Identification and Lullback Information in the GLSEM

by
Phoebus J. Dhrymes, Columbia University

September 1993

Discussion Paper Series No. 667

Identification and Kullback Information in the GLSEM*

Phoebus J. Dhrymes

Columbia University

originated 08/31/93

Abstract

In this paper we derive very succinctly the necessary and sufficient (nas) conditions for identification in the general linear structural econometric model (GLSEM) by use of the **Kullback information** apparatus.

1 Introduction

The purpose of this paper is to intoroduce more widely, in econometrics, the use of **Kullback information**. We do so in the context of the standard GLSEM, by showing how the identification problem becomes almost a routine by product of the convergence properties of the (log) likelihood function (LF).

2 Formulation of the Problem and Notation

Consider the standard GLSEM

$$YB^* = XC + U, \quad \text{or} \quad ZA^* = U, \quad A^* = (B^{*\prime}, C^\prime)^\prime, \quad Z = (Y, X), \quad (1)$$

* This is a preliminary version and is not to be quoted except with the explicit permission of the author

where Y is $T \times m$, X is $T \times G$ and contain, respectively, the current endogenous and predetermined variables of the system; evidently, B^* and C are $m \times m$, $G \times m$, respectively, and contain the unknown parameters of the model; U is the $T \times m$ matrix of the "structural" errors whose rows are taken to be i.i.d., with[1]

$$E(u_{t\cdot}^{'}) = 0, \quad Cov(u_{t\cdot}^{'}) = \Sigma > 0.$$

In this context it is customary to impose

Convention 1. In the i^{th} equation it is possible to, and we do, set the coefficient of y_i equal to unity.

The convention above allows us to rewrite Eq. (1) as

$$Y = YB + XC + U = ZA + U, \tag{2}$$

where

$$A = \begin{pmatrix} B \\ C \end{pmatrix}, \quad b_{ii} = 0, \quad i = 1, 2, \ldots, m. \tag{3}$$

We shall not be very exacting about the assumptions made regarding the presence or absence of lagged dependent variables, since we do not focus on the distributional aspects of the problem and, at any rate, these problems and their solution are, by now, rather well known.[2]

In this context, "identification" is obtained by "exclusion restrictions", although, of course, more general schemes are possible; the latter is easily incorporated in our framework, although for simplicity of exposition we shall operate with the "exclusions" option. Consequently, we have

Convention 2. In the i^{th} equation there are m_i ($\leq$ $m - 1$), and G_i ($\leq$ G) "explanatory" variables, **which are endogenous and predetermined, respectively.**

In order to implement this convention, we introduce the device of selection matrices,[3] as follows. Let L_{1i}, be a permutation of m_i of the

[1] The simplicity of this specification is retained so as to have exact correspondence with the historical evolution of this subject.

[2] The requisite central limit theorems (CLT) for solving the distributional problems in the static or dynamic (GLSEM) models, or models with autoregressive errors are given, respectively, in Dhrymes (1989), Ch. 4, pp. 257ff, and Ch.5, pp. 323ff. All of the distributional results asserted herein remain valid even if the model is dynamic (but stable) with i.i.d. structural errors. The minimum requirement is that $(1/\sqrt{T})\sum_{t=1}^{T}(I \otimes x_{t\cdot})^{'}u_{t\cdot}^{'}$ should obey a martingale difference CLT with a Lindeberg condition (Ch 5. pp 323ff).

[3] The device of selection matrices was first introduced, in this context, by Dhrymes

columns of the identity matrix I_m, and L_{2i}, a permutation of G_i of the columns of I_G, such that

$$YL_{1i} = Y_i, \qquad XL_{2i} = X_i, \; i = 1, 2, \ldots, m. \tag{4}$$

Giving effect to Convention 2, the i^{th} equation may be written as

$$y_{\cdot i} = Y_i \beta_{\cdot i} + X_i \gamma_{\cdot i} + Y_i^* \beta_{\cdot i}^* + X_i^* \gamma_{\cdot i}^* + u_{\cdot i}, \qquad i = 1, 2, \ldots, m, \tag{5}$$

where the notation $y_{\cdot i}$, $u_{\cdot i}$ means the i^{th} column of Y and U, respectively, and $\beta_{\cdot i}$, $\gamma_{\cdot i}$ contain, respectively, the elements in the i^{th} column of $B(b_{\cdot i})$ and $C(c_{\cdot i})$ not known a priori to be zero. Evidently, $\beta_{\cdot i}^*$ and $\gamma_{\cdot i}^*$ represent the elements of the two columns, respectively, set to zero by the prior restrictions. It follows immediately that

$$b_{\cdot i} = L_{1i}\beta_{\cdot i}, \qquad c_{\cdot i} = L_{2i}\gamma_{\cdot i}, \qquad L_{1i}^{'} b_{\cdot i} = \beta_{\cdot i}, \qquad L_{2i}^{'} c_{\cdot i} = \gamma_{\cdot i}. \tag{6}$$

Define

$$L_i = \begin{bmatrix} L_{1i} & 0 \\ 0 & L_{2i} \end{bmatrix}, \quad L_i^* = \begin{bmatrix} L_{1i}^* & 0 \\ 0 & L_{2i}^* \end{bmatrix}, \; i = 1, 2, \ldots, m, \tag{7}$$

and note that the i^{th} column of A, in Eq. (3) is given by

$$a_{\cdot i} = \begin{pmatrix} b_{\cdot i} \\ c_{\cdot i} \end{pmatrix}, \; i = 1, 2, \ldots, m.$$

The unknown structural parameters of the i^{th} equation are rendered, in this notation, as

$$\delta_{\cdot i} = L_i^{'} a_{\cdot i}, \quad i = 1, 2, \ldots m, \tag{8}$$

and for the system as a whole we have

$$\delta = L^{'} a, \quad a = \text{vec}(A), \quad \text{where } L = \text{diag}(L_1, L_2, \ldots, L_m). \tag{9}$$

Finally, we append the following standard assumptions:

A1. The error process $\{u_{t\cdot}^{'} : t \geq 1\}$ is a sequence of i.i.d. random vectors distributed as $N(0, \Sigma)$, $\Sigma > 0$.

A2. If the GLSEM is dynamic, it is stable in the sense that the roots of its characteristic equation lie outside the unit circle (no unit roots).

A3. The exogenous variables of the system lie in a compact subset $\Xi \subset R^s$.

(1973). Greater detail regarding their meaning and function may be found in that reference, as well as in Dhrymes (1978).

A4. The parameter space, $\Theta \subset R^k$ is compact, i.e. the admissible values of the elements of A^* and Σ lie in a compact set.

We may thus write the likelihood function of the observations as

$$L^*(\theta) = (2\pi)^{-(mT/2)}|\Sigma|^{-(T/2)}|B^{*\prime}B^*|^{(T/2)}\exp\left(-\frac{T}{2}\right)\mathrm{tr}\Sigma^{-1}S, \quad \text{where} \tag{10}$$

$$S = \frac{1}{T}A^{*\prime}\tilde{M}_{zz}A^*, \quad \tilde{M}_{zz} = \frac{1}{T}Z^{\prime}Z, \quad \theta = (\mathrm{vec}(A^*)^{\prime}, \mathrm{vec}(\Sigma)^{\prime})^{\prime} \tag{11}$$

and a zero subscript (or superscript) will indicate the true parameter vector.

3 Kullback Information and Minimum Contrast (MC) Estimators

3.1 Kullback Information

In the framework created in the previous section, the probability space(s) indexed on the parameter θ will be termed an **econometric model.** Basically, this is the probability space $(\Omega, \mathcal{A}, \mathcal{P}_\theta)$ which is induced by the probability space of the error process indexed on the parameter θ, given the space of the exogenous variables Ξ. To avoid excessive notation we suppress the latter space.[4] We have[5]

Definition 1. In the context created above, the **Kullback information** of $\mathcal{P}_{\theta_0}$ on $\mathcal{P}_\theta$, or, for brevity's sake, of θ_0 on θ is defined by

$$K(\theta, \theta_0) = \int_\Omega \left(\frac{L^*(\theta_0}{L^*(\theta)}\right) d\mathcal{P}_{\theta_0}. \tag{12}$$

If there exists a dominant measure μ such that $d\mathcal{P}_\theta = f_\theta d\mu$, in the sense that $\mathcal{P}_\theta(A) = \int_A f_\theta d\mu$, the Kullback information may be rendered as

$$K(\theta, \theta_0) = \int_\Omega \left(\frac{L^*(\theta_0)}{L^*(\theta)}\right) f_{\theta_0} d\mu.$$

Remark 1. We note that in the case under consideration, if $A \in \mathcal{A}$, so that $\mathcal{P}(A)$ gives the probability that the dependent variables of the

[4] For the excessively purist reader this may be rationalized as an argument conditioned on a specific sequence in Ξ.

[5] The discussion of this section in part based on Chs. 2 and 3, vol. II, of Dacunha-Castell and Duflo (1986).

problem obey $Y \in A$, then

$$\mathcal{P}_\theta(A) = \int_A L^*(\theta) d\mu, \tag{13}$$

where μ is ordinary Lebesgue measure. Consequently, the Kullback information expression of Eq. (11) may also be written as

$$K(\theta, \theta_0) = \int_\Omega \left(\frac{L^*(\theta_0}{L^*(\theta)} \right) L^*(\theta_0) d\mu = E_0 L^*(\theta_0) - E_0 L^*(\theta) \geq 0. \tag{14}$$

This shows that the Kullback information is a nonnegative function and, further, **that it attains its global minimum when** $\theta = \theta_0$.

3.2 MC Estimators

Definition 2. Consider the probability space (Ω, $\mathcal{A}$, $\mathcal{P}$), and the econometric model (Ω, $\mathcal{A}$, $\mathcal{P}_\theta$), $\theta \in \Theta \subset R^k$, with the "true" parameter, θ_0, being an **interior** point of Θ. A **contrast function** of this model, relative to θ_0, is a function

$K: \Theta \times \Theta \longrightarrow R,$

say $K(\theta, \theta_0)$, having a strict minimum at the point $\theta = \theta_0$, in the sense that $K(\theta_0, \theta_0) < K(\theta, \theta_0)$, for all $\theta \in \Theta$, $\theta \neq \theta_0$.

Definition 3. In the context of Definition 2, let $X = \{X'_{t\cdot} : t = 1, 2, 3, \ldots, T\}$ be a sequence of random vectors (elements), and consider the (nested) sequence of subalgebras [6]

$$\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots \subset \mathcal{G}_T \subset \cdots \mathcal{A}.$$

A **contrast**, relative to θ_0 and K, is a function [7]

$H: \mathcal{N} \times \Theta \times \Omega \longrightarrow R,$

independent of θ_0, such that

i. for every $\theta \in \Theta$, $H_T(\theta, \omega)$ is $\mathcal{G}_T$-measurable;

[6] Basically, the motivation for the sequence of subalgebras is to provide the minimal probability space on which to describe certain sequences of r.v. Thus, for example, if we take $\mathcal{G}_0 = \{\emptyset, \Omega\}$, the trivial σ-algebra used to describe "constants", and $\mathcal{G}_T = \sigma(X_1, X_2, \ldots, X_T)$, we will have produced the sequence referred to in the text, which is quite suitable for studying the samples $\{X_{(T)} : T \geq 1\}$.

[7] In the description of the function, $\mathcal{N}$ represents the integers, i.e. $\mathcal{N} = \{1, 2, \ldots\}$.

ii. $H_T(\theta,\cdot)$ converges to the contrast function $K(\theta,\,\theta_0)$, at least in probability. [8]

A **minimum contrast estimator** (MC) associated with H is a function,

$\hat{\theta}:\ \mathcal{N}\times\Omega \longrightarrow \Theta\,,$

such that

$$H_T(\hat{\theta}_T) \;=\; \inf_{\theta\in\Theta} H_T(\theta).$$

The definition above makes possible the following important

Theorem 1. In the context of Definitions 2 and 3, suppose, further,

i. $\Theta \subset R^k$ is closed and bounded (compact);

ii. $K(\theta,\theta_0)$, and $H_T(\theta,\omega)$ are continuous in θ;

iii. (identification condition) if $K(\theta^1,\theta_0) = K(\theta^2,\theta_0)$ **then** $\theta^1 = \theta^2$;

iii. letting

$$c_T(\delta) \;=\; \sup_{|\theta_1-\theta_2|\le\delta} \mid H_T(\theta_1) - H_T(\theta_2) \mid,$$

there exist sequences $\{\epsilon_k:\ \epsilon_k > 0, k\ge 1\}$, and $\{\delta_k:\ \delta_k > 0, k\ge 1\}$, both (monotonically) tending to zero with k, such that the sets $F_T = \{\omega : c_T(\delta_T) > \epsilon_T\}$ obey $\mathcal{P}(F_T) \le 2\epsilon_T$, and hence $\lim_{n\to\infty}\mathcal{P}(F_T) = 0$.

Then, **every** MC estimator is **consistent.**

Proof: Since $K(\theta,\theta_0)$ is continuous and $K(\theta,\theta_0) = 0$, there exists $\epsilon > 0$, such that

$$K(\theta,\theta_0) > 2\epsilon, \quad \text{for} \quad \theta \in \bar{B}, \tag{15}$$

where

$$B \;=\; \{\theta : \mid \theta - \theta_0 \mid < \epsilon\}. \tag{16}$$

Since B is open , $\Theta^* = \Theta \cap \bar{B}$ is compact; consequently, there exists a **countable** set D, that is everywhere dense in Θ^*, say

$$D \;=\; \{\theta_i :\ i \ge 1\}.$$

---

[8] When a statement like this is made, or when an expectation is taken, we shall always mean that the operations entailed are performed in accordance with the probability measure $\mathcal{P}_{\theta_0}$.

Moreover, for $\epsilon_k < \epsilon$, there exists a finite open cover of Θ^*, say

$$\Theta^* \subset \bigcup_{i=1}^{N} A_i, \quad \text{with } A_i = \{\theta : \mid \theta - \theta_i \mid < \epsilon_k\}. \tag{17}$$

Next, note that we can write

$$H_T(\theta) = H_T(\theta_i) - [H_T(\theta_i) - H_T(\theta)].$$

Consequently, for sufficiently large n, we obtain

$$\begin{aligned} H_T(\theta) &\geq H_T(\theta_i) - \mid H_T(\theta_i) - H_T(\theta) \mid \\ \inf_{\theta \in \Theta^*} H_T(\theta) &\geq \inf_{1 \leq i \leq N} H_T(\theta_i) - \sup_{\theta_i \in D} \sup_{|\theta_i - \theta| < \delta_k} \mid H_T(\theta_i) - H_T(\theta) \mid \\ &\geq \inf_{1 \leq i \leq N} H_T(\theta_i) - c_T(\delta_T), \end{aligned} \tag{18}$$

where $\delta_T < \delta_k$ and hence $c_T(\delta_T) \leq c_T(\delta_k)$. Now if $\hat{\theta}_T$ is the MC estimator, i.e. if $H_T(\hat{\theta}_T) = \inf_{\theta \in \Theta} H_T(\theta)$ we must show that its probability limit is θ_0. It is clear that $\hat{\theta}_T \in \bar{B}$ if and only if $\inf_{\theta \in \Theta^*} H_T(\theta) < H_T(\theta_0)$. This is so since, by the continuity of $H_T(\theta)$, if the condition above holds, there exists a neighborhood of θ_0, say $N(\theta_0; \epsilon) = \{\theta : \mid \theta - \theta_0 \mid < \epsilon\}$, such that

$$\inf_{\theta \in \Theta^*} H_T(\theta) < H_T(\bar{\theta}), \quad \text{for } \bar{\theta} \in N(\theta_0; \epsilon),$$

and it is this type of neighborhood that constitutes the set B. Define now the sets

$$\begin{aligned} B_T &= \{\omega : \hat{\theta}_T \in \Theta^*\}, \quad C_T = \{\omega : \inf_{\theta \in \Theta^*} [H_T(\theta) - H_T(\theta_0)] < 0\} \\ D_T &= \{\omega : \inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] - c_T(\delta_T) < 0\}, \end{aligned} \tag{19}$$

and note that

$$B_T \subset C_T \subset D_T.$$

Define the sets

$$E_T = \{\omega : \inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] < \epsilon_T\}, \quad F_T = \{\omega : c_T(\delta_T) > \epsilon_T\}, \tag{20}$$

and note that for $c_T(\delta_T) \leq \epsilon_T$

$$D_T \cap \bar{F}_T = \{\omega : \inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] < c_T(\delta_T), \text{ and } c_T(\delta_T) \leq \epsilon_T\} \subseteq E_T. \tag{21}$$

Since

$$D_T = (D_T \cap \bar{F}_T) \cup (D_T \cap F_T) \subset (E_T \cup F_T), \tag{22}$$

it follows that

$$\mathcal{P}(B_T) \leq \mathcal{P}(E_T \cup F_T) \leq \mathcal{P}(E_T) + \mathcal{P}(F_T). \tag{23}$$

By iii., of the premises of the proposition, $\mathcal{P}(F_T) \longrightarrow 0$, and, by Definitions 2 and 3

$$\inf_{1\leq i\leq N}[H_T(\theta_i) - H_T(\theta_0)] \xrightarrow{P} \inf_{1\leq i\leq N} K(\theta,\theta_i) - K(\theta_0,\theta_0) \;\geq 2\epsilon,$$

whence we conclude

$$\lim_{n\to\infty} \mathcal{P}_{\theta_0}(E_T) = 0, \quad \text{and hence } \lim_{n\to\infty} \mathcal{P}_{\theta_0}(B_T) = 0.$$

But his means that $\lim_{n\to\infty} \mathcal{P}_{\theta_0}(\bar{B}_T) = 1$, so that $\hat{\theta}_T$ is consistent for θ_0.

q.e.d.

Corollary 1. In the context of Theorem 1, suppose that

$$H_T(\theta) - H_T(\theta_0) \stackrel{\text{a.c.}}{\to} K(\theta,\theta_0)$$

uniformly for $\theta \in \Theta$. Then the MC estimator converges to θ_0 with probability one, i.e, it is **strongly consistent** for θ_0.

Proof: Proceed as in the proof of Theorem 1, and define the sets B, Θ^*, B_T, C_T, as defined therein. If the convergence

$$H_T(\theta) - H_T(\theta_0) \stackrel{\text{a.c.}}{\to} K(\theta,\theta_0),$$

is **uniform in** θ then

$$\inf_{\theta\in\Theta^*}[H_T(\theta) - H_T(\theta_0)] \stackrel{\text{a.c.}}{\to} \inf_{\theta\in\Theta^*} K(\theta,\theta_0) \geq 2\epsilon > 0. \tag{24}$$

Consequently,

$$\overline{\lim_{T\to\infty}} C_T \;=\; C^*, \quad \text{obeys} \quad \mathcal{P}(C^*) = 0. \tag{25}$$

Since, by construction, $B_T \subset C_T$, we have that

$$B^* = \overline{\lim_{n\to\infty}} B_T \subseteq \overline{\lim_{n\to\infty}} C_T = C^*; \tag{26}$$

hence, in view of Eq. (25) we conclude that $\mathcal{P}(B^*) = 0$. But this means that the ML estimator,

$$\inf_{\theta\in\Theta} H_T^*(\theta) \;=\; H_T^*(\hat{\theta}_T),$$

obeys $\hat{\theta}_T \in B$ with probability one, or that it converges a.c. to the true parameter θ_0.

q.e.d.

4 Application to the GLSEM

Let $L_T(\theta) = (1/T)L^*(\theta)$, to be referred to subsequently as LF, and note that $H_T(\theta) = -L_T(\theta)$ is a **constrast** in the sense of Definition 3. In fact if we define

$$H_T^* = H_T(\theta) - H(\theta_0), \tag{27}$$

we see that that the ML estimator of θ is a **MC estimator**. Using the results in Ch. 4 Dhrymes (1984), we note that

$$|H_T(\theta)| \leq k_1 + k_2 \parallel \tilde{M}_{zz} \parallel = g(Y, X)$$

which is **an integrable function and does not depend on** θ; moreover, H_T, and hence H_T^* satisfy the conditions of Theorem 1 above. In Ch. 2 of Dhrymes (1993) it is shown that $\tilde{M}_{zz}$ converges a.c. (almost certainly), and we colcude that

$$\tilde{M}_{zz} \stackrel{\text{a.c.}}{\to} \begin{bmatrix} \Omega_0 & 0 \\ 0 & 0 \end{bmatrix} + (\Pi_0,\ I)' M_{xx}(\Pi_0,\ I) \ = \ M_{zz},$$

$$\tilde{M}_{xx} \ = \ \frac{1}{T}X'X, \quad \tilde{M}_{xx} \stackrel{\text{a.c.}}{\to} M_{xx}. \tag{28}$$

By the compactness of the parameter space Θ, the stability of the model, and the assumptions regarding the exogenous variables of the model $H_T^*(\theta) \stackrel{\text{a.c.}}{\to} K(\theta, \theta_0)$, uniformly in Θ. Hence, by Corollary 1, we conclude that the ML estimator converges a.c. to the true parameter vector, provided the identification condition in iii. of Theorem 1 is satisfied.

Our next task is to obtain an expression of the contrast function K and to verify that it is the Kullback information. This last requirement is obvious, since $H_T^*(\theta, \theta_0)$ converges by the Kolmogorov criterion, see Proposition 22, Ch. 3 Dhrymes (1989), to the limit of $E_0[H_T^*(\theta, \theta_0)$; hence, the limit is nonnegative and assumes its global minimum at $\theta = \theta_0$, as required of contrast functions. As for the form of the Kullback information, we find

$$\begin{aligned} A^{*\prime}\tilde{M}_{zz}A^* \stackrel{\text{a.c}}{\to} \ & \Sigma_0 + (A^* - A_0^*)' \begin{bmatrix} \Omega_0 & 0 \\ 0 & 0 \end{bmatrix} (A^* - A_0^*) \\ & + (A^* - A_0^*)'(\Pi_0, I)' M_{xx}(\Pi_0, I)(A^* - A_0^*) \\ & + \Sigma_0(B_0^{*-1}B^* - I) + (B^{*\prime}B_0^{*\prime -1} - I)\Sigma_0, \end{aligned}$$

and

$$A_0^{*\prime}\tilde{M}_{zz}A_0^* \stackrel{\text{a.c}}{\to} \Sigma_0. \tag{29}$$

Consequently,

$$\inf_{\theta\in\Theta} H_T^*(\theta) = L_T(\theta_0) - \sup_{\theta\in\Theta} L_T(\theta) \stackrel{\text{a.c.}}{\to} \inf_{\theta\in\Theta} K(\theta,\, \theta_0). \tag{30}$$

But,

$$\begin{aligned} K(\theta,\, \theta_0) &= -\frac{1}{2}m - \frac{1}{2}\ln \mid \Sigma_0 \mid + \frac{1}{2}\ln \mid B_0^* B_0^{*'} \mid + \frac{1}{2}\ln \mid \Sigma \mid - \frac{1}{2}\ln \mid B^* B^{*'} \mid \\ &\quad + \frac{1}{2}\mathrm{tr}\Sigma^{-1}(A^* - A_0^*)' P_0^*(A^* - A_0^*) \\ &\quad + \frac{1}{2}\mathrm{tr}\Sigma^{-1}(B^{*'} B_0^{*'-1})\Sigma_0(B_0^{*-1}B^*), \\ P_0^* &= (\Pi_0,\; I)' M_{xx}(\Pi_0,\; I). \end{aligned} \tag{31}$$

Noting that $\Omega_0 = B_0^{*'-1}\Sigma_0 B_0^{*-1}$ and, therefore, that $B_0^{*'}\Omega_0 B_0^* = \Sigma_0$, we can rewrite the Kullback Information of Eq. (31) as

$$\begin{aligned} K(\theta_0,\, \theta) &= -\frac{1}{2}m - \frac{1}{2}\ln \mid \Sigma^{-1} \mid - \frac{1}{2}\ln \mid \Omega_0 \mid - \frac{1}{2}\ln \mid B^* B^{*'} \mid + \frac{1}{2}\mathrm{tr}\Sigma^{-1} J^* \\ J^* &= B^{*'}\Omega_0 B^* + (A^* - A_0^*)' P_0^*(A^* - A_0^*) \end{aligned} \tag{32}$$

The expression above may be (partially) minimized with respect to Σ^{-1}, yielding the first order conditions,

$$\frac{\partial K}{\partial \mathrm{vec}(\Sigma^{-1})} = -\frac{1}{2}\mathrm{vec}(\Sigma)' + \frac{1}{2}\mathrm{vec}(J^*)' = 0,$$

whence we obtain

$$\Sigma = J^*.$$

Noting that

$$\frac{1}{2}\ln \mid \Sigma \mid + \frac{1}{2}\ln \mid (B^* B^{*'})^{-1} \mid = \frac{1}{2}\ln \mid B^{*'-1}\Sigma B^{*-1} \mid,$$

and inserting the minimizer in Eq. (32), we obtain the "concentrated" Kullback information expression,

$$K^*(\theta, \theta_0) = \frac{1}{2}\ln\left(\frac{\mid \Omega_0 + B^{*'-1}(A^* - A_0^*)' P_0^*(A^* - A_0^*) B^{*-1} \mid}{\mid \Omega_0 \mid}\right). \tag{33}$$

Remark 2. Since the expression in the large round bracket is equal to or greater than unity, it is **globally** minimized when we take $A^* = A_0^*$;

when we do so the fraction becomes unity, in which case the Kullback information becomes null. Referring back to the partial minimization with respect to Σ, we see that when the choice $A^* = A_0^*$ is made, the expression therein implies $\Sigma = \Sigma_0$. However, in Eq. (33) **it is not transparent that the global minimizer is unique**. This is so since the matrix P_0^* is of dimension $G+m$, but of rank G! Hence, its null space is **of dimension** m and thus contains m linearly independent vectors, say the columns of some matrix N_0. If J is an arbitrary $m \times m$ **nonsingular matrix** consider the choice $A^* = A_0^* + N_0 J$, which implies $P_0^*(A^* - A_0^*) = P_0^* N_0 J = 0$. Consequently, the Kullback information of Eq. (33) does not satisfy the condition in item iii. of Theorem 1, unless further restrictions are placed on the structure, as indicated in Conventions 1 and 2. Suppose that in order to make A^* **admissible** the **restrictions required** were such that the **intersection of the null space of** P_0^* and **the class of admissible structures has** A_0^* **as its only member**. Evidently, this would establish identification!

5 Alternative Derivation of the Identification Conditions for the GLSEM

The preceding discussion affords us a singularly felicitous venue for establishing the identification conditions for the GLSEM, in a Kullback information context. In Remark 2 we have established that in order to have identification, any matrix A^* for which the (concentrated) Kullback information attains its minimum, must have the property that $A^* = A_0^*$, where A_0^* is the "true" parameter matrix. This means that a necessary and sufficient condition for identification is that $\Psi = (A^* - A_0^*)' P_0^* (A^* - A_0^*) = 0$, for every admissible matrix A^*. Noting that, **subject to normalization**, $A^* - A_0^* = A_0 - A$, where now $A = (B', C')'$, $B^* = I - B$, and A_0 is the true parameter matrix, we may rewrite Ψ in terms of A and A_0; moreover, since we are dealing with a positive semidefinite matrix, the condition $\Psi = 0$ is equivalent to

$$\mathrm{tr}(\Psi) = \sum_{i=1}^{m} (a_{\cdot i}^0 - a_{\cdot i})' P_0^* (a_{\cdot i}^0 - a_{\cdot i}).$$

Reintroducing the selection matrices L_i, and $L = \mathrm{diag}(L_1, L_2, \ldots, L_m)$, of the preceding sections we note that

$$a_{\cdot i}^0 - a_{\cdot i} = L_i(\delta_{\cdot i}^0 - \delta_{\cdot i}), \quad \mathrm{tr}\Psi = (\delta^0 - \delta) L' (I_m \otimes P_0^*) L (\delta^0 - \delta).$$

In this framework a necessary and sufficient condition for identification of the parameters of the system is that $L'(I_m \otimes P_0^*) L$ be a **positive**

definite matrix. The i^{th} diagonal block of that matrix, however, is of the form

$$L_i'(\Pi, I)' M_{xx}(\Pi, I) L_i = S_i' M_{xx} S_i,$$

which is thus required to be nonsingular, i.e. it is required that

$$\text{rank}(S_i) = \text{rank}(\Pi L_i, L_{2i}) = m_i + G_i, \quad \text{for every i} = 1, 2, \ldots \text{m}. \quad (34)$$

From Dhrymes (1973), or Proposition 2 Ch. 6 Dhrymes (1978), we note that these were **precisely the conditions for the identification of the** i^{th} structural equation.

Thus, we have derived the necessary and sufficient conditions for the identification of the equations of a GLSEM, solely in terms of the identification requirements placed on Kullback Information!

REFERENCES

Dacunha-Castelle, D. and M. Duflo (1986), *Probability and Statistics*, vols. I and II, New York: Springer-Verlag.

Dhrymes, P.J. (1973), "Restricted and Unrestricted Reduced Forms: Asymptotic Distributions and Relative Efficiencies", *Econometrica*, vol. 41, pp.119-134.

——————— (1978), *Introductory Econometrics*, New York: Springer-Verlag.

——————— (1984), *Mathematics for Econometrics*, 2nd edition, New York: Springer-Verlag.

——————— (1989), *Topics in Advanced Econometrics: Probability Foundations*, New York: Springer Verlag.

——————— (1993), *Topics in Advanced Econometrics: Linear and Nonlinear Simultaneous Equations*, New York: Springer-Velag (in press).