

**Identification and Kullback
Information in the GLSEM**

by

Phoebus J. Dhrymes, Columbia University

March 1996

Discussion Paper Series No. 9596-11

Identification and Kullback Information in the GLSEM*

Phoebus J. Dhrymes

Columbia University

August 20, 1994; revised March 30, 1996

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 illustrate the usefulness of **Kullback Information** in deriving identification conditions. We do so in the context of the standard GLSEM, by showing how the resolution of the identification problem becomes almost a routine by product of the convergence properties of the (log) likelihood function (LF), and the consistency property of the maximum likelihood (ML) estimator.

* This paper is prepared for the Conference in honor of Carl F. Crist, and is inspired by his exceptionally lucid discussion of the identification problem in Christ (1966).

2 Formulation of the Problem and Notation

Consider the standard GLSEM and the corresponding reduced form (RF)

$$YB^* = XC + U, \text{ or } ZA^* = U, \text{ with RF } Y = X\Pi + V, \quad (1)$$

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

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., with¹

$$E(u'_t) = 0, \quad \text{Cov}(u'_t) = \Sigma > 0.$$

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

$$Y = YB + XC + U = ZA + U, \quad \text{where}$$

$$A = \begin{pmatrix} B \\ C \end{pmatrix}, \quad b_{ii} = 0, \quad i = 1, 2, \dots, m. \quad (2)$$

The model is assumed to be **dynamic**; thus, the predetermined variables are given by

$$x_t = (y_{t-1}, y_{t-2}, \dots, y_{t-k}, p_t),$$

where p_t is an s -element **row** vector representing the **exogenous** variables; the latter may be assumed (a): to be generated by a **nonstochastic** process, or (b): by a zero mean stochastic process, independent of the error process, which is square integrable and obeys, at least, a second moment ergodic condition

$$\frac{1}{T} \sum_{t=1}^T p'_t p_t \xrightarrow{\text{a.s.}} \lim_{T \rightarrow \infty} \sum_{t=1}^T E p'_t p_t = M_{pp} > 0, \quad (3)$$

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

the term a.c. meaning almost certainly, also rendered as almost surely (a.s.) in the literature.

In this context, “identification” is obtained by “exclusion restrictions”, although, of course, more general schemes are possible; this alternative 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 pre-determined, respectively.**

In order to implement this convention, we introduce the device of selection matrices,² 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

$$YL_{1i} = Y_i, \quad XL_{2i} = X_i, \quad i = 1, 2, \dots, m. \quad (4)$$

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

$$y_i = Y_i\beta_i + X_i\gamma_i + Y_i^*\beta_i^* + X_i^*\gamma_i^* + u_i, \quad i = 1, 2, \dots, m, \quad (5)$$

where the notation y_i , u_i means the i^{th} column of Y and U , respectively, and β_i , γ_i contain, respectively, the elements in the i^{th} column of B , b_i , and C , c_i , not known a priori to be zero. Evidently, β_i^* and γ_i^* represent the elements of the two columns, respectively, set to zero by the prior restrictions. It follows immediately that

$$b_i = L_{1i}\beta_i, \quad c_i = L_{2i}\gamma_i, \quad L'_{1i}b_i = \beta_i, \quad L'_{2i}c_i = \gamma_i. \quad (6)$$

Define

$$L_i = \begin{bmatrix} L_{1i} & 0 \\ 0 & L_{2i} \end{bmatrix}, \quad L_i^* = \begin{bmatrix} L_{1i}^* & 0 \\ 0 & L_{2i}^* \end{bmatrix}, \quad i = 1, 2, \dots, m, \quad (7)$$

and note that the i^{th} column of A , in Eq. (2), is given by

$$a_i = \begin{pmatrix} b_i \\ c_i \end{pmatrix}, \quad i = 1, 2, \dots, m.$$

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

$$\delta_i = L_i' a_i, \quad i = 1, 2, \dots, m, \quad (8)$$

² 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).

and for the system as a whole we have³

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

Finally, we append the following standard assumptions:

- A1. The error process $\{u'_t : t \geq 1\}$ 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 $\Xi \subset R^s$, or alternatively they are generated by a zero mean stochastic process independent of the structural errors, are square integrable and obey the condition in Eq. (3).
- A4. The parameter space, $\Theta \subset R^n$ 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

$$L^*(\theta) = (2\pi)^{-(mT/2)} |\Sigma|^{-(T/2)} |B^{*'} B^*|^{(T/2)} \exp\left(-\frac{T}{2}\right) \text{tr} \Sigma^{-1} S, \quad \text{where}$$

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

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

3 Kullback Information and Minimum Contrast (MC) Estimators

3.1 Kullback Information

We begin by defining the general concept of Kullback information (KI).

³ Note that the exclusion restrictions for the system as a whole may be written as

$$L^{*'} a = 0, \quad L^* = \text{diag}(L_1^*, L_2^*, \dots, L_m^*).$$

If the prior restrictions are not imposed **directly**, as has been the universal practice in this literature and as we shall do below, but rather in form of Lagrange multipliers, we shall be afforded a routine instrumentality for testing the validity of some or all of the **overidentifying** restrictions. For a discussion of this approach in the case of 2SLS and 3SLS see Dhrymes (1994a).

Definition 1. Let Ω be a countable (discrete) set and let \mathcal{A} be the set of all subsets of Ω ; let P and Q be two probabilities defined on \mathcal{A} , such that $P(\omega_i) = 0$ whenever $Q(\omega_i) = 0$, together with the conventions that $0 \ln 0 = 0$, and $0/0 = 0$. The **Kullback Information** of P on Q is given by

$$K(P, Q) = \sum_{i=1}^{\infty} P(\omega_i) \ln \left(\frac{P(\omega_i)}{Q(\omega_i)} \right).$$

Definition 2 (Generalization). Let P and Q be probability measures defined on the measurable space (Ω, \mathcal{A}) , and suppose that Q is absolutely continuous with respect to P . Suppose further that P and Q are dominated by a measure μ , in the sense that there exists a measure μ , and integrable functions f and g , such that for every set $A \in \mathcal{A}$

$$P(A) = \int_A f d\mu, \quad Q(A) = \int_A g d\mu.$$

The Kullback Information of P on Q is defined by

$$K(P, Q) = \int_{\Omega} f \ln \left(\frac{f}{g} \right) d\mu.$$

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, (as well as that of the exogenous variables) indexed on the parameter θ which comprises the parameter triplet (B, C, Σ) , or in the case of nonstochastic exogenous variables, given the space of the exogenous variables Ξ .⁴

In the context created above, 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 may be rendered as

$$K(\theta, \theta_0) = \int_{R^m} \ln \left(\frac{f_{\theta_0}}{f_{\theta}} \right) f_{\theta_0} d\mu. \quad (11)$$

⁴ The discussion of this section is, in part, based on Chs. 2 and 3, vol. II, of Dacunha-Castell and Dufflo (1986).

Remark 1. We note that in the case under consideration, for $A \in \mathcal{A}$, $\mathcal{P}_\theta(A)$ gives the probability that the dependent variables of the problem obey $y \in B$, where $A = y^{-1}(B)$; thus, if L^* is the likelihood, not the loglikelihood, function of the observations then

$$\mathcal{P}_\theta(A) = \int_B L^*(\theta) d\mu, \quad (12)$$

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

$$K(\theta, \theta_0) = \int_{R^m} \ln \left(\frac{L^*(\theta_0)}{L^*(\theta)} \right) L^*(\theta_0) d\mu = E_0 L(\theta_0) - E_0 L(\theta) \geq 0, \quad (13)$$

where $L(\theta) = \ln L^*(\theta)$. 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 3. Consider the probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and the econometric model $(\Omega, \mathcal{A}, \mathcal{P}_\theta)$, $\theta \in \Theta \subset R^n$, 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 4. In the context of Definition 3, let $X = \{X'_t : t = 1, 2, 3, \dots, T\}$ be a sequence of random vectors (elements), and consider the (nested) sequence of subalgebras⁵

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

A **contrast**, relative to θ_0 and K , is a function⁶

⁵ 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 $\mathcal{G}_T = \sigma(X_1, X_2, \dots, X_T)$, we will have produced the sequence referred to in the text, which is quite suitable for studying the samples $\{X_{(T)} : T \geq 1\}$.

⁶ In the description of the function, \mathcal{N}_+ represents the natural numbers, i.e. $\mathcal{N}_+ = \{1, 2, \dots\}$.

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

independent of θ_0 , such that

- i. for every $\theta \in \Theta$, $H_T(\theta, \omega)$ is \mathcal{G}_T -measurable;
- ii. $H_T(\theta, \cdot)$ converges to the contrast function $K(\theta, \theta_0)$, at least in probability.⁷

A **minimum contrast (MC) estimator** 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 3 and 4, suppose, further,

- i. $\Theta \subset R^n$ is closed and bounded (compact);
- ii. $K(\theta, \theta_0)$, and $H_T(\theta, \omega)$ are continuous in θ ;
- iii. letting

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

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

- iv. (identification condition) if $\inf_{\theta \in \Theta} K(\theta, \theta_0) = K(\theta^*, \theta_0)$ **then** $\theta^* = \theta_0$.

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

$$K(\theta, \theta_0) > 2\epsilon, \text{ for } \theta \in \bar{B}, \tag{14}$$

⁷ 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}_{θ_0} .

where

$$B = \{\theta : |\theta - \theta_0| < \epsilon\}. \quad (15)$$

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

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

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

$$\Theta^* \subset \bigcup_{i=1}^N A_i, \quad \text{with } A_i = \{\theta : |\theta - \theta_i| < \epsilon_T\}. \quad (16)$$

Next, note that we can write $H_T(\theta) = H_T(\theta_i) - [H_T(\theta_i) - H_T(\theta)]$, so that $H_T(\theta) \geq H_T(\theta_i) - |H_T(\theta_i) - H_T(\theta)|$. Consequently, we obtain

$$\begin{aligned} \inf_{\theta \in \Theta^*} H_T(\theta) &\geq \inf_{1 \leq i \leq N} H_T(\theta_i) - \sup_{\theta_i \in D} \sup_{|\theta_i - \theta| < \delta_T} |H_T(\theta_i) - H_T(\theta)| \\ &\geq \inf_{1 \leq i \leq N} H_T(\theta_i) - c_T(\delta_T). \end{aligned} \quad (17)$$

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 $N(\theta_0; \epsilon) = \{\theta : |\theta - \theta_0| < \epsilon\}$, 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

$$\begin{aligned} B_T &= \{\omega : \hat{\theta}_T \in \Theta^*\}, \quad C_T = \{\omega : \inf_{\theta \in \Theta^*} [H_T(\theta) - H_T(\theta_0)] < 0\} \\ D_T &= \{\omega : \inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] - c_T(\delta_T) < 0\}, \\ E_T &= \{\omega : \inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] < \epsilon_T\}, \quad F_T = \{\omega : c_T(\delta_T) > \epsilon_T\}, \end{aligned} \quad (18)$$

and note that

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

If $\hat{\theta}_T$ converges in Θ^* then for all (sufficiently large) T

$$\mathcal{P}(E_T) > 0, \quad \text{and indeed } \lim_{T \rightarrow \infty} \mathcal{P}(E_T) > 0. \quad (19)$$

We show that this implies a contradiction. To this end, note that for $c_T(\delta_T) \leq \epsilon_T$

$$D_T \cap \bar{F}_T = \{\omega : \inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] < c_T(\delta_T), \text{ and } c_T(\delta_T) \leq \epsilon_T\} \subseteq E_T. \quad (20)$$

Since

$$D_T = (D_T \cap \bar{F}_T) \cup (D_T \cap F_T) \subset (E_T \cup F_T), \quad (21)$$

it follows that

$$\mathcal{P}(B_T) \leq \mathcal{P}(C_T) \leq \mathcal{P}(E_T \cup F_T) \leq \mathcal{P}(E_T) + \mathcal{P}(F_T). \quad (22)$$

By iii of the premises of the proposition, $\mathcal{P}(F_T) \rightarrow 0$, and by Definition 3 and Corollary 4 Dhrymes (1989) p. 147,

$$\inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] \xrightarrow{P} \inf_{1 \leq i \leq N} K(\theta, \theta_i) - K(\theta_0, \theta_0) \geq 2\epsilon. \quad (23)$$

This is a **contradiction** if $\mathcal{P}(E_T) > 0$, since for all T the left member above is negative while the right member is positive. Hence

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

But this means that $\lim_{T \rightarrow \infty} \mathcal{P}(\bar{B}_T) = 1$; since $\bar{B}_T = \{\omega : \hat{\theta}_T \in N(\theta_0; \epsilon)\}$, and ϵ is arbitrary, we have that $\hat{\theta}_T$ is consistent for θ_0 .

q.e.d.

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

$$H_T(\theta) - H_T(\theta_0) \xrightarrow{a.c.} K(\theta, \theta_0)$$

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: We note that under the premises of the corollary, $\mathcal{P}(\overline{\lim}_{n \rightarrow \infty} F_n) = 0$. 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 $\hat{\theta}_T$ converges in Θ^* . We show a contradiction. Consider the “event”⁸

$$\{\omega : \hat{\theta}_T \in \Theta^*, i.o.\} = \overline{\lim}_{T \rightarrow \infty} B_T.$$

In view of the preceding we obtain

$$\overline{\lim}_{T \rightarrow \infty} B_T \subset \overline{\lim}_{T \rightarrow \infty} C_T \subset \left(\overline{\lim}_{T \rightarrow \infty} E_T \cup \overline{\lim}_{T \rightarrow \infty} F_T \right);$$

⁸ The notation *i.o.* means infinitely often.

Since $\mathcal{P}(\overline{\lim}_{T \rightarrow \infty} F_T) = 0$,

$$\mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} B_T\right) \leq \mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} C_T\right) \leq \mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} E_T\right).$$

If $\mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} E_T\right) > 0$, we have a contradiction since the premises imply

$$\inf_{1 \leq i \leq N} [H_T(\theta_i) - H_T(\theta_0)] \xrightarrow{\text{a.c.}} \inf_{1 \leq i \leq N} K(\theta_i, \theta_0) \geq 2\epsilon > 0. \quad (24)$$

Consequently, $\mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} B_T\right) = 0$. Or more directly, since the convergence is **uniform in θ** ,

$$\inf_{\theta \in \Theta^*} [H_T(\theta) - H_T(\theta_0)] \xrightarrow{\text{a.c.}} \inf_{\theta \in \Theta^*} K(\theta, \theta_0) > 2\epsilon,$$

which is a contradiction. This is so since the left member is negative for all T but it converges a.c. to the right member which is positive. Thus,

$$\mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} C_T\right) = 0, \quad \text{and, consequently,} \quad \mathcal{P}\left(\overline{\lim}_{T \rightarrow \infty} B_T\right) = 0. \quad (25)$$

Since

$$\overline{\lim}_{T \rightarrow \infty} B_T = \underline{\lim}_{T \rightarrow \infty} \bar{B}_T, \quad \text{and} \quad \bar{B}_T = \{\omega : \hat{\theta}_T \in N(\theta_0; \epsilon)\}, \quad (26)$$

the argument above implies $\mathcal{P}(\underline{\lim}_{T \rightarrow \infty} \bar{B}_T) = 1$, which means that the event $\{\omega : \hat{\theta}_T \in N(\theta_0; \epsilon), \text{ i.o.}\}$ has probability one, i.e. that the ML estimator, $\inf_{\theta \in \Theta} H_T(\theta) = H_T(\hat{\theta}_T)$, obeys

$$\hat{\theta}_T \xrightarrow{\text{a.c.}} \theta_0,$$

thus converging a.c. to the true parameter θ_0 .

q.e.d.

Remark 2. Notice that in the proof of Theorem 1 we place no restrictions on how⁹ the parameters of the problem enter the function H_T ; nor do we require that the likelihood function be of a specific form, or that the observations be i.i.d.. The strongest condition imposed is the “smoothness” condition in iii which will be satisfied if H is **bounded by an integrable** function of the observations and, at any rate, is rather mild by the standards of the literature of econometrics. Thus the results of Theorem 1 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. For an application to the general

⁹ Meaninging whether they enter the model linearly or nonlinearly.

nonlinear model or the general nonlinear simultaneous equations model with additive errors see, for example, Dhrymes (1994c).

Indeed, the class of estimators referred to in the literature as M-estimators are MC estimators. It would certainly facilitate matters, at least in a pedagogical sense, if M-estimators were discussed in the context created above; the use of the contrast function will render the discussion of identification and consistency far simpler to carry out.

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 of the ML estimator in the context of the standard GLSEM. First, we show that the LF of 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

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

and note that $H_T^*(\theta) = -L_T(\theta)$ is a **contrast** in the sense of Definition 4. In fact, we shall not violate the sense of Definition 4, if we put

$$H_T(\theta) = H_T^*(\theta) - H_T^*(\theta_0), \quad (28)$$

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 Ch. 4 Dhrymes (1984), that

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

which is¹⁰ **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 identification is

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

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 **dynamic and stable**,

$$\begin{aligned}
H_T &\xrightarrow{\text{a.c.}} K(\theta, \theta_0), \\
K(\theta, \theta_0) &= \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \ln |B^{*'} B^*| + \frac{1}{2} \text{tr} \Sigma^{-1} Q \\
&\quad - \frac{m}{2} - \frac{1}{2} \ln |\Sigma_0| + \frac{1}{2} \ln |B_0^{*'} B_0^*|, \\
Q &= [B^{*'} \Omega_0 B^* + (A - A_0)' P_0 (A - A_0)] \\
P_0 &= (\Pi_0, I_G)' M_{xx} (\Pi_0, I_G), \quad \frac{X' X}{T} \xrightarrow{\text{a.c.}} M_{xx}. \quad (29)
\end{aligned}$$

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

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

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

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

We have

$$\frac{U' U}{T} = \frac{1}{T} \sum_{t=1}^T u_t' u_t \xrightarrow{\text{a.c.}} \Sigma_0. \quad (31)$$

This is so since $\{u_t : t \geq 1\}$ 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. 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 is of the form

$$\frac{U' X}{T} = \frac{1}{T} \sum_{j=1}^T u_t' x_t \xrightarrow{\text{a.c.}} 0, \quad (32)$$

by the discussion in the Appendix. Consequently, the second component, obtained from the reduced form representation, obeys

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

Finally, since by the preceding discussion and that in the Appendix

$$\begin{aligned} \tilde{M}_{zz} &= \frac{1}{T}Z'Z = \frac{1}{T} \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \stackrel{\text{a.c.}}{\rightarrow} \begin{bmatrix} \Omega_0 & 0 \\ 0 & 0 \end{bmatrix} + P_0, \\ P_0 &= (\Pi_0, I_G)' M_{xx}(\Pi_0, I_G), \end{aligned} \quad (34)$$

we may establish, after some manipulation, that (**uniformly**) for every $\theta \in \Theta$

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

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

$$\begin{aligned} L_T(\theta) &\stackrel{\text{a.c.}}{\rightarrow} \bar{L}(\theta, \theta_0), \\ &= -\frac{m}{2} - \ln|\Sigma| + \frac{1}{2}\ln|B^{*\prime}B^*| + \text{tr}\Sigma^{-1}Q, \quad \text{where} \\ Q &= B^{*\prime}\Omega_0 B^* + (A_0^* - A^*)' P_0 (A_0^* - A^*). \end{aligned} \quad (36)$$

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

$$\begin{aligned} H_T(\theta) &\stackrel{\text{a.c.}}{\rightarrow} \bar{L}(\theta_0, \theta_0) - \bar{L}(\theta, \theta_0) \\ &= -\frac{1}{2} \left(m + \ln|\Sigma_0| - \ln|B_0^{*\prime}B_0^*| \right) + \frac{1}{2} \left(\ln|\Sigma| - \ln|B^{*\prime}B^*| \right) \\ &\quad + \text{tr}\Sigma^{-1}Q. \end{aligned} \quad (37)$$

Defining now

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

we shall now show that the function K , above, is the asymptotic KI of the problem, i.e. the limit of $E_{\theta_0}L_T(\theta_0) - E_{\theta_0}L_T(\theta)$. To this end, we observe that

$$\begin{aligned} E_{\theta_0}L_T(\theta_0) &= -\frac{m}{2}\ln(2\pi) - \frac{m}{2} - \frac{1}{2}\ln|\Sigma_0| + \frac{1}{2}\ln|B_0^{*\prime}B_0^*| \\ E_{\theta_0}L_T(\theta) &= -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| + \frac{1}{2}\ln|B^{*\prime}B^*| + \text{tr}\Sigma^{-1}Q. \end{aligned} \quad (39)$$

Consequently, defining the sample based KI by $K_T(\theta_0, \theta) = E_{\theta_0} L_T(\theta_0) - E_{\theta_0} L_T(\theta)$, we find

$$K_T(\theta, \theta_0) = -\frac{1}{2} \left(\ln |\Sigma_0| - \ln |B_0^* B_0^{*'}| \right) - \frac{m}{2} + \frac{1}{2} \left(\ln |\Sigma| - \ln |B^* B^{*'}| \right) \\ + \frac{1}{2} \text{tr} \Sigma^{-1} \tilde{Q}, \quad \tilde{P}_0 = (\Pi_0, I_G)' \tilde{M}_{xx} (\Pi_0, I_G) \\ \tilde{Q} = B^{*'} \Omega_0 B^* + (A_0^* - A^*)' \tilde{P}_0 (A_0^* - A^*), \quad \tilde{M}_{xx} = \frac{X' X}{T}, \quad (40)$$

and it may be verified quite easily that the asymptotic KI is given by

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

An immediate consequence of the preceding is

Corollary 2. Under assumptions A.1 through A.4, 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

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

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

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

If $\bar{\theta}$ is the point at which K attains its global minimum then: (a) $\hat{\theta}_T \xrightarrow{\text{a.c.}} \bar{\theta}$, 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 $\hat{\theta}_T \xrightarrow{\text{a.c.}} \theta_0$, 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

$$K(\theta, \theta_0) = -\frac{1}{2} m - \frac{1}{2} \ln |\Sigma_0| + \frac{1}{2} \ln |B_0^* B_0^{*'}| + \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \ln |B^* B^{*'}| \\ + \frac{1}{2} \text{tr} \Sigma^{-1} Q.$$

Noting that $\Omega_0 = B_0^{*'}\Sigma_0 B_0^{*-1}$ and, therefore, that $B_0^{*'}\Omega_0 B_0^* = \Sigma_0$, we can rewrite the (asymptotic) KI of Eq. (40) as

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

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

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

whence we obtain

$$\Sigma = Q.$$

Noting that

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

and inserting the minimizer in Eq. (41), we obtain the ‘‘concentrated’’ KI expression,

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

Remark 3. Since the expression in the large round bracket is equal to or greater than unity, K^* 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. (42) **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. (41) 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**,¹¹ the **restrictions required** were such that the **intersection of the null space of P_0 and the class of admissible structures**

¹¹ In this context a matrix A^* is said to be admissible, as in the standard context, if and only if it satisfies all prior restrictions.

has A_0^* as its only member, in the sense that $P_0(A_0^* - A^*) = 0$ if and only if $A_0^* = A^*$. Evidently, this would establish identification!

In Remark 3 we had 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 (nas) 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

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

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

$$a_{.i} - a_{.i}^0 = L_i(\delta_{.i} - \delta_{.i}^0), \quad \text{tr}\Psi = (\delta - \delta^0)' L'(I_m \otimes P_0)L(\delta - \delta^0). \quad (44)$$

In the framework of Eq. (43) a nas condition for identification of the parameters of the system is that the **block diagonal** matrix $L'(I_m \otimes P_0)L$ 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

$$\text{rank}(S_i) = \text{rank}(\Pi L_i, L_{2i}) = m_i + G_i, \quad \text{for every } i = 1, 2, \dots, m. \quad (45)$$

By Theorem 5 and Corollary 1, in Chapter 3 Dhrymes (1994c), the conditions above are the nas conditions for the identification of the parameters in the i^{th} equation, and the system as a whole. Thus, we have derived the nas 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 the strong consistency for the ML estimator. We may summarize the preceding discussion in

Theorem 3. Consider the GLSEM of Eq. (1), subject to Conventions 1, 2 and assumptions A.1 through A.4.; suppose further that it satisfies the KI based identification condition (KIBIC)

$$\inf_{\theta \in \Theta} K(\theta, \theta_0) = K(\theta^*, \theta_0), \text{ implies } \theta^* = \theta_0.$$

Then

- i. the ML estimator obeys $\hat{\theta}_T \xrightarrow{\text{a.c.}} \theta_0$;
- ii. the KIBIC holds if and only if Eq. (44) holds, i.e. if the standard identification requirements are valid.

Remark 4. Notice that the identification result for the GLSEM of Eq. (1) **in no way depends** on whether the error process is Gaussian; nor does it depend on whether the structural errors are i.i.d.. It only depends on whether the conditions of Theorem 1 are applicable; even if the errors were not normal **but, engaging in a quasi-ML exercise**, we define the function to be minimized by

$$H_T(\theta) = \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \ln |B^{*'} B^*| + \frac{1}{2} \text{tr} \Sigma^{-1} A^{*'} \tilde{M}_{zz} A^* \\ - \frac{1}{2} \ln |\Sigma_0| - \frac{1}{2} \ln |B_0^{*'} B_0^*| + \frac{1}{2} \text{tr} \Sigma_0^{-1} A_0^{*'} \tilde{M}_{zz} A_0^*$$

we shall establish, as its a.c. limit, precisely the function $K(\theta, \theta_0)$ above, even if normality is **not assumed**. The only difference in this context is that we **must show** that K is a contrast function; thus, the normality assumption is only incidental in that it allows us to write the particular LF and, given that we are dealing with a LF function, we are assured that K is KI and as such it is a contrast function. However, the results obtained are of much wider applicability, and depend only on the properties of the function to be minimized in order to obtain the MC estimator, and on whether the limit to which it converges is a contrast function. If these two conditions hold then detailed identification conditions may be routinely obtained for the GLSEM, or for nonlinear models. It is this feature of the MC estimator framework that makes it very appealing.

Finally if the errors are dependent, certain complications will arise which are related to subsidiary issues, such as the convergence of the function to be minimized to an appropriate probability or a.c. limit. Whereas these issues may be easily dealt with by proper modification of the model specification, they lie outside the purview of this paper.

REFERENCES

- Christ, C. F. (1966), *Econometric Models and Methods*, New York: Wiley.
- 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: vol. I, Probability Foundations*, New York: Springer Verlag.
- (1994a), "Specification Tests in Simultaneous Equations Models", *Journal of Econometrics*, vol. 64, pp. 45-76.
- (1994b), "On the Estimation of Systems of Equations with Autoregressive Errors and Singular Covariance Matrices", *Econometric Theory*, vol. 10, pp. 254-285.
- (1994c), *Topics in Advanced Econometrics: vol. II, Linear and Nonlinear Simultaneous Equations*, New York: Springer-Verlag.
- Hall, P. and C.C. Heyde (1980), *Martingale Theory and its Application*, New York: Academic Press.

APPENDIX

In this appendix we provide the details of the convergence arguments for the loglikelihood function, and demonstrate the asymptotic normality of the ML estimator. We do so for the case where the model is dynamic and stable and the **exogenous** variables are generated by a zero mean, square integrable stochastic process obeying the condition in Eq. (3); or, alternatively, they lie in a compact set $\Xi \subset R^s$, i.e. for every t , $p_t \in \Xi$, and have the property

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T p_t' p_t = M_{pp} > 0.$$

Consider the model in Eq. (1) as amplified in the discussion following, and write its reduced form as

$$y_t' = \Pi_1 y_{t-1}' + \dots + \Pi_k y_{t-k}' + \Pi_0 p_t' + D u_t',$$

or in more compact form

$$\Pi(L) y_t' = \Pi_0 p_t' + D' u_t', \quad \Pi(L) = I_m - \sum_{j=1}^k \Pi_j L^j,$$

where L , in the context of this discussion only, is the usual lag operator. Let $H(L)$ be the matrix adjoint to $\Pi(L)$ and $h(L) = |\Pi(L)|$. Since the model is **stable** we may write

$$[\Pi(L)]^{-1} = \frac{H(L)}{h(L)},$$

so that each element of the inverse matrix operator is a ratio of polynomial operators of degree $(m-1)k$ and mk , respectively, and it possesses an absolutely converging power series expansion. We may thus obtain the **final form**

$$y_t' = \frac{H_1(L)}{h(L)} p_t' + \frac{H_2(L)}{h(L)} u_t', \quad \text{with } H_1(L) = H(L)\Pi_0, \quad H_2(L) = H(L)D'.$$

Putting

$$p_t^{*'} = \frac{H_1(L)}{h(L)} p_t' = \sum_{s=0}^{\infty} F_{1s} p_{t-s}', \quad \sum_{j=0}^{\infty} \|F_{1j}\| < \infty$$

$$u_t^{*'} = \frac{H_2(L)}{h(L)} u_t' = \sum_{s=0}^{\infty} F_{2s} u_{t-s}', \quad \sum_{j=0}^{\infty} \|F_{2j}\| < \infty,$$

the final form may be written, more compactly in vector and matrix form, as

$$y'_t = p_t^* + u_t^*, \text{ or } Y = P^* + U^*$$

the first term being the “systematic” or exogenous component, the second being the “random component”.

From Eqs. (29) through (34) it is clear that the basic entities whose behavior needs to be determined are:

$$\frac{U'U}{T}, \quad \frac{X^*U}{T}, \quad \frac{X'X}{T}.$$

Examining the first term, we find

$$\frac{U'U}{T} = \sum_{t=1}^T u'_t u_t \xrightarrow{\text{a.c.}} \Sigma_0,$$

by Proposition 23 Dhrymes (1989) p. 188.

4.3 Convergence of $U'X/T$

Before we proceed, we note that having assumed the existence of the a.c. limit

$$\frac{1}{T} \sum_{t=1}^T p'_t p_t \xrightarrow{\text{a.c.}} \lim_{T \rightarrow \infty} \sum_{t=1}^T E p'_t p_t = M_{pp} > 0,$$

we must conclude that the a.c. limits

$$\hat{M}_{i,j} = \frac{1}{T} \sum_{t=1}^T p'_{t-i} p_{t-j} \xrightarrow{\text{a.c.}} M_{ij}$$

exist as well for all i, j , owing to the fact that

$$p'_{t-i} p_{t-j} + p'_{t-j} p_{t-i} \leq p'_{t-i} p_{t-i} + p'_{t-j} p_{t-j}.$$

We begin by noting that

$$\begin{aligned} \frac{U'X}{T} &= \frac{1}{T} \left(\sum_{t=1}^T u'_t p_{t-1}^*, \dots, \sum_{t=1}^T u'_t p_{t-k}^*, \sum_{t=1}^T u'_t p_t \right) \\ &\quad + \frac{1}{T} \left(\sum_{t=1}^T u'_t u_{t-1}^*, \dots, \sum_{t=1}^T u'_t u_{t-k}^*, 0 \right), \end{aligned}$$

and defining the entities

$$\zeta_{t1} = \text{vec}(u'_t p_t), \quad \zeta_{t2} = \text{vec}(u'_t p_{t-i}^*), \quad \zeta_{t3} = \text{vec}(u'_t u_{t-i}^*).$$

What is common amongst the three entities above is that they are zero mean **uncorrelated** random elements; the covariance matrix of these terms is given, respectively, by

$$\text{Cov}(\zeta_{t1}) = E p_{t \cdot} p_{t \cdot}' \otimes \Sigma, \quad \text{Cov}(\zeta_{t2}) = E p_{t-i \cdot}^* p_{t-i \cdot}^* \otimes \Sigma,$$

$$\text{Cov}(\zeta_{t3}) = \Phi_{00} \otimes \Sigma, \quad \Phi_{00} = \sum_{s=0}^{\infty} F_{2s} \Sigma F_{2s}'.$$

Moreover, by the result given in the appendix of Dhrymes (1994b), all such matrices are dominated by $C_1 t^\alpha$, $\alpha \in [0, \frac{1}{2})$, where C_1 is a matrix of finite constants. Consequently, by Proposition 26 Dhrymes (1989) p. 193,

$$\frac{1}{T} \sum_{t=1}^T \zeta_{ti} \xrightarrow{\text{a.c.}} 0, \quad i = 1, 2, 3, \quad \text{and thus} \quad \frac{U' X}{T} \xrightarrow{\text{a.c.}} 0.$$

4.3.1 Convergence of $X' X / T$

In this section the basic entities are

$$\frac{P_{-i}^* P_{-j}^*}{T}, \quad \frac{P_{-i}^* P}{T}, \quad \frac{P' P}{T}, \quad \frac{U_{-i}^* U_{-j}^*}{T}, \quad \frac{P_{-i}^* U_{-j}^*}{T}, \quad \frac{P' U_{-j}^*}{T} \quad i, j = 1, 2, \dots, k.$$

The first three of these entities may be handled by the following generic argument. Recall that

$$\hat{M}_{i,j} = \frac{1}{T} \sum_{t=1}^T p_{t-i \cdot} p_{t-j \cdot}' \xrightarrow{\text{a.c.}} M_{i,j},$$

and consider

$$D_T = \frac{P_{-i}^* P_{-j}^*}{T} - M_{ij}^*, \quad M_{ij}^* = \sum_{s=0}^{\infty} \sum_{r=0}^{\infty} F_{1s} M_{i+s, j+r} F_{1r}'.$$

We shall show that $D_T \xrightarrow{\text{a.c.}} 0$. We first note that

$$\| D_T \| \leq \sum_{s=0}^{\infty} \sum_{r=0}^{\infty} \| F_{1s} \| \| F_{1r} \| \| \hat{M}_{i+s, j+r} - M_{i+s, j+r} \|,$$

$$\hat{M}_{i+s, j+r} = \frac{1}{T} \sum_{t=1}^T p_{t-i-s \cdot} p_{t-j-r \cdot}', \quad \text{and} \quad \hat{M}_{i+s, j+r} \xrightarrow{\text{a.c.}} M_{i+s, j+r}.$$

In view of the last equation above, there exists a constant k and $\delta > 0$, (but arbitrarily small) such that for all $T > T(s, r)$

$$A_{T(s,r)} = \{ \omega : \| \hat{M}_{i+s, j+r} - M_{i+s, j+r} \| < k \}, \quad \text{and} \quad \mathcal{P}(A_{T(s,r)}) \geq 1 - \delta.$$

Thus, for $T > T_0$ and

$$T_0 = \sup_{s,r \geq N} T(s,r)$$

we have, with probability at least $1 - \delta$,

$$\begin{aligned} \| D_T \| &\leq \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} \| F_{1s} \| \| F_{1r} \| \| \hat{M}_{i+s,j+r} - M_{i+s,j+r} \| \\ &+ k \sum_{s=N}^{\infty} \sum_{r=N}^{\infty} \| F_{1s} \| \| F_{1r} \| . \end{aligned}$$

Letting $T \rightarrow \infty$ we find

$$\| D_T \| \xrightarrow{\text{a.c.}} k^* \leq k \sum_{s=N}^{\infty} \sum_{r=N}^{\infty} \| F_{1s} \| \| F_{1r} \| .$$

Finally, letting $N \rightarrow \infty$, we obtain the desired conclusion, viz.

$$\frac{P_{-i}^{*'} P_{-j}^*}{T} \xrightarrow{\text{a.c.}} M_{ij}^*, \quad i, j = 1, 2, \dots, k.$$

Consider now the entities

$$\frac{U_{-i}^{*'} U_{-j}^*}{T}, \quad \frac{P_{-i}^{*'} U_{-j}^*}{T}, \quad \frac{P_{-i}^{*'} U_{-j}^*}{T},$$

and note that u_t^* and $u_t^{*'} u_t^*$ are **mixingales**, see Hall and Heyde (1980), (HH), p. 19. The parameters ψ_m , c_n in HH's notation are given by

$$c_n = \| \Sigma \|, \quad \psi_m = \left(\sum_{j=m}^{\infty} \| F_{2s} \|^2 \right)^{1/2},$$

in the case of u_t^* , and by

$$c_n = K, \quad \psi_m = \sum_{j=m}^{\infty} \left(\sum_{j=m}^{\infty} \| F_{2s} \| \right)^2,$$

in the case of $\text{vec}(u_t^{*'} \otimes u_t^*)$, where K is a bound on fourth order moments; this is a uniform finite bound in view of the fact that the structural error process is normal.¹² Thus, the a.c. convergence of the first entity

¹² This is the only instance, in this Appendix, where the normality assumption regarding the error process plays a role by ensuring, without additional conditions, that the fourth order moments are uniformly bounded; note that we would have obtained the same result if we only assumed the errors to be independent with uniformly bounded fourth moments, or even weaker conditions.

is implied by Theorem 2.1 in HH, p. 41, if $\sum_{s=m}^{\infty} \|F_{2s}\|$ converges to zero at the rate of m^{-a} , where $a > (1/4)$. This is so since

$$m^{(1/2)-2\alpha} (\ln m)^2 \rightarrow 0,$$

under the conditions specified. Thus, we obtain

$$\frac{U_{-i}^* U_{-j}^*}{T} \xrightarrow{\text{a.c.}} \Phi_{ij}, \quad \Phi_{ij} = \sum_{s=0}^{\infty} F_{2s} \Sigma F_{2,i-j+s}, \quad i, j = 1, 2, \dots, k.$$

If we take the exogenous variables to be nonstochastic and lie in a compact set Ξ , it is clear that convergence to zero is implied for $P_{-i}^* U_{-j}^* / T$ by the result in HH cited above, because $p_{i,t}^* u_{i,t}^*$ is a mixingale, and thus we have convergence a.c. to its mean, which is zero. In the case of stochastic exogenous variables the same result will continue to hold, in view of the assumption that $p_{i,t}$ is independent of $u_{i,t}$; this is so since if we condition on $\mathcal{G}_t = \sigma(p_s, s \leq t)$ the sequence $(p_{i,t}^* \otimes u_{i,t}^* | \mathcal{G}_t)$ is a mixingale, as in the nonstochastic case, and the result continues to hold. With this argument we have completed the derivation of the required results; specifically, we have obtained

$$\begin{aligned} \frac{U'U}{T} &\xrightarrow{\text{a.c.}} \Sigma_0, \quad \frac{X'X}{T} \xrightarrow{\text{a.c.}} M_{xx}, \\ M_{xx} &= \begin{bmatrix} M_{11}^* + \Phi_{11} & M_{12}^* + \Phi_{12} & \dots & M_{1k}^* + \Phi_{1k} & M_{10}^* \\ M_{21}^* + \Phi_{21} & M_{22}^* + \Phi_{22} & \dots & M_{2k}^* + \Phi_{2k} & M_{20}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ M_{k1}^* + \Phi_{k1} & M_{k2}^* + \Phi_{k2} & \dots & M_{kk}^* + \Phi_{kk} & M_{k0}^* \\ M_{01}^* & M_{02}^* & \dots & M_{0k}^* & M_{00}^* \end{bmatrix}, \\ \frac{P'P_{-j}^*}{T} &\xrightarrow{\text{a.c.}} M_{0j}^*, \quad M_{j0}^* = M_{0j}^*, \quad \frac{P'P}{T} \xrightarrow{\text{a.c.}} M_{00}, \quad \Phi_{ii} = \Phi_{00}. \end{aligned}$$

4.3.2 Limiting Distribution

Admitting stochastic exogenous variables introduces a certain complication into the problem of the limiting distribution of ML estimators; this section is devoted to the resolution of this issue. It is shown in Dhrymes (1994c) p. 214, that the limiting distribution in question depends crucially on

$$\frac{1}{\sqrt{T}} (I_G \otimes X') u = \frac{1}{\sqrt{T}} \sum_{t=1}^T (I_G \otimes x'_t) u'_t = \frac{1}{\sqrt{T}} \sum_{t=1}^T \zeta'_t.$$

In that discussion, the exogenous variables were specified to be non-stochastic and lie in a compact subset of R^s . Now, construct the stochastic basis

$$\mathcal{A}_t = \sigma(p_r., r \leq t + 1, u_s., s \leq t)$$

and note that $\{(\zeta_t., \mathcal{A}_t) : t \geq 1\}$ is a **martingale difference**, since it is integrable and $E(\zeta_t. | \mathcal{A}_{t-1}) = 0$. Moreover, it may be shown that it satisfies a **Lindeberg** condition, and in the preceding discussion we have shown that

$$\frac{X'X}{T} \xrightarrow{\text{a.c.}} M_{xx} > 0,$$

and obtained the precise form of that matrix. Hence, by Proposition 21 Dhrymes (1989) p. 337, we conclude that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T (I_G \otimes x'_t) u'_t \xrightarrow{d} N(0, \Sigma \otimes M_{xx}).$$

The remaining arguments remain unaffected, and may be obtained following the discussion in Dhrymes (1994c) p. 212ff.

1995-1996 Discussion Paper Series
Department of Economics
Columbia University
1022 International Affairs Bldg.
420 West 118th Street
New York, N.Y., 10027

The following papers are published in the 1995-96 Columbia University Discussion Paper series which runs from early November to October 31 (Academic Year). Domestic orders for discussion papers are available for purchase at \$8.00 (US) each and \$140.00 (US) for the series. Foreign orders cost \$10.00 (US) for individual paper and \$185.00 for the series. To order discussion papers, please send your check or money order payable to Department of Economics, Columbia University to the above address. Be sure to include the series number for the paper when you place an order.

- 9596-01 Protectionist Response to Import Competition in Declining Industries Reconsidered
 J.P. Choi
- 9596-02 New Estimates on Climate Demand: Evidence from Location Choice
 M. Cragg
 M. Kahn
- 9596-03 Enforcement by Hearing
 C.W. Sanchirico
- 9596-04 Preferential Trading Areas and Multilateralism: Strangers, Friends or Foes?
 J. Bhagwati
 A. Panagariya
- 9596-05 Simplification, Progression and a Level Playing Field
 William Vickrey
- 9596-06 The Burden of Proof in Civil Litigation
 C. Sanchirico
- 9596-07 Market Structure and the Timing of Technology Adoption
 J.P. Choi
 M. Thum
- 9596-08 The Emergence of the World Economy
 R. Findlay
- 9596-09 The Global Age: From a Skeptical South to a Fearful North
 J. Bhagwati
- 9596-10 A Conformity Test for Cointegration
 P.J. Dhrymes
- 9596-11 Identification and Kullback Information in the GLSEM
 P.J. Dhrymes