Identification and Kullback Information in the GLSEM

by

Phoebus Dhrymes

Discussion Paper No. 702

Identification and Kullback Information in the GLSEM*

Phoebus J. Dhrymes

Columbia University

August 20, 1994

Abstract

In this paper we employ the **Kullback Information** apparatus in (a) obtaining the strong consistency of the maximum likelihood (ML) estimator in the standard version of the general linear structural econometric model (GLSEM); (b) deriving very succinctly the necessary and sufficient (nas) conditions for identification by the use of exclusion restrictions. The arguments given in (a), however, are equally applicable to a wide class of nonlinear models and the arguments in (b) are equally applicable in the context of more general types of restrictions.

1 Introduction

The purpose of this paper is to introduce 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).

* This paper is prepared for the Conference in honor of Carl F. Christ.

2 Formulation of the Problem and Notation

Consider the standard GLSEM

$$YB^{*}=XC+U, \text{ or } ZA^{*}=U, \text{ and its reduced form } Y=X\Pi+V,$$

$$A^{*}=(B^{*\prime}, -C^{\prime})^{\prime}, \quad Z=(Y,X), \quad \Pi=CD, \quad V=UD, \quad D=B^{*-1} \tag{1}$$

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]

$$E(u_{t\cdot}^{\prime})=0, \quad \mathrm{Cov}(u_{t\cdot}^{\prime})=\Sigma>0.$$

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 the structural form in 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

[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) Chapter 4 pp. 257ff, and Chapter 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 $(1/\sqrt{T})\sum_{t=1}^{T}(I\otimes x_{t\cdot})^{\prime}u_{t\cdot}^{\prime}$ should obey a martingale difference CLT with a Lindeberg condition (Chapter 5 pp. 323ff).

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 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

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

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

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

Finally, we append the following standard assumptions:

[3] The device of selection matrices was first introduced, in this context, by Dhrymes (1973). Greater detail regarding their meaning and function may be found in that reference, as well as in Dhrymes (1978).

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 $\Xi \subset R^s$.

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, B^* is a nonsingular matrix and Σ is positive definite.

We may thus write the likelihood function of the observations as

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

$$S = \frac{1}{T}A^{*\prime}\tilde{M}_{zz}A^*, \quad \tilde{M}_{zz} = \frac{1}{T}Z'Z, \quad \theta = (\mathrm{vec}(A^*)', \mathrm{vec}(\Sigma)')' \quad (10)$$

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 which comprises the parameter triplet (B, C, Σ), 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

$$K(\theta, \theta_0) = \int_\Omega \left(\frac{\mathcal{P}_{\theta_0}}{\mathcal{P}_\theta}\right) d\mathcal{P}_{\theta_0}, \quad (11)$$

[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 is, in part, based on Chs. 2 and 3, vol. II, of Dacunha-Castell and Duflo (1986).

where θ_0 is the "true" parameter vector, it being understood that if $\mathcal{P}_\theta(\omega) = 0$, then $\mathcal{P}_{\theta_0}(\omega) = 0$ and $0/0$ is defined to be zero.

In the context of this discussion, it is to be understood that the dependent variables of the problem are viewed as measurable functions defined on the sample space, i.e.

$$y : \Omega \longrightarrow R^m ,$$

so that everything may be expressed in terms of the econometric models $(\Omega, \mathcal{A}, \mathcal{P}_{\theta_0})$ and $(\Omega, \mathcal{A}, \mathcal{P}_\theta)$. If there exists a dominant measure μ such that $d\mathcal{P}_\theta = f_\theta d\mu$, in the sense that $\mathcal{P}_\theta(A) = \int_A f_\theta d\mu$, for every $\mathcal{A}$-measurable set A, by a simple change in variable procedure, the Kullback information (KI) may be rendered as

$$K(\theta, \theta_0) = \int_\Omega \left(\frac{f(\theta_0)}{f(\theta)} \right) f(\theta_0) d\mu. \tag{12}$$

Remark 1. To connect the notion of a dominant measure and KI to the case under consideration, recall that

$$y : \Omega \longrightarrow R^m ,$$

introduce the measure space $(R^m, \mathcal{B}(R^m))$, and the σ-algebra, $\sigma(y) = \mathcal{G} \subset \mathcal{A}$ induced by y. Clearly, for every $B \in \mathcal{B}(R^m)$, $A = y^{-1}(B) \in \mathcal{G}$. Conversely, for every $A \in \mathcal{G}$, $y(A) = B \in \mathcal{B}(R^m)$. Let L^* be the likelihood (**not** the loglikelihood) function of y and note that for any $A \in \mathcal{G}$, $\mathcal{P}_\theta(A)$ gives the probability that the dependent variables of the problem obey $y \in B$, where $A = y^{-1}(B)$; thus,

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

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

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

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 $(\Omega, \mathcal{A}, \mathcal{P})$, and the econometric model $(\Omega, \mathcal{A}, \mathcal{P}_\theta)$, $\theta \in \Theta \subset R^k$, 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]

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

such that

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

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,

[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. $\mathcal{N} = \{1, 2, \ldots\}$.

[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 $\mathcal{P}_{\theta_0}$.

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. letting

$$c_n(\delta) = \sup_{|\theta_1 - \theta_2| \le \delta} | H_n(\theta_1) - H_n(\theta_2) |,$$

there exist sequences $\{\epsilon_n : \epsilon_n > 0, n \ge 1\}$, and $\{\delta_n : \delta_n > 0, n \ge 1\}$, both (monotonically) tending to zero with n, such that the sets $F_n = \{\omega : c_n(\delta_n) > \epsilon_n\}$ obey $\mathcal{P}(F_n) \le 2\epsilon_n$, and hence $\lim_{n\to\infty} \mathcal{P}(F_n) = 0$;

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

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

Proof: We proceed by contradiction; thus suppose the estimator does not converge to θ_0. Since $K(\theta, \theta_0)$ is continuous and $K(\theta_0, \theta_0) = 0$, there exists $\epsilon > 0$, such that

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

where

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

We shall obtain a contradiction if $\hat{\theta}_T$ converges in Θ, **but** outside the set B. 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

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

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

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

Next, note that we can write $H_T(\theta) = H_T(\theta_i) - [H_T(\theta_i) - H_T(\theta)]$, so that $H_T(\theta) \ge H_T(\theta_i) - | H_T(\theta_i) - H_T(\theta) |$. Consequently, for sufficiently large T, we obtain

$$\begin{aligned} \inf_{\theta \in \Theta^*} H_T(\theta) &\ge \inf_{1 \le i \le N} H_T(\theta_i) - \sup_{\theta_i \in D} \sup_{|\theta_i - \theta| < \delta_T} | H_T(\theta_i) - H_T(\theta) | \\ &\ge \inf_{1 \le i \le N} H_T(\theta_i) - c_T(\delta_T). \end{aligned} \tag{18}$$

Let $\hat{\theta}_T$ be the MC estimator, i.e. $H_T(\hat{\theta}_T) = \inf_{\theta \in \Theta} H_T(\theta)$; we 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

$$B_T = \{\omega : \hat{\theta}_T \in \Theta^*\}, \quad C_T = \{\omega : \inf_{\theta\in\Theta^*}[H_T(\theta) - H_T(\theta_0)] < 0\}$$

$$D_T = \{\omega : \inf_{1\leq i\leq N}[H_T(\theta_i) - H_T(\theta_0)] - c_T(\delta_T) < 0\}, \tag{19}$$

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, $\mathcal{P}(F_T) \longrightarrow 0$; hence by Definition 2, and Corollary 4 Dhrymes (1989) p. 147,

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

which is a **contradiction**; whence we conclude

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

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

q.e.d.

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

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

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. Suppose we have convergence as in the premise, but $\hat{\theta}_T$ does not converge to θ_0. We show a contradiction. If the convergence[9]

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

is **uniform in** θ then

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

which is a contradiction.

Consequently,

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

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

$$B^* = \limsup_{T \to \infty} B_T \subseteq \limsup_{T \to \infty} C_T = C^*; \tag{26}$$

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.

Remark 2. Notice that in the proofs above we do not require that the likelihood function be quadratic in the parameters of interest, nor do we require that the observations be i.i.d.; thus the results are applicable to a wide variety of contexts that can be shown to satisfy conditions i. through iii. of Theorem 1 and, for strong consistency, the premise of Corollary 1.

[9] Note that $K(\theta_0, \theta_0) = 0$.

4 Identification and Strong Consistency of ML in the GLSEM

4.1 Strong Consistency

In this section we employ the Kullback information (KI) developed in the preceding sections to establish identification criteria, as well as the strong consistency properties of the ML estimator of parameters in the standard GLSEM. First, we show that the LF of the observations on the model of Eq. (1) satisfies the conditions in Theorem 1. Evidently, condition i of the theorem is satisfied, in view of assumption A.4. Define

$$L(\theta) = \frac{1}{T} \ln L^*(Y, X; \theta), \quad H_T^*(\theta) = -L(\theta). \tag{27}$$

and note that $H_T^*(\theta) = -L_T(\theta)$ is a **constrast** in the sense of Definition 3. In fact, we shall not violate the sense of Definition 3, if we put

$$H_T(\theta) = H_T^*(\theta) - H_T^*(\theta_0), \tag{28}$$

since the minimization procedure does not involve $H_T^*(\theta_0)$. This is quite evident from the fact that

$$\inf_{\theta \in \Theta} H_T(\theta) = \inf_{\theta \in \Theta} H_T^*(\theta).$$

By assumptions A.3 and A.4 we may conclude, using the results in Chapter 4 Dhrymes (1984), that

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

which is[10] **an integrable function and does not depend on** θ; thus, H_T satisfies condition iii of Theorem 1 as well. As for condition iv (the identification condition), this is of course a condition that must be imposed in the ML context as well, otherwise no estimation is possible. The point of this section is to illustrate how condition iv yields the standard results of the identification discussions in the GLSEM, the argumentation for which normally consumes several pages.

Our next task is to determine the limit to which H_T converges. We have

Proposition 1. Under conditions A.1 through A.4, and assuming the GLSEM is **static**

$$H_T \stackrel{\text{a.c.}}{\to} K(\theta, \theta_0), \quad \text{where} \tag{29}$$

[10] We note that $\tilde{M}_{zz} = (Z'Z/T)$.

$$
\begin{aligned}
K(\theta,\theta_0) &= \frac{1}{2}\ln|\Sigma| - \frac{1}{2}\ln|B^{*'}B^*| + \frac{1}{2}\mathrm{tr}\Sigma^{-1}Q \\
&\quad -\frac{m}{2} - \frac{1}{2}\ln|\Sigma_0| + \frac{1}{2}\ln|B_0^{*'}B_0^*|, \quad \text{where} \\
Q &= B^{*'}\Omega_0 B^* + (A_0^* - A^*)'P_0(A_0^* - A^*) \\
P_0 &= (\Pi_0, I_G)'M_{xx}(\Pi_0, I_G), \qquad \frac{X'X}{T} \overset{\text{OL } \underline{\text{or}} \text{ a.c.}}{\to} M_{xx}.
\end{aligned}
$$

Proof: From the nature of the LF, we need only determine the limit of

$$S(\theta) = \frac{1}{T}A^{*'}Z'ZA^*, \quad A^* = (B^{*'}, -C')'. \tag{30}$$

Since $ZA^* = ZA_0^* - Z(A_0^* - A^*) = U - Z(A_0^* - A^*)$, we need only determine the limiting behavior of

$$\frac{U'U}{T}, \; \frac{U'Z}{T}, \; \text{and} \; \frac{Z'Z}{T}.$$

We have

$$\frac{U'U}{T} = \frac{1}{T}\sum_{t=1}^{T} u_{t\cdot}'u_{t\cdot} \overset{\text{a.c.}}{\to} \Sigma_0. \tag{31}$$

This is so since $\{u_{t\cdot} : t \geq 1\}$ is a sequence of i.i.d. random elements (vectors) with mean zero and covariance matrix Σ_0. The almost certain (a.c.) convergence follows by Proposition 23 Dhrymes (1989) p. 188.[11]

[11] If the model contains lagged dependent variables, the error process in the **final form** is representable as

$$u_{t\cdot}^{*'} = \sum_{j=0}^{\infty} F_j u_{t\cdot}', \quad \sum_{j=0}^{\infty} \| F_j \| < \infty,$$

which shows it to be a strictly stationary process, with covariance matrix

$$\Omega^* = \sum_{j=0}^{\infty} F_j \Sigma_0 F_j', \quad \text{such that} \quad \| \Omega^* \| < \infty.$$

Moreover, it may be shown that the sequence $\{u_{t\cdot}^{*'}u_{t\cdot}^* : t \geq 1\}$ is at least **covariance stationary** and its expectation, Ω^*, is finite. The convergence of $\left(\sum_{t=1}^{T} u_{t\cdot}^{*'}u_{t\cdot}^*/T\right)$, to Ω^* is then a consequence of Proposition 33 Dhrymes (1989) p. 362. Thus, even though for the sake of simplicity we deal with a static model, all arguments given in this paper are easily adapted, *mutatis mutandis*, to dynamic models and more complex forms of error specification. It is just a question of how opaque one wants to make one's presentation.

Next, we consider the limiting behavior of $U'Z/T$, which consists of two components, $U'X/T$, and $U'Y/T$. The first component obeys

$$\frac{U'X}{T} = \frac{1}{T}\sum_{j=1}^{T} u_{t\cdot}' x_{t\cdot} \stackrel{\text{a.c.}}{\to} 0, \tag{32}$$

by the Kolmogorov criterion, see Proposition 22 Dhrymes (1989) p. 186. The second component, obtained from the reduced form representation, obeys

$$\frac{U'Y}{T} = \left(\frac{1}{T}U'X\right)\Pi_0 + D_0'\left(\frac{U'U}{T}\right)D_0 \stackrel{\text{a.c.}}{\to} D_0'\Sigma_0 D_0 = \Omega_0, \quad D_0 = B_0^*, \tag{33}$$

by the preceding discussion. Finally, since

$$\tilde{M}_{zz} = \frac{1}{T}Z'Z = \frac{1}{T}\begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \stackrel{\text{a.c.}}{\to} \begin{bmatrix} \Omega_0 & 0 \\ 0 & 0 \end{bmatrix} + P_0, \tag{34}$$

we may establish, after some manipulation, that for every $\theta \in \Theta$

$$S(\theta) \stackrel{\text{a.c.}}{\to} B^{*\prime}\Omega_0 B^* + (A_0^* - A^*)' P_0 (A_0^* - A^*) = Q. \tag{35}$$

Hence, we conclude that, **uniformly in** Θ, we have that

$$L(\theta) \stackrel{\text{a.c.}}{\to} \bar{L}(\theta, \theta_0), \tag{36}$$

$$= -\frac{m}{2} - \ln|\Sigma| + \frac{1}{2}\ln|B^{*\prime}B^*| + \mathrm{tr}\Sigma^{-1}Q, \quad \text{where}$$

$$Q = B^{*\prime}\Omega_0 B^* + (A_0^* - A^*)' P_0 (A_0^* - A^*).$$

It follows, therefore, that **uniformly in** Θ

$$H_T(\theta) \stackrel{\text{a.c.}}{\to} \bar{L}(\theta_0, \theta_0) - \bar{L}(\theta, \theta_0) = -\frac{1}{2}\left(1 + \ln|\Sigma_0| - \ln|B_0^{*\prime}B_0^*|\right)$$

$$+\frac{1}{2}\left(\ln|\Sigma| - \ln|B^{*\prime}B^*|\right) + \mathrm{tr}\Sigma^{-1}Q. \tag{37}$$

Defining

$$K(\theta, \theta_0) = \bar{L}(\theta_0, \theta_0) - \bar{L}(\theta, \theta_0) \tag{38}$$

we shall now show that K is the asymptotic KI of the problem, i.e. the limit of $E_{\theta_0}L(\theta_0) - E_{\theta_0}L(\theta)$. To this effect we note that

$$E_{\theta_0}L(\theta_0) = -\frac{m}{2}\ln(2\pi) - \frac{m}{2} - \frac{1}{2}\ln|\Sigma_0| + \frac{1}{2}\ln|B_0^{*\prime}B_0^*|, \tag{39}$$

$$E_{\theta_0}L(\theta) = -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| + \frac{1}{2}\ln|B^{*\prime}B^*| - \frac{1}{2}\mathrm{tr}\Sigma^{-1}Q,$$

Consequently, defining the sample based KI by $K_T(\theta,\theta_0) = L(\theta_0)-L(\theta)$, we find

$$K_T(\theta,\theta_0) = -\frac{1}{2}\left(\ln|\Sigma_0| - \ln|B_0^{*\prime}B^*|\right) - \frac{m}{2} + \frac{1}{2}\left(\ln|\Sigma| - \ln|B^{*\prime}B^*|\right)$$

$$+\frac{1}{2}\mathrm{tr}\Sigma^{-1}\tilde{Q}, \quad P_0 = (\Pi_0, I_G)'\tilde{M}_{xx}(\Pi_0, I_G), \tag{40}$$

$$\tilde{Q} = B^{*\prime}\Omega_0 B^* + (A_0^* - A^*)'\tilde{P}_0(A_0^* - A^*) \text{ and } \tilde{M}_{xx} = \frac{X'X}{T},$$

and it is easily verified that the asymptotic KI is given by

$$\lim_{T\to\infty} K_T(\theta,\theta_0) = K(\theta,\theta_0) = \bar{L}(\theta_0,\theta_0) - \bar{L}(\theta,\theta_0). \tag{41}$$

q.e.d.

An immediate consequence of the preceding is

Corollary 2. Under assumptions A.1 through A.4 and Conventions 1 and 2, and assuming the GLSEM is **identified**, the ML estimator $\hat{\theta}_T$ defined by the operation $\inf_{\theta\in\Theta} H_T(\theta)$ obeys

$$\inf_{\theta\in\Theta} H_T(\theta) \stackrel{\text{a.c.}}{\to} \inf_{\theta\in\Theta} K(\theta,\theta_0) = K(\bar{\theta},\theta_0) = 0, \text{ and } \hat{\theta}_T \stackrel{\text{a.c.}}{\to} \theta_0.$$

Proof: The function H_T satisfies all conditions of Theorem 1, as well the conditions of Corollary 1; consequently,

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

If $\bar{\theta}$ is the point at which K attains its global minimum then: (a) $\hat{\theta}_T \stackrel{\text{a.c.}}{\to} \bar{\theta}$, and (b) $K(\bar{\theta},\theta_0) = 0$. But, from the properties of KI we also have $K(\theta_0,\theta_0) = 0$. By the identification condition, we have $\bar{\theta} = \theta_0$. Thus $\hat{\theta}_T \stackrel{\text{a.c.}}{\to} \theta_0$, and the ML estimator of the parameters of the standard GLSEM is strongly consistent.

q.e.d.

4.2 Identification

In this section we derive the detailed identification conditions for (each of) the equations of the GLSEM, as implications of the identification requirement of the preceding discussion. We recall that

$$K(\theta,\, \theta_0) \;=\; -\frac{1}{2}m - \frac{1}{2}\ln\mid \Sigma_0 \mid + \frac{1}{2}\ln\mid B_0^* B_0^{*'} \mid + \frac{1}{2}\ln\mid \Sigma \mid - \frac{1}{2}\ln\mid B^* B^{*'} \mid$$

$$+\frac{1}{2}\mathrm{tr}\Sigma^{-1}Q.$$

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 (asymptotic) KI of Eq. (41) as

$$K(\theta_0, \theta) = -\frac{1}{2}m - \frac{1}{2}\ln\mid \Sigma^{-1} \mid - \frac{1}{2}\ln\mid \Omega_0 \mid - \frac{1}{2}\ln\mid B^* B^{*'} \mid + \frac{1}{2}\mathrm{tr}\Sigma^{-1}Q. \tag{42}$$

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

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

whence we obtain

$$\Sigma \;=\; Q.$$

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. (42), we obtain the "concentrated" KI expression,

$$K^*(\theta, \theta_0) \;=\; \frac{1}{2}\ln\left(\frac{\mid \Omega_0 \;+\; B^{*'-1}(A_0^* - A^*)' P_0 (A_0^* - A^*) B^{*-1} \mid}{\mid \Omega_0 \mid}\right). \tag{43}$$

Remark 3. 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. (43) **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 $P_0(A_0^* - A^*) = P_0^* N_0 J = 0$. Consequently, the Kullback information of

Eq. (43) does not satisfy the (identification) condition in item iv of Theorem 1, unless certain restrictions are placed on the structure, as indicated in Conventions 1 and 2. Suppose that in order to make A^* **admissible,**[12] 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!

In Remark 3, we established that in order to have identification, any matrix A^* for which the (concentrated) Kullback information attains its global 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

$$\Psi = (A_0^* - A^*)' P_0^* (A_0^* - A^*) = 0,$$

for every admissible matrix A^*. To implement this requirement we have at our disposal Conventions 1 and 2. By Convention 1 (normalization) we may set $B^* = I_m - B$, with $b_{ii} = 0$, for all i, and similarly for $B_0^* = I_m - B_0$. Consequently, $A_0^* - A^* = A - A_0$, where now $A = (B', \ C')'$, and A_0 is the true parameter matrix; thus, 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

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

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

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

In this framework a necessary and sufficient condition for identification of the parameters of the system is that $L'(I_m \otimes P_0)L$, which is a **block diagonal** matrix be **positive definite.** The i^{th} diagonal block of that matrix, however, is of the form

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

Thus, identification of the system is obtained if and only if

$$\mathrm{rank}(S_i) = \mathrm{rank}(\Pi L_i, L_{2i}) = m_i + G_i, \quad \text{for every i} = 1, 2, \ldots \text{m}. \qquad (44)$$

[12] In this context a matrix A^* is said to be admissible, as in the standard context, if and only if it satisfies all prior rerstrictions.

By Theorem 5 and Corollary 1, in Chapter 3 Dhrymes (1994), the conditions above are the necessary and sufficient conditions for the identification of the parameters in the i^{th} equation, and the system as a whole. Consequently, we have derived the necessary and sufficient conditions for the identification of the equations of a GLSEM by a very simple argument, based solely on the identification requirements placed on KI, and almost as a by-product of the argument showing strong consistently for the ML estimator.

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.

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

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

——————— (1994), *Topics in Advanced Econometrics: Linear and Nonlinear Simultaneous Equations*, New York: Springer-Verlag.