of the data matrix will be set forth. In the fourth section, several sufficient conditions will be derived which demonstrate that overidentified systems can <u>sometimes</u> be estimated by FIML when the data matrix is less than full column rank. These sufficient conditions will form the basis for the development of several necessary conditions, in the fifth section. Finally, in the sixth section, we will summarize our results and discuss the possibilities for future research in this area.

2. MODEL AND DEFINITIONS

Consider the linear structural system

$$By_t + \Gamma z_t = u_t \qquad (t = 1,2,\ldots T) \tag{2.1}$$

where y_t is an Mx1 vector of endogenous variables at time t, z_t is a Kx1 vector of exogenous variables, u_t is an Mx1 vector of structural disturbances, B is the MxM matrix of coefficients for the endogenous variables, and Γ is a MxK matrix of coefficients of the predetermined variables. It will sometimes be convenient to represent the system

$$Ax_t = u_t \qquad (t = 1,2,\ldots,T)$$

where $x_t' = (y_t', z_t')$ is the 1x(M+K) vector of all variables at time t and $A = (B:\Gamma)$ is the complete Mx(M+K) matrix of coefficients. A more compact representation is

$$AX' = U'$$

where $X = (Y:Z)$ is the Tx(M+K) matrix of observations on all variables and U is the TxM matrix of structural disturbances for all periods.

In order that the endogenous vector y_t be unique for every predetermined vector z_t it is assumed that B is nonsingular or $|B| \neq 0$. Thus (2.1) may be premultiplied by B^{-1} to obtain the reduced form system

$$y_t = -B^{-1}\Gamma z_t + B^{-1}u_t$$

$$= \Pi z_t + v_t$$

or more compactly

$$Y' = \Pi Z' + V'$$

where $\Pi = -B^{-1}\Gamma$ is the MxK reduced form coefficient matrix, $v_t = B^{-1}u_t$ is the Mx1 vector of reduced form disturbances, and V is the TxM matrix of all reduced form disturbances.

The coefficients are assumed to satisfy certain linear homogeneous restrictions, which can be represented

$$\alpha_i'\Phi_i = 0 \qquad (i = 1,2,\ldots,M) \tag{2.2}$$

where the 1x(M+K) vector α_i' denotes the i-th row of A (i.e. the coefficients of the i-th equation) and Φ_i is a $(M+K)xR_i$ known matrix with each of the R_i rows corresponding to a homogeneous restriction on α_i. Of course the familiar exclusion restrictions can be written in this form. It is assumed that (2.2) provides sufficient restrictions to identify each equation.

The structural disturbance vectors u_t are assumed to be identically and independently distributed multivariate normal with zero mean vector and MxM covariance matrix Σ. The covariance matrix is assumed to be unrestricted, except that it be positive definite, which implies that (2.1) involves no identities or other exact relationships. By properties of the multi-variate normal distribution

$$v_t \sim \text{i.i.d. } N(0,\Omega)$$

and

$$y_t|z_t \sim N(\Pi z_t,\Omega) \tag{2.3}$$

where $\Omega = B^{-1}\Sigma B'^{-1}$ denotes the reduced form covariance matrix. Since Σ is unrestricted except to be positive definite, then Ω is also unrestricted except to be positive definite.

From (2.3), the joint probability distribution of the endogenous variables given the predetermined variables is

$$p(Y|Z);\Pi,\Omega) = \frac{1}{(2\pi)^{nT/2}|\Omega|^{T/2}} \exp\{-\frac{1}{2}\sum_{t=1}^{T}(y_t-\Pi z_t)'\Omega^{-1}(y_t-\Pi z_t)\}.$$

Thus, the log-likelihood function can be written

$$(2.4) \qquad L = \text{const.} - \frac{T}{2} \ell n|\Omega| - \frac{1}{2}tr\{\Omega^{-1}(Y'-\Pi Z')(Y-Z\Pi')\}$$

or equivalently

$$(2.5) \qquad L = \text{const.} - \frac{T}{2} \ell n|\Sigma| - \frac{1}{2}tr\{\Sigma^{-1}AX'XA'\} + \frac{T}{2}\ell n|B|^2$$

since $\Pi = -B^{-1}\Gamma$ and $\Omega = B^{-1}\Sigma B'^{-1}$. The maximum likelihood procedure is to find A and Σ which yield a maximum for (2.4) and hence (2.5), while satisfying the prior restrictions $|B| \neq 0$, $\alpha_i'\phi_i = 0$, and Σ positive definite.

Suppose that A and Σ maximize (2.5) while satisfying the prior restrictions, then so will $A^* = DA$ and $\Sigma^* = D\Sigma D'$ for any nonsingular MxM diagonal matrix D. But such a multiplication simply amounts to a rescaling of each equation, thus the need arises for a "normalization" rule in order to establish a scale for each equation. The most familiar procedure is to set the diagonal elements of B to unity. This approach can do more than rescale each equation, however, since it rules out $\beta_{ii} = 0$ which may be possible under the other prior restrictions. Accordingly, we will introduce the length normalization $\alpha_i'\alpha_i = 1$ $(i = 1,2,\ldots,M)$, which will only restrict the scale for any A and Σ satisfying the other prior restrictions.[1]

In order to simplify terminology we introduce the following definitions.

DEFINITION 2.1: An Mx(M+K) coefficient matrix $A = (B:\Gamma)$ is said to be <u>admissible</u> if and only if it satisfies all pertinent prior restrictions, namely (1) the homogeneous restriction $\alpha_i'\phi_i = 0$, (2) the normalization rules $\alpha_i'\alpha_i = 1$, and (3) $|B| \neq 0$.

DEFINITION 2.2: The Mx(M+K) matrix $\hat{A} = (\hat{B}:\hat{\Gamma})$ and the MxM matrix $\hat{\Sigma}$ are said to be maximum likelihood estimates of A and Σ if and only if $\hat{A}$ and $\hat{\Sigma}$ maximize the log-likelihood function (2.5) subject to the restriction that $\hat{A}$ be admissible and $\hat{\Sigma}$ be positive definite.

DEFINITION 2.3: The equation system (2.1) is said to be estimable if and only if there exist maximum likelihood estimates of A and Σ.

3. PRELIMINARY RESULTS

For an estimate of Σ to yield a maximum of (2.5) it must satisfy the first-order condition

$$0 = \frac{\partial L}{\partial \Sigma}$$

$$= -\frac{T}{2}\Sigma^{-1} - \frac{1}{2}\Sigma^{-1}(AX'XA')\Sigma^{-1}.$$

Solving this condition we find that Σ maximizes the log-likelihood only if

$$\Sigma = \frac{1}{T} AX'XA' \tag{3.1}$$

and is positive definite only if $|AX'XA'| \neq 0$. Provided $|AX'XA'| \neq 0$, then (3.1) may be substituted into (2.5) to obtain the concentrated log-likelihood function

$$L_c = \text{const.} - \frac{T}{2} \ell n \left|\frac{1}{T} AX'XA'\right| - \frac{T}{2} \ell n |B|^2 . \tag{3.2}$$

Thus A and Σ will maximize (2.5) subject to A being admissible and Σ positive definite if and only if A maximizes (3.2) subject to A being

admissible and $|AX'XA'| \neq 0$.

An important issue is whether admissible A which yields $|AX'XA'| = 0$ is a maximum likelihood estimate. It is sometimes argued, in such cases, that the log-likelihood is infinity and hence a maximum. Actually, neither (2.5) nor (3.2) can be evaluated when $|AX'XA'| = 0$, since $\Sigma = \frac{1}{T} AX'XA'$ is not invertible and $\ell n|\Sigma| = \ell n 0$ is undefined. Because such points cannot yield a maximum of the log-likelihood then we have ruled them out in this paper by imposing the prior restriction that Σ be positive definite.[2] In fact, as the following theorem shows, the existence of such a point not only rules out that point as a maximum of the log-likelihood, but all admissible A as well.

THEOREM 3.1: A necessary and sufficient condition for the system (2.1) to be estimable is that

$$|AX'XA'| \neq 0$$

for every admissible A.

PROOF: (Necessity). Assume to the contrary that A^0 is admissible but $|A^0X'XA^{0\prime}| = 0$. Let A^1 be a maximum likelihood estimate of A, then A^1 is admissible and $|A^1X'XA^{1\prime}| \neq 0$. Define $A^2 = \theta A^1 + (1-\theta)A^0$ for $0 \leq \theta \leq 1$, so A^2 will satisfy the homogeneous restrictions. Accordingly, let $\Sigma^0 = \frac{1}{T} A^0X'XA^{0\prime}$, $\Sigma^1 = \frac{1}{T} A^1X'XA^{1\prime}$, and $\Sigma^2 = \frac{1}{T} A^2X'XA^{2\prime}$.

Now, the elements of Σ^2 and hence $|\Sigma^2|$ are polynomials in θ and nonzero for $\theta = 1$ or $A^2 = A^1$, hence $|\Sigma^2| \neq 0$ for all θ except on a set of measure zero.[3] Since B^2 is linear in θ, then $|B^2|$ is also a polynomial in θ, and we see that $|B^2|$ and $|\Sigma^2|$ are both continuous in θ. Thus, by choosing θ sufficiently small we can find A^2 satisfying the homogeneous

restrictions such that $|B^2|$ is close to $|B^0|$ and $|\Sigma^0| \neq 0$. But this means we can find admissible A^2 with $|\Sigma^2| \neq 0$ which yields a log-likelihood exceeding that associated with A^1, which contradicts the assumption.

(Sufficiency). Suppose $|AX'XA'| \neq 0$ for every admissible A, then the log-likelihood associated with each admissible A will be non-zero and finite. This guarantees the existence of a maximum (perhaps not unique) of the log-likelihood subject to the restriction that A be admissible. END OF PROOF.

The above theorem, while very general, is somewhat difficult to apply since it requires knowledge of all admissible A, which usually is an infinite set. Thus, we are interested in developing conditions that do not require knowledge of A, which is the task of the remainder of this paper. First, we have the following rather straightforward corollaries, which result from the fact that $|AX'XA'| \neq 0$ if and only if $\rho(AX'XA') = M$.

COROLLARY 3.1: A necessary condition for the system (2.1) to be estimable is

$$\rho(X) \geq M.$$

COROLLARY 3.2: For (2.1) to be estimable it is necessary that

$$T \geq M.$$

Now $|AX'XA'| \neq 0$ if and only if $XA'\lambda \neq 0$ for all Mx1 vectors $\lambda \neq 0$, thus the system is estimable if and only if $XA'\lambda \neq 0$ for all admissible A and $\lambda \neq 0$. Suppose X is full column rank M+K or

equivalently $X\mu \neq 0$ for $\mu \neq 0$, then $XA'\lambda \neq 0$ for admissible A and $\lambda \neq 0$ since admissible A' has full column rank M or $\mu = A'\lambda \neq 0$ for $\lambda \neq 0$. Thus, we have the following result, which will prove to be the basis of all subsequent sufficient conditions.

COROLLARY 3.3: A sufficient condition for the estimability of (2.1) is that

$$\rho(X) = M+K$$

Corollaries 3.1 and 3.3 establish an upper and lower limit for the estimability of the system (2.1) in relation to the rank of X. In the present paper, we are interested in how knowledge about the identifying restrictions might eliminate the gap. In particular, we wish to examine the role of overidentifying restrictions and whether the imposition of overidentifying restrictions can reduce the upper limit as given by Corollary 3.3. The effect of overidentifying restrictions can best be studied after having established some basic results on just-identified systems.

As was discussed above, maximum likelihood estimates of A and Σ can be obtained by maximizing the reduced form log-likelihood (2.4) with respect to Π and Ω, subject to the restrictions that A be admissible. In the case of a just-identified system it is well known that there is a one-to-one relationship between restricted A and unrestricted Σ on the one hand and unrestricted Π and Ω on the other. Thus, the prior restrictions on the structural system imply no restrictions on the reduced form system and we can maximize (2.4) directly with respect to Π and Ω. If $\rho(Z) = K$, then we may solve the first-order conditions to obtain

$$\hat{\Pi}' = (Z'Z)^{-1}Z'Y \tag{3.3}$$

and

$$\hat{\Omega} = \frac{1}{T}(Y'-\hat{\Pi}X')(Y-Z\hat{\Pi}') \tag{3.4}$$

as estimates of Π and Ω. The estimate $\hat{\Pi}$ and the homogeneous restrictions imply $\hat{A} = (\hat{B}:\hat{\Gamma})$ uniquely and $\hat{\Sigma} = \hat{B}^{-1}\hat{\Omega}\hat{B}'^{-1}$ is also unique.

THEOREM 3.2: Suppose the system (2.1) is just-identified and $\rho(Z) = K$, then $T \geq M + K$ is a necessary condition for estimability.

PROOF: Let $\rho(Z) = K$, then substitution of (3.3) and (3.4) yields

$$\hat{\Omega} = \frac{1}{T}[Y'-Y'Z(Z'Z)^{-1}Z'][Y-Z(Z'Z)^{-1}Z'Y]$$

$$= \frac{1}{T}Y'[I_T-Z(Z'Z)^{-1}Z']Y$$

$$= \frac{1}{T}Y'DY.$$

Now $D = [I_T-Z(Z'Z)^{-1}Z']$ is symmetric idempotent which means $\rho(D) = tr(D) = T - K$. Thus, if $T < M + K$ then $\rho(D) < M$ whereupon $\rho(\hat{\Omega}) < M$ and $\rho(\hat{\Sigma}) < M$ since $\hat{\Sigma} = \hat{B}^{-1}\hat{\Omega}\hat{B}'^{-1}$. END OF PROOF.

It is clear that $X = (Y:Z)$ has full column rank M+K with unit probability when $\rho(Z) = K$ and $T \geq M + K$, since the elements of Y are stochastic and not perfectly correlated. By Theorem 3.2, if $\rho(Z) = K$ and the just-identified system is estimable, then $T \geq M + K$ and $\rho(X) = M + K$ with unit probability. Thus we have the following corollary in terms of the rank of X.

COROLLARY 3.4: Suppose the system (2.1) is just-identified and $\rho(Z) = K$, then with unit probability a necessary condition for estimability is $\rho(X) = M + K$.

An interesting question in the just-identified case is whether $\hat{\Pi}$ given by (3.2) is maximum likelihood when $T < M + K$ but $\rho(Z) = M$. Clearly, $|\hat{\Sigma}| = |\hat{\Omega}| = 0$ and the value of the likelihood function (2.5) is undefined. On the other hand, the estimate $\hat{\Pi}$ would be the same even if the true value of Ω were known. Thus, even though we cannot get a positive definite estimate of Σ we can treat $\hat{\Pi}$ and hence A as maximum likelihood estimates. This particular result is due to the fact that Σ does not enter into the estimation of A when the system is just-identified. When over-identifying restrictions are added, however, estimation of A requires an estimate of Σ and the condition for maximum likelihood estimation of A and Σ will be the same.

4. SUFFICIENT CONDITIONS

In the previous section we established that a sufficient condition for the system (2.1) to be estimable is that the data matrix (X) have full column rank. For just-identified systems this was also discovered to be a necessary condition for estimability. In this section we will develop sufficient conditions which demonstrate that overidentified systems may sometimes be estimable when $\rho(X) < M + K$. As will be shown, however, the system can always be reduced in such cases into a system where the data matrix (X*) of the transformed system is full column rank.

Suppose Φ_i, the restriction matrix for equation i, has rank R_i. Define an $(M+K) \times (M+K-R_i)$ matrix P_i such that

$$P_i'\Phi_i = 0 \tag{4.1}$$

and

$$\rho(\Phi_i) + \rho(P_i) = M + K \tag{4.2}$$

Now P_i' forms a basis for the null space of Φ_i, so $\alpha_i'\Phi_i = 0$ for nonzero α_i if and only if

$$\alpha_i = P_i\bar{\alpha}_i$$

for some $(M+K-R_i)x1$ vector $\bar{\alpha}_i$. Thus we have solved the homogeneous restrictions on each equation so as to represent all parameters in the equation (α_i) in terms of several unrestricted parameters ($\bar{\alpha}_i$).

A complete matrix A that will satisfy the homogeneous restrictions can be formed in a similar fashion. If $\alpha_i = P_i\bar{\alpha}_i$ for i=1,2,...,M then

$$\begin{aligned} A' &= (\alpha_1,\alpha_2,\ldots,\alpha_M) \\ &= (P_1\bar{\alpha}_1,P_2\bar{\alpha}_2,\ldots,P_M\bar{\alpha}_M) \\ &= (P_1:P_2:\ldots:P_M)\begin{pmatrix} \bar{\alpha}_1 & 0 & \ldots & 0 \\ 0 & \bar{\alpha}_2 & \ldots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \ldots & \bar{\alpha}_M \end{pmatrix} \\ &= P\bar{A}' \end{aligned}$$

where

$$P = (P_1:P_2:\ldots P_M)$$

and

$$\bar{A}' = \begin{pmatrix} \bar{\alpha}_1 & 0 & \cdots & 0 \\ 0 & \bar{\alpha}_2 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & \bar{\alpha}_M \end{pmatrix}$$

Clearly, a matrix A will satisfy the homogeneous restrictions on each equation if and only if $A' = P\bar{A}'$ for some $\bar{A}'$.[4]

THEOREM 4.1: A sufficient condition for the system (2.1) to be estimable is that $\rho(XP) = \rho(P)$.

PROOF: If $\rho(XP) = \rho(P) = M + K$, then $\rho(X) = M + K$ and by Lemma 3.1 the system is estimable. Thus we will suppose $\rho(XP) = \rho(P) = M+K^* < M+K$ or

$$P = QP^*$$

where P^* is formed from $M+K^*$ independent rows of P and Q is an $(M+K) \times (M+K^*)$ matrix. Let A be admissible and hence satisfy the homogeneous restrictions, then

$$\begin{aligned} XA'\lambda &= XPA'\lambda \\ &= XQP^*A'\lambda \\ &= X^*P^*A'\lambda \\ &= X^*A^{*\prime}\lambda \end{aligned}$$

where $X^* = XQ$ and $A^{*\prime} = P^*A'$. Now P^* is full row rank thus $\rho(X^*) = \rho(X^*P^*) = \rho(XP) = M+K^*$ and $X^*\mu^* \neq 0$ for any $(M+K^*)\times 1$ vector $\mu^* \neq 0$.[5] Since $|B| \neq 0$ for admissible $A = (B:\Gamma)$, then $A'\lambda = P\bar{A}'\lambda = QP^*\bar{A}'\lambda = QA^{*\prime}\lambda \neq 0$ and $\mu^* = A^{*\prime}\lambda \neq 0$ for admissible A and $\lambda \neq 0$. Thus $XA'\lambda = XP\bar{A}'\lambda = X^*P^*\bar{A}'\lambda = X^*A^{*\prime}\lambda \neq 0$ for all admissible A and $\lambda \neq 0$, or $|AX'XA'| \neq 0$

for all admissible A. END OF PROOF.

This result shows that (2.1) is sometimes estimable when $\rho(X) < M+K$, provided the homogeneous restrictions yield $\rho(XP) = \rho(P)$. The system can always be transformed in such a case, however, so that in some respect Theorem 3.3 holds. The M+K* variables of the transformed system are defined by $x_t^* = Q'x_t$ and form the Tx(M+K*) data matrix $X^* = XQ$. An Mx(M+K*) coefficient matrix for the new system, which obeys the original homogeneous restrictions, is defined by $A^* = \bar{A}P^{*\prime}$. In the preceding proof we demonstrated that $\rho(X^*) = M+K^*$ when $\rho(XP) = \rho(P)$, thus we see that Theorem 4.1 is really just an application of Theorem 3.3 to the transformed system.

Now $\rho(P) = M+K^*$ indicates the number of columns of admissible A that may have independent coefficients. Let A be admissible and hence satisfy the homogeneous restrictions, then

$$\begin{aligned} A &= \bar{A}P' \\ &= \bar{A}(P_B':P_\Gamma') \\ &= (B:\Gamma) \end{aligned}$$

for some $\bar{A}$, where P_B' denotes the first M columns of P' and P_Γ' the remainder. But $B = \bar{A}P_B'$ and hence P_B' have full column rank M, so we can rearrange and partition the columns of P' and hence Γ such that

$$\begin{aligned} A &= \bar{A}(P_B':P_{\Gamma 1}':P_{\Gamma 2}') \\ &= (B:\Gamma_1:\Gamma_2) \end{aligned}$$

where $(P_B':P_{\Gamma 1}')$ are $\rho(P) = M+K^*$ independent columns of P'. Because of columns of $P_{\Gamma 2}'$ are spanned by the columns of $(P_B:P_{\Gamma 1}')$ then

$$P'_{\Gamma 2} = (P'_B : P'_{\Gamma 1}) \begin{pmatrix} Q'_0 \\ Q'_1 \end{pmatrix}$$

and

$$\begin{aligned} \Gamma_2 &= \bar{A} P'_{\Gamma 2} \\ &= \bar{A}(P'_B : P'_{\Gamma 1}) \begin{pmatrix} Q'_0 \\ Q'_1 \end{pmatrix} \\ &= (B : \Gamma_1) \begin{pmatrix} Q'_0 \\ Q'_1 \end{pmatrix} \end{aligned}$$

where Q_0 and Q_1 are known matrices. Thus, admissible A has at most M+K* columns with independent coefficients, since Γ_2 is always determined by the choice of $(B:\Gamma_1)$.[6]

If the K predetermined variables are all independent and T > K, then we could expect $\rho(Z) = K$ and we can state a sufficient condition in terms of T. Let $P^{*\prime} = (P'_B : P'_{\Gamma 1})$ and accordingly

$$Q = \begin{pmatrix} I_M & 0 \\ 0 & I_{K^*} \\ Q_o & Q_1 \end{pmatrix}$$

Rearranging and partitioning Z conformably, then we have

$$\begin{aligned} X^* &= XQ \\ &= (Y : Z_1 : Z_2) \begin{pmatrix} I_M & 0 \\ 0 & I_{K^*} \\ Q_0 & Q_1 \end{pmatrix} \\ &= (Y + Z_2 Q_0 : Z_1 + Z_2 Q_1) \\ &= (Y^* : Z^*) \end{aligned}$$

and $A^* = (B:\Gamma_1)$. Suppose $Z = (Z_1:Z_2)$ is full column rank K, then

$$\rho(Z_1:Z_2)\begin{pmatrix} I_{K^*} \\ Q_1 \end{pmatrix} = \rho\begin{pmatrix} I_{K^*} \\ Q_1 \end{pmatrix} = K^*$$

or $\rho(Z^*) = K^*$. It is clear that the M columns of Y^* will be independent of any given T-M columns with unit probability. Thus, if $T \geq \rho(P) = M+K^*$, then $\rho(X^*) = \rho(X^*P^*) = M+K^*$ with unit probability and we have the following corollary.

COROLLARY 4.1: Suppose $\rho(Z) = K$, then with unit probability a sufficient condition for the system (2.1) to be estimable is that $T > \rho(P)$.

The preceding corollary is appropriate in the absence of extreme multicollinearity in the data matrix of predetermined variables, which is usually the case. When this matrix is less than full rank, then we may sometimes apply the following result.

THEOREM 4.2: A sufficient condition for the system (2.1) to be estimable is that

$$\rho(X) = \rho(Z) + M.$$

PROOF: Suppose $\rho(X) = \rho(Z) + M$, then the columns of Y are all independent of the columns of Z. Let μ_B be an Mxl vector and μ_Γ a Kxl vector, then

$$(Y:Z)\begin{pmatrix} \mu_B \\ \mu_\Gamma \end{pmatrix} \neq 0$$

for $\mu_B \neq 0$. Now (2.1) is estimable if and only if $|AX'XA'| \neq 0$ for all admissible A or equivalently $XA'\lambda \neq 0$ for all admissible A and Mx1 vectors $\lambda \neq 0$. But $|B| \neq 0$ and $\mu_B = B\lambda \neq 0$ for admissible $A = (B:\Gamma)$ and $\lambda \neq 0$, whereupon

$$(Y:Z)\begin{pmatrix} B \\ \Gamma \end{pmatrix}\lambda \neq 0$$

and (2.1) is estimable. END OF PROOF.

This result states that the system can be estimated when $\rho(X) < M+K$, provided the M columns of Y are mutually independent and also independent of the columns of Z. In this case, however, we do not really have M+K independent variables but only $\rho(X)$. Let $\rho(X) = \rho(Z) + M < M+K$ then $M+K - \rho(X)$ columns of Z (denoted Z_2) are spanned by the remaining $M - \rho(X)$ (denoted Z_1) or

$$Z_2 = Z_1 D$$

for some matrix D. Following a conformable partitioning of Γ we have

$$XA' = (Y:Z_1:Z_2)\begin{pmatrix} B' \\ \Gamma_1' \\ \Gamma_2' \end{pmatrix}$$

$$= (Y:Z_1)\begin{pmatrix} B' \\ \Gamma_1' + D'\Gamma'_2 \end{pmatrix}$$

Therefore we can define an equivalent transformed system with data matrix $X^+ = (Y:Z_1)$ and coefficient matrix $A^+ = (B:\Gamma_1 + \Gamma_2 D)$. Now $(Y:Z_1)$ is full column rank, and we see that this result is also just an application of Theorem 3.3 to a reduced system.[6]

5. NECESSARY CONDITIONS

In the preceding section we developed two conditions, $\rho(XP) = \rho(P)$ and $\rho(X) = \rho(Z) + M$, either of which is sufficient for the estimability of the system (2.1), even when the data matrix X is less than full rank. It was discovered in both cases, however, that the system could be transformed into a reduced system where the data matrix of the new system was full column rank. In this section, we will show that for the original system (2.1) to be estimable it is necessary that at least one of the two conditions must be met. Thus, in a certain respect, FIML estimation always requires a data matrix of full column rank.

THEOREM 5.1: With unit probability a necessary condition for (2.1) to be estimable is that

$$\rho(X) = \rho(Z) + M$$

or

$$\rho(XP) = \rho(P).$$

PROOF: Assume to the contrary that $\rho(X) < \rho(Z) + M$ and $\rho(XP) < \rho(P) \leq M+K$. Given these conditions, we will demonstrate the existence of A satisfying $|AX'XA'| = 0$ that are admissible with unit probability. Thus, we will show that the system (2.1) is not estimable when the conditions of the assumptions are not.

As in the proof of Theorem 4.1, for A satisfying the homogeneous restrictions we can write

$$\begin{aligned} XA'\lambda &= XP\bar{A}'\lambda \\ &= XQP^*\bar{A}'\lambda \\ &= X^*P^*\bar{\alpha} \end{aligned}$$

where X* and P* are the same as before but $\lambda' = (1,1,\ldots,1)$ and $\bar{\alpha}' = (\bar{\alpha}_1', \bar{\alpha}_2', \ldots, \bar{\alpha}_M')$. Since P* is full row rank, then $\rho(X^*) = \rho(X^*P^*) = \rho(XP) < \rho(P) = \rho(P^*)$ and $X^*\mu^* = 0$ for some $(M+K^*)x1$ vector $\mu^* \neq 0$. Moreover, $P^*\bar{\alpha} = \mu^*$ for some $\bar{\alpha} \neq 0$, whereupon

$$\begin{aligned} 0 &= X^*P^*\bar{\alpha} \\ &= XA'\lambda \end{aligned}$$

Thus we can find $A' = (P, \bar{\alpha}_1, P_2\bar{\alpha}_2, \ldots, P_M\bar{\alpha}_M)$, which will satisfy the homogeneous restrictions, and $\lambda' = (1,1,\ldots,1) \neq 0$ such that $XA'\lambda = 0$.

Since $\rho(X) < \rho(Z) + M$, then X may be partitioned $X = (Y_1 : Y_2 : Z_1 : Z_2)$ where $(Y_1 : Z_1)$ has $\rho(X)$ independent columns. Following a conformable partitioning of μ, then we may write

$$\begin{aligned} 0 &= X\mu \\ &= Y_1\mu_{11} + Y_2\mu_{12} + Z_1\mu_{21} + Z_2\mu_{22} \\ &= (Y_1 : Z_1)\mu_1 + (Y_2 : Z_2)\mu_2 \end{aligned}$$

where $\mu_1' = (\mu_{11}', \mu_{21}')$ and $\mu_2' = (\mu_{12}', \mu_{22}')$. If we discard all but $\rho(X)$ independent rows of X then the solution to this system may be represented

$$\begin{aligned} \mu_1 &= (Y_1 : Z_1)^{-1}(Y_2 : Z_2)\mu_2 \\ &= \frac{\text{adj } (Y_1 : Z_1)}{\det (Y_1 : Z_1)} (Y_2 : Z_2)\mu_2 \\ &= \frac{1}{\Delta} g_1 (Y; \mu_2) \end{aligned}$$

where $\Delta = \det (Y_1 : Z_1)$ and $g(\cdot)$ is a vector of polynomial functions in Y and the arbitrary parameters μ_2.

Let R be the right-inverse of Q, then $\mu^* = R\mu$ is a solution to $\mu = Q\mu^*$ and $X^*\mu^* = 0$ when $X\mu = 0$. Partitioning $P^* = (P_1^*:P_2^*)$, where P_1^* has $\rho(P^*)$ independent columns, and conformably $\bar{\alpha}' = (\bar{\alpha}_1':\bar{\alpha}_2')$ then

$$\begin{aligned}\mu^* &= P^*\bar{\alpha} \\ &= P_1^*\,\bar{\alpha}_1 + P_2^*\,\bar{\alpha}_2\end{aligned}$$

has the solution

$$\bar{\alpha}_1 = P_1^{*-1}(\mu^* - P_2^*\bar{\alpha}_2).$$

Thus, following substitution, a solution of $X^*P^*\bar{\alpha} = XP\bar{\alpha} = 0$ may be represented

$$\begin{aligned}(5.1)\qquad \bar{\alpha}_1 &= P_1^{*-1}(R\mu - P_2^*\bar{\alpha}_2) \\ &= \frac{1}{\Delta}\,P_1^{*-1}\left(R\begin{bmatrix} g_1(Y;\mu_2) \\ \mu_2 \end{bmatrix} - \Delta R_2^*\bar{\alpha}_2\right) \\ &= \frac{1}{\Delta}\,h_1(Y;\mu_2,\bar{\alpha}_2)\end{aligned}$$

where $h_1(\cdot)$ is a vector-valued polynomial function in Y and the parameters μ_2 and $\bar{\alpha}_2$.

Now $|B|$ is a polynomial in the elements of B. But B is linear in $\bar{\alpha}$ when $A' = (P_1\bar{\alpha}_1:P_2\bar{\alpha}_2:\ldots:P_M\bar{\alpha}_M)$, whereupon $|B(\bar{\alpha})|$ is a polynomial in the elements of $\bar{\alpha}$. By substitution, then

$$|B(\alpha)| = (\tfrac{1}{\Delta})^M d(Y;\mu_2,\bar{\alpha}_2)$$

when $XP\bar{\alpha} = 0$, where $d(\cdot)$ is a polynomical in Y and the parameters μ_2 and $\bar{\alpha}_2$.

Let $A^0 = (P_1\bar{\alpha}_1^0 : P_2\bar{\alpha}_2^0 : \ldots : P_M\bar{\alpha}_M^0)$ be admissible and $\lambda' = (1,1,\ldots,1)$ then $\mu^0 = A^{0\prime}\lambda^0$ and $\bar{\alpha}^{0\prime} = (\bar{\alpha}_1^{0\prime}, \bar{\alpha}_2^{0\prime}, \ldots, \bar{\alpha}_M^{0\prime})$. Clearly, we can choose one column of Y_2 such that

$$0 = (Y_1 : Z_1)\mu_1^0 + (Y_2 : Z_2)\mu_2^0$$

thus for $\mu_2 = \mu_2^0$, $\bar{\alpha}_2 = \bar{\alpha}_2^0$, and some Y_2, then (5.1) yields $\bar{\alpha} = \bar{\alpha}^0$ and hence $A = A^0$. Since $\Delta = \det (Y_1 : Z_1)$ is unchanged and $|B^0| \neq 0$, then $d(Y; \mu_2^0, \bar{\alpha}_2^0) \neq 0$ for some Y and hence all Y except on a set of measure zero. Now the elements of Y are continuous random variables and not perfectly correlated, so $|B| \neq 0$ with unit probability for $A = (B : \Gamma)$ determined by (5.1).[7] Thus we can find $\bar{\alpha}$ such that $XP\bar{\alpha} = 0$ and $|B(\bar{\alpha})| \neq 0$ with unit probability or equivalently $A' = (P_1\bar{\alpha}_1, P_2\bar{\alpha}_2, \ldots, P_M\bar{\alpha}_M)$ which satisfy the homogeneous restrictions with certainty, $|B| \neq 0$ with unit probability, and $XA'\lambda = 0$ for $\lambda' = (1,1,\ldots,1)$. END OF PROOF.

Suppose $\rho(Z) = K$ then $\rho(X) = \rho(Z) + M$ implies X is full column rank M+K, whereupon $\rho(XP) = \rho(P)$ and we have the following corollaries.

COROLLARY 5.1: Let $\rho(Z) = K$, then with unit probability $\rho(XP) = \rho(P)$ is necessary for the system (2.1) to be estimable.

COROLLARY 5.2: Assume $\rho(Z) = K$, then with unit probability a necessary condition for the estimability of (2.1) is that $T \geq \rho(P)$.

Combining Corollary 5.2 with Corollary 4.1, we discover that in the absence of multicollinearity among the predetermined variables $T \geq \rho(P)$ is a necessary and sufficient condition with unit probability.

A rather obvious consequence of Theorem 5.1 is the following result.

COROLLARY 5.3: Suppose $\rho(X) < \rho(Z) + M$, then with unit probability

$$\rho(XP) = \rho(P)$$

is necessary for the system (2.1) to be estimable.

Sargan studied estimability when the homogeneous restrictions are of the exclusion type. In this case, he showed that $\rho(X) = M+K$ is a necessary condition (with unit probability) if Z has at least T-M+1 independent columns when $T < M+K$. But $\rho(P) = M+K$ for the exclusion restriction case, as will be shown below, and $\rho(Z) \geq T - M + 1$ implies $\rho(X) < \rho(Z) + M$, since X has only M more columns than Z. Thus, the above corollary is a generalization of Sargan's result to the homogeneous case.

When the homogeneous restrictions are in fact of the exclusion type, then we have the following simple necessary condition, which is less restrictive than Sargan's condition.

THEOREM 5.2: Suppose all homogeneous restrictions are of the exclusion type, then with unit probability a necessary condition for the estimability of (2.1) is that

$$\rho(X) = \rho(Z) + M.$$

PROOF: Assume that all homogeneous restrictions are all of the exclusion type, then Φ_i will have a column e_j for each α_{ij} restricted to be zero. Consequently, since $P_i'\Phi_i = 0$ and $\rho(P_i) + \rho(\Phi_i) = M+K$, then P_i can be formed with a column e_k for each α_{ik} that is <u>not</u> restricted to be zero. No column of A can be entirely zero restricted, so $(e_1, e_2, \ldots, e_{M+K})$ must all be included as columns of $P = (P_1 : P_2 : \ldots : P_M)$.

Thus $\rho(X) = \rho(Z) + M$ will be met when $\rho(XP) = \rho(P)$, since P will be full row rank M+K and $\rho(XP) = \rho(X)$. END OF PROOF.

Sargan's result follows as a simple corollary to this theorem.

COROLLARY 5.4: Suppose that all homogeneous restrictions are of the exclusion type, $T < M+K$, and $\rho(Z) \geq T - M + 1$, then with unit probability the system (2.1) is not estimable.

PROOF: Let $T < M+K$ and $\rho(Z) \geq T - M + 1$, then $\rho(X) < \rho(Z) + M$ since X has only M more columns than Z. But when all homogeneous restrictions are of the exclusion type, then by Theorem 4.4, $\rho(X) = \rho(Z) + M$ is necessary. END OF PROOF.

6. CONCLUSION

Klein has asserted that FIML estimation of the linear simultaneous equation model requires that the data matrix be full column rank. When the identifying restrictions are of the exclusion type, Sargan has shown that this conjecture is indeed correct. In the more general case of linear coefficient restrictions on single equations, we proved that FIML estimation is sometimes possible when the data matrix is less than full column rank. We also showed, however, that this can only occur when the system can be transformed into an equivalent system with fewer variables and the data matrix of the reduced system is full column rank.[8] In a very real sense, then, Klein's assertion is still correct in the more general case.

Of course when the system (2.1) is subjected to more complicated prior restrictions, the results obtained above need not apply. In the

case of linear restrictions on coefficients of more than one equation, simple examples can be constructed which allow FIML estimation when the data matrix is less than full rank and cannot be transformed to be so.[9] Other restrictions which might relax the data matrix rank requirement are nonlinear coefficient restrictions and constraints on the coveriance matrix. A more careful treatment should be undertaken to determine the precise relationship between the rank of X and estimability when the system is subject to these more complicated restrictions.

Perhaps the most frequently occuring extension of (2.1) is the model which is linear in the parameters but nonlinear in the variables. The system can then be written

$$Aq(y_t,x_t) = u_t \qquad (t = 1,2,\ldots,T)$$

where $q(\cdot)$ is a vector of (possibly nonlinear) functions of the endogenous and exogenous variables. As before, the question is whether the system can be estimated by FIML when the data matrix of observations $Q' = (q(y_1,z_1), \quad q(y_2,z_2),\ldots,q(y_T,z_T))$ is less than full rank. It is my conjecture that this is indeed possible, provided, as in the linear model, the system can be transformed such that the data matrix of the transformed variable is full rank. This case also deserves an extended analysis to verify or reject this conjecture.

FOOTNOTES

[1]The results given below are unchanged if we adopt the normalization $\beta_{ii} = 1$, however, the development is more straightforward with the length normalization.

[2]Sargan argues that the maximum likelihood estimator is unidentifiable when admissible A yields $|AX'XA'| \neq 0$. Rothenberg [6] has shown that the model is locally unidentifiable at some point (A,Σ) if the information matrix is singular at that point. But when $\Sigma = \frac{1}{T} AX'XA'$ is singular it is easy to show that the information matrix evaluated at (A,Σ) is also singular. Thus, if one accepts the argument that admissible A and singular $\Sigma = \frac{1}{T} AX'XA'$ are maximum likelihood estimates, then we have Sargan's conclusion.

[3]Fisher [2] has shown that if an analytic function is nonzero somewhere in its domain, then it is nonzero for all values of its argument except on a set of measure zero.

[4]There is sometimes a choice as to which parameters will be unrestricted and which will be eliminated (i.e. represented as linear combinations of the unrestricted parameters). Fortunately P_i must span the same column space and hence $P = (P_1:P_2:\ldots:P_M)$ and XP must have the same rank regardless of the choice of parameters to be eliminated for each equation.

[5]Suppose D is full row rank, then we can partition $D = (D_1:D_2)$ and $CD = (CD_1:CD_2)$ where D_1 is square nonsingular. Now $\rho(CD) \leq \rho(D)$ and $\rho(CD_1:CD_1) \geq \rho(CD_1)$. But $\rho(CD_1) = \rho(C)$ when C_1 is square nonsingular thus $\rho(CD) = \rho(C)$.

[6]It would be misleading, however, to conclude that multicollinearity buys us degrees of freedom for nothing. In order for the system to be identified in the presence of a collinear predetermined variable we must impose additional "over-identifying" restrictions.

[7]By the same arguments we could show that $\beta_{ii} \neq 0$ with unit probability, thus, we could satisfy the normalization rule $\beta_{ii} = 1$ with unit probability.

[8]No gain is achieved over other estimators such as TSLS and 3SLS, however, since we could also apply them to the transformed system.

[9]Consider (2.1) when M = 3, K = 3 and

$$A = \begin{pmatrix} 1 & \beta & 0 & \gamma & 0 & 0 \\ 0 & 1 & \beta & 0 & \gamma & 0 \\ \beta & 0 & 1 & 0 & 0 & \gamma \end{pmatrix}$$

where we have imposed the normalization $\beta_{ii} = 1$. The system will be estimable if $XA'\lambda \neq 0$ for all β, γ and $\lambda \neq 0$ or $\lambda'\lambda = 1$. Suppose $\rho(X) = T = 5$, then $XA'\lambda = 0$ supplies 5 equations and $\lambda'\lambda = 1$ another. Thus, we have 6 equations (which will generally be independent) in only 5 unknowns (β, γ, λ_1, λ_2, λ_3), which cannot be solved.

REFERENCES

[1] Brundy, J. and D.W. Jorgenson, "Efficient Estimation of Simultaneous Equation Systems by Instrumental Variables," Review of Economics and Statistics, 53(1971), 207-224.

[2] Fisher, F.M., The Identification Problem in Econometrics. New York: McGraw-Hill, 1966.

[3] Klein, L.R., "Forecasting and Policy Evaluation Using Large Scale Econometric Models: The State of the Art," in M.D. Intrilligator (ed.), Frontiers of Quantitative Economics. Amsterdam: North-Holland, 1971.

[4] Klein, L.R. and M. Nakamura, "Singularity in the Equation Systems of Econometrics: Some Aspects of the Problem of Multicollinearity," International Economic Review, 3(September 1963), 274-299.

[5] Koopmans, T.C., H. Rubin and R.B. Leipnik, "Measuring the Equation Systems of Dynamic Economics," in Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10. New York: John Wiley and Sons, 1950.

[6] Rothenberg, T.J., "Identification in Parametric Models, "Econometrica,39(1971), 557-592.

[7] Sargan, J.D., "Asymptotic Theory and Large Models," International Economic Review, 16(February 1975), 75-91.