8-1-1966

Structural Restrictions and Estimation Efficiency in Linear Econometric Models

Thomas J. Rothenberg

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 213

STRUCTURAL RESTRICTIONS AND ESTIMATION EFFICIENCY IN LINEAR ECONOMETRIC MODELS

Thomas Rothenberg

August 23, 1966

STRUCTURAL RESTRICTIONS AND ESTIMATION EFFICIENCY IN LINEAR ECONOMETRIC MODELS

by

Thomas Rothenberg*

I. INTRODUCTION

One of the important stochastic models used in econometric research is the system of linear regression equations commonly known as the simultaneous equations model. This model, in its simplest form, relates a vector of random variables y linearly to a vector of predetermined variables x and an additive random error vector u :

$$By + \Gamma x = u$$

where B and Γ are matrices of parameters. The error vector u is usually assumed to be normally distributed with mean zero and covariance matrix Σ . A major econometric problem is to estimate the structural parameters B , Γ, and Σ (or certain functions of them) based on a random sample on x and y . In particular, often the aim is to estimate the so-called reduced-form parameters

$$\Pi = - B^{-1}\Gamma \quad \text{and} \quad \Omega = B^{-1}\Sigma B'^{-1}$$

---

* The author is Assistant Professor of Economics at Northwestern University. This paper is a revised version of a chapter of the author's doctoral dissertation submitted to Massachusetts Institute of Technology. Part of the research for this paper was undertaken while the author was a member of the Cowles Foundation research staff in Spring, 1965.

which completely characterize the conditional distribution of y given x.

If B, Γ, and Σ are entirely unknown, the statistical analysis of the model is elementary: The structural parameters are unidentified and cannot be estimated; the reduced-form parameters can be estimated by the method of least squares which yields minimum-variance-unbiased estimates. However, the traditional treatment of the simultaneous equations model usually assumes that certain elements of B, Γ, and Σ are known a priori and need not be estimated. In this case, if enough structural information is known, it is possible to estimate B, Γ, and Σ; furthermore, it is possible to use the information to obtain reduced-form estimates more efficient than those given by least squares.

In its traditional form, the simultaneous equations problem is a special case of the following statistical problem: Let $f(x, \theta)$ be the likelihood function for a sample vector x and an unknown parameter vector θ. However, it is known a priori that θ belongs to some subset A of the possible parameter space. (In the simultaneous equations model, θ consists of the elements of Π and Ω; the restrictions on θ result from the a priori information on B, Γ, and Σ.) Under the classical minimum-variance-unbiased approach to statistical estimation theory, three interesting questions can be asked of such a model:

(1) When will the restrictions improve the efficiency of estimating θ?

(2) What is the optimal estimate of θ ?

(3) By how much is efficiency increased as a result of using the a priori information?

In an earlier study [14], the present author has investigated these questions by generalizing the classical Cramer-Rao inequality. Our purpose here is to apply these results to the simultaneous equations problem in an attempt to give a unified classical treatment of that topic. The value of overidentifying restrictions in increasing the efficiency of reduced-form estimation is analyzed in some detail. In addition, the well-known identification problem is viewed in a new light.[1]

2. THE MODEL[2]

Let y_t be a G-dimensional vector of random endogenous variables which are related to the K-dimensional vector of predetermined variables x_t by a system of G linear equations with additive random errors u_t :

$$By_t + \Gamma x_t = u_t \quad . \tag{2.1}$$

---

1 Our approach follows closely that of Koopmans, Rubin, and Leipnik [11]. Indeed a major part of this paper consists of rederiving and finding an explicit expression for equation (3.127) of that article. Our approach is also similar in spirit to that of Chipman [5] and Klein [9]. A few of the results of the present paper appear in a recent article by Rothenberg and Leenders [13] although the derivation and approach is quite different.

2 For a more complete description of the model, see Koopmans, Rubin, and Leipnik [11, pp. 54-75].

The $G \times G$ matrix B and the $G \times K$ matrix Γ contain unknown parameters which are to be estimated from the sample which consists of n observations on the $G + K$ variables y_t and x_t. The matrix B is assumed to be nonsingular. The error vectors $u_1, \ldots, u_n$ are assumed to be independently distributed normal random variables each with mean vector zero and nonsingular[1] covariance matrix Σ. The predetermined variables x_t are assumed to be distributed independently of the errors u_s for $t \leq s$.[2]

Since B is nonsingular, the structural form (2.1) may be written in the reduced form

$$(2.2)\qquad \begin{aligned} y_t &= - B^{-1}\Gamma x_t + B^{-1}u_t \\ &= \Pi x_t + v_t \end{aligned}$$

where Π is a $G \times K$ matrix and v_t is a normally distributed vector with mean zero and nonsingular covariance matrix Ω. The reduced-form parameters (Π, Ω) are related to the structural parameters (B, Γ, Σ) by the equations

$$(2.3)\qquad \begin{aligned} \Pi &= - B^{-1}\Gamma \\ \Omega &= B^{-1}\Sigma B'^{-1} . \end{aligned}$$

---

1 The assumption that Σ is nonsingular means (2.1) can contain no identities. This assumption is made solely for ease of exposition. In Appendix C the analysis is extended to include identities.

2 That is, the errors are independent of current and past predetermined variables. This permits lagged endogenous variables to be included among the predetermined variables.

The joint conditional density function for $y_1, \ldots, y_n$ given $x_1, \ldots, x_n$ is[1]

$$(2.4) \qquad f(y;\Pi, \Omega) = (2\pi)^{-\frac{1}{2}nG} |\Omega|^{-\frac{1}{2}n} \exp\left\{-\frac{1}{2}\sum_t (y_t - \Pi x_t)'\Omega^{-1}(y_t - \Pi x_t)\right\} .$$

This can be written more conveniently in terms of the observation matrices X, an $n \times K$ matrix of observations on the K predetermined variables, and Y, an $n \times G$ matrix of observations on the G endogenous variables. We shall assume that

$$(2.5) \qquad X'X = \sum_{t=1}^{n} x_t x_t' ,$$

the matrix of sums of squares and crossproducts of the predetermined variables, is nonsingular. The density function (2.4) can now be written in logarithmic form as

$$(2.6) \qquad \log f = K - \frac{1}{2} n \log \det \Omega - \frac{1}{2} \operatorname{tr} [\Omega^{-1}(Y' - \Pi X')(Y - X\Pi')]$$

where K is a constant.

The probability law for the endogenous variables (given X) is uniquely determined by the parameter matrices (Π, Ω). Yet for any pair (Π^o, Ω^o) there are an infinite set of different matrices (B, Γ, Σ)

---

1 More precisely, (2.4) is the conditional density for the endogenous variables $y_1, \ldots, y_n$ given those elements of $x_1, \ldots, x_n$ which are either (1) exogenous (i.e., not lagged values of y) or (2) lagged values y_t with $t \leq 0$. For a derivation of (2.4) see [11, pp. 72-73].

which satisfy (2.3). Hence the structure (2.1) is not identified unless some a priori constraints are placed on the parameters. We shall concentrate on the case where the identification restrictions take the form of knowing specific elements of B, Γ , and Σ . This seems to be the case most often met in practice and the one most often dealt with in the econometric estimation literature. The case of more general restrictions is discussed briefly in Section 8 but the results are very formal compared to the detailed analysis of the rest of the paper.

Knowledge of specific elements of (B, Γ, Σ) will be referred to as "zero-order" restrictions. The knowledge that a given element of (B, Γ, Σ) is zero will be called a homogeneous zero-order restriction. The knowledge that a parameter is a given nonzero number will be called a nonhomogeneous zero-order restriction.

Let α be a column vector consisting of those elements of (B, Γ, Σ) which are not known a priori. Let θ be a vector consisting of all the elements of (Π, Ω) . Then (2.3) is a set of equations of the form

$$\theta = h(\alpha) \tag{2.7}$$

and (2.4) is a probability function expressed in terms of θ alone. If the rank of the transformation h is less than the number of elements of θ , the allowed parameter space of θ is restricted by (2.7). Thus, if enough structural parameters are known a priori, the transformation (2.7) serves to express constraints on the reduced-form parameters θ .

We are therefore in a position to apply the results of a previous study [14] concerning the efficient estimation of the basic parameters θ and the constraint parameters α. Specifically, the purpose of the present paper is to use the general theory of constrained estimation to answer the following questions: (1) What increase in efficiency is gained in estimating the reduced form parameters θ by imposing the restrictions (2.7)? (2) Of what value are restrictions on the elements of Σ ? (3) When will the structural parameters α be identified? (4) What methods of estimation yield efficient estimates of θ and α ?

It will be convenient to distinguish between those elements of α and θ which refer to the regression coefficients (B, Γ, Π) and those elements which refer to the covariance matrices (Σ, Ω). Hence, the vectors α and θ are partitioned as

$$(2.8) \qquad \alpha = \begin{bmatrix} \delta \\ \sigma \end{bmatrix} \qquad \theta = \begin{bmatrix} \pi \\ \omega \end{bmatrix}$$

where δ is an r-dimensional column vector of all the <u>unknown</u> elements of B and Γ, σ is an r*-dimensional column vector of all the <u>unknown</u> elements of Σ, π is a GK-dimensional vector of all the elements of Π, and ω is a G^2-dimensional column vector of all the elements of Ω.

On forming vectors out of the elements of matrices it is necessary to specify the order in which the elements are listed. We shall follow the convention of first taking the elements of the first row, then the elements of the second row, then the elements of the third row, etc. Any matrix element that is known <u>a priori</u> will simply be omitted. Let π_i be the transpose of the i^{th} row of Π and let ω_i be the i^{th} column of Ω. Let β_i, γ_i, and σ_i be column vectors consisting of the unknown elements of the i^{th} row of B, Γ, and Σ, respectively. Then π, ω, δ, and σ may be written as

$$\pi = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_G \end{bmatrix}, \quad \omega = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_G \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_G \end{bmatrix}, \quad \sigma = \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_G \end{bmatrix},$$

where δ_i is the vector of unknown regression coefficients in the i^{th} structural equation:

$$\delta_i = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} .$$

In this study we shall concentrate on the efficient estimation of δ and π, treating σ and ω as nuisance parameters.

3. THE THEORY OF EFFICIENT ESTIMATION

3.1. Structural Restrictions Ignored

Before examining the usual situation where a number of structural parameters are known *a priori*, we shall first consider the case where B , Γ , and Σ are completely unrestricted. In this case structural estimation is impossible since the structure is not identified. Furthermore, equations (2.3) impose no restrictions on the parameter space of (Π, Ω) . Thus the density (2.4) may be considered as the likelihood function for (Π, Ω) with the structure completely ignored.

In order to study the efficiency of unconstrained reduced-form estimation, we shall use some general results of classical estimation theory. Suppose $f(x, \theta)$ is the likelihood function for a sample vector x and a parameter vector θ . By the famous Cramer-Rao inequality, a lower bound to the covariance matrix of any unbiased estimator of θ which does not use a priori restrictions is given by R_n^{-1} where

$$R_n = -E\left[\frac{\partial^2 \log f}{\partial\theta_i\,\partial\theta_j}\right] \tag{3.1}$$

is the information matrix for f . Furthermore, a lower bound to the asymptotic covariance matrix of any consistent estimator of θ is given by R^{-1} where

$$R = \lim_{n\to\infty} \frac{1}{n} R_n \tag{3.2}$$

is the asymptotic information matrix. Under independent sampling and certain regularity conditions the maximum-likelihood and minimum-chi-square estimators are asymptotically efficient; that is, their asymptotic covariance matrices are equal to the lower bound R^{-1} .

The information matrix for the reduced-form parameters can be derived from (2.6). Partitioning according to π and ω , we get the square matrix of order $GK + G^2$

(3.3)
$$R_n = -E\begin{bmatrix} \frac{\partial^2 \log f}{\partial\pi\, \partial\pi'} & \frac{\partial^2 \log f}{\partial\pi\, \partial\omega'} \\ \frac{\partial^2 \log f}{\partial\omega\, \partial\pi'} & \frac{\partial^2 \log f}{\partial\omega\, \partial\omega'} \end{bmatrix} = \begin{bmatrix} R_{11}^n & R_{12}^n \\ R_{21}^n & R_{22}^n \end{bmatrix}$$
$$= \begin{bmatrix} \Omega^{-1} \otimes \mathcal{M}_n & 0 \\ 0 & \frac{1}{2}n(\Omega^{-1} \otimes \Omega^{-1}) \end{bmatrix}$$

where $\otimes$ represents the Kronecker product and where

(3.4) $$\mathcal{M}_n = E[X'X] .$$

The asymptotic information matrix is given by

(3.5) $$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} \Omega^{-1} \otimes \mathcal{M} & 0 \\ 0 & \frac{1}{2}(\Omega^{-1} \otimes \Omega^{-1}) \end{bmatrix}$$

where

(3.6) $$\mathcal{M} = \lim \frac{1}{n}\mathcal{M}_n = \lim \frac{1}{n} E[X'X] .$$

We shall assume that the stochastic process which generates X is sufficiently regular to insure that $\mathcal{M}$ is finite, positive definite, and equal to Plim $X'X/n$.[1]

Unless X is nonstochastic, the finite-sample bound R_n^{-1} is not attainable. However, the asymptotic bound R^{-1} is attained by the covariance matrix of the least-squares estimators

$$\hat{\Pi} = P = Y'X(X'X)^{-1}$$

$$\hat{\Omega} = S = \frac{1}{n} Y'[I - X(X'X)^{-1}X']Y \ .$$

Thus the least-squares estimators P and S (which are also the unconstrained maximum likelihood estimators) are asymptotically efficient with asymptotic covariance matrix given by

$$(3.7) \qquad R^{-1} = \begin{bmatrix} \Omega \otimes \mathcal{M}^{-1} & 0 \\ 0 & 2(\Omega \otimes \Omega) \end{bmatrix} .$$

3.2. Structural Restrictions Utilized

If the knowledge that some structural parameters are known *a priori* is utilized in estimating θ , the asymptotic covariance matrix of the optimal estimator is less than R^{-1} . Furthermore, if there are enough restrictions put on the structure, the structural parameters α can also be estimated. These results concerning simultaneous equation

[1] See, for example, [11, pp. 133-136]. This assumption implies that, if X contains lagged endogenous variables, the difference equation system (2.2) is stable.

systems follow from the general theory of constrained estimation presented in [14]. The relevant results of that paper may be summarized as follows:

THEOREM: Let $f(x, \theta)$ be the likelihood function for a parameter vector θ and let R be its nonsingular asymptotic information matrix. Suppose it is known that θ may be written in the form

$$\theta = h(\alpha)$$

where α is a vector of unknown parameters. Let H, the matrix of partial derivatives of the transformation h evaluated at the true value α^o, have full column rank. Then, under certain regularity conditions, the following hold:

a) The constraint parameter α is locally identified[1] and can be consistently estimated. A lower bound for the asymptotic covariance matrix of any consistent estimator of α is given by

$$M = (H'RH)^{-1} .$$

b) The parameter space for θ is restricted if the rank of H is less than the dimension of θ. A lower bound for the asymptotic covariance matrix of any consistent estimator of θ is given by

$$N = H(H'RH)^{-1}H' .$$

---

[1] Let α^o and θ^o be the true parameter vectors. The structure α is said to be uniquely identified if α^o is the only solution to $\theta^o = h(\alpha)$. The structure α is said to be locally identified if there is some open neighborhood of α^o for which α^o is the only solution. Cf. Rothenberg [14] and Fisher [6, Section 5.8].

c) The bounds M and N are attained by the asymptotic covariance matrices of the constrained maximum-likelihood and minimum-chi-square estimators.

d) The gain in reduced-form efficiency due to imposing the constraints is given by the positive semidefinite matrix $R^{-1} - N$. This matrix is nonzero as long as the rank of H is less than the dimensionality of θ .

Our analysis of the simultaneous equations model will involve the evaluation of M and N for the process having likelihood function (2.4) and the constraints implied by (2.3). The asymptotic information matrix R has already been given in (3.5). The remaining task is to derive the partial derivative matrix for the transformation (2.3) which relates the reduced-form parameters θ to the structural parameters α . Partitioning according to (2.8), we define the $(GK + G^2) \times (r + r^*)$ matrix

$$(3.7) \qquad H = \begin{bmatrix} \frac{\partial\pi}{\partial\delta'} & \frac{\partial\pi}{\partial\sigma'} \\ \frac{\partial\omega}{\partial\delta'} & \frac{\partial\omega}{\partial\sigma'} \end{bmatrix} \equiv \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

where H is evaluated at the true parameter vector α^o .

Unless H has full column rank $(r + r^*)$, the structural parameters will be unidentified.[1] We shall consider here the case of an

---

1 See, for example, [14, Section 6.5]. Even if H does not have full column rank, θ will be restricted if H has less than full row rank. We shall return to this point in Section 5.

identified structure and shall therefore assume that H has rank $r + r^*$. An implication of this assumption is that the number of unknown elements of (B, Γ, Σ) must be no greater than $GK + G^2$. Since there are a total of $(2G^2 + GK)$ structural parameters, the number of zero-order restrictions must be at least as large as G^2. For the moment we shall also assume that H_{11} has full column rank r. The meaning of this assumption will become clear as we proceed.

If M is partitioned according to (δ, σ)

$$(3.8) \qquad M = (H'RH)^{-1} \equiv \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

then N can be written in partitioned form as

$$(3.9) \qquad N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} = \begin{bmatrix} H_{11} & 0 \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} H'_{11} & H'_{21} \\ 0 & H'_{22} \end{bmatrix}$$

since from (2.3) it is clear that H_{12} is a matrix of zeros. Thus a lower bound for the asymptotic covariance matrix of a consistent estimator of π is given by the $GK \times GK$ matrix

$$(3.10) \qquad N_{11} = H_{11} M_{11} H'_{11} \quad .$$

The matrix M_{11} is the lower bound for the asymptotic covariance matrix of a consistent estimator of the structural regression coefficients δ. Using the fact that R_{12}, R_{21}, and H_{12} are zero, we can write

(3.8) as

$$M = \left(\begin{bmatrix} H'_{11} & H'_{21} \\ 0 & H'_{22} \end{bmatrix} \begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix} \begin{bmatrix} H_{11} & 0 \\ H_{21} & H_{22} \end{bmatrix} \right)^{-1}$$

(3.11)
$$= \begin{bmatrix} H'_{11}R_{11}H_{11} + H'_{21}R_{22}H_{21} & H'_{21}R_{22}H_{22} \\ H'_{22}R_{22}H_{21} & H'_{22}R_{22}H_{22} \end{bmatrix}^{-1} .$$

The matrix M_{11} is then given by

(3.12)
$$M_{11} = [H'_{11}R_{11}H_{11} + H'_{21}R_{22}H_{21} - H'_{21}R_{22}H_{22}(H'_{22}R_{22}H_{22})^{-1}H'_{22}R_{22}H_{21}]^{-1}$$

$$= [M^{(1)} + M^{(2)} + M^{(3)}]^{-1}$$

as long as H'_{22} has full row rank.[1] It is shown in Appendix A that this is necessarily the case.

In order to derive an explicit expression for M_{11} and N_{11} we must calculate the matrix H. Before turning to this task, however, let us see what can be learned from the form of (3.12). It is useful to think of M_{11}^{-1} as the sum of two terms: $M^{(1)}$ and $M^{(2)} + M^{(3)}$. It can be verified that both of these terms are positive semidefinite. If H_{11} has full column rank r, then $M^{(1)} = H'_{11}R_{11}H_{11}$ is strictly positive definite. The sum $M^{(2)} + M^{(3)}$, however, may be singular and in an important special case is in fact zero.

---

[1] See Goldberger [7, p. 27] for the calculation of partitioned inverses.

Suppose that there is no a priori information on the elements of Σ except that the matrix is positive definite. In that case the equation

$$\Omega = B^{-1}\Sigma B'^{-1}$$

is a one-to-one transformation from the G^2-dimensional parameter space of Σ to the G^2-dimensional parameter space of Ω. The matrix H_{22} will be nonsingular and given by the expression[1]

$$(3.13) \qquad H_{22} = B^{-1} \otimes B^{-1} .$$

But if H_{22} is nonsingular it is clear that

$$M^{(2)} + M^{(3)} = H_{21}'R_{22}H_{21} - H_{21}'R_{22}H_{22}H_{22}^{-1}R_{22}^{-1}H_{22}'^{-1}H_{22}'R_{22}H_{21}$$

$$= H_{21}'R_{22}H_{21} - H_{21}'R_{22}H_{21}$$

is zero. Thus the expression (3.12) for M_{11} greatly simplifies. Furthermore, if H_{22} is nonsingular, H can have full column rank if and only if H_{11} has full column rank.

The above discussion may be summarized as follows: If Σ is unrestricted, H_{22} is a square nonsingular matrix. The matrix H will have full column rank $r + r^*$ (and the structure identified) only if H_{11} has full column rank r. In this case, a lower bound for the asymptotic covariance matrix of a consistent estimator of δ is given by

$$(3.14) \qquad M_{11} = (H_{11}'R_{11}H_{11})^{-1}$$

---

[1] The derivation appears in Appendix A .

and a lower bound for the asymptotic covariance matrix of a consistent estimator of π is given by

$$(3.15) \qquad N_{11} = H_{11}(H_{11}' R_{11} H_{11})^{-1} H_{11}' .$$

If Σ is restricted, M_{11} is given by (3.12) instead of (3.14). Furthermore, it is possible for H to have full column rank even if H_{11} does not. That is, Σ - restrictions can help identify an otherwise unidentified structure.

The case where there are no restrictions on the structural covariance matrix Σ is perhaps the most important one in econometric practice. Whereas economic theory often suggests that certain regression coefficients are zero, rarely does the economist possess reliable prior information on the error variances and covariances. Nevertheless, it is possible that in some instances it is known that certain covariance terms are zero. For example, it might be argued that two structural equations describing two separated sectors have independently distributed error terms. In such cases it is important to realize that the asymptotic efficiency of the estimates of δ and π can be increased by making use of the Σ-restrictions. Suppose that H_{11} has full column rank so that the structure is identified on the basis of restrictions on B and Γ alone. Then, by (3.12) M_{11} is the inverse of a matrix which is the sum of a positive definite matrix and a positive semi-definite matrix. The inverse of such a sum is less positive definite than the inverse of only the first matrix as long as the second matrix is nonzero.[1] Hence an important question concerning the value of restrictions on Σ is whether $M^{(2)} + M^{(3)}$ is zero. If not, we can conclude that a priori restrictions on the matrix Σ reduces the asymptotic covariance matrices M_{11} and N_{11}.[2]

For the next two sections we shall concentrate on the case where there are no Σ-restrictions. The matrices M_{11} and N_{11} are formed using (3.14) and (3.15) after an explicit expression for H_{11} is found. In Section 6 we return to the case of Σ-restrictions (which require the evaluation of H_{21} and H_{22}) . There it will be proven that a priori information on the structural covariance matrix does indeed improve the asymptotic efficiency in estimating the regression coefficients.

4. THE DERIVATION OF H_{11}

The matrix H depends on the exact nature of the a priori restrictions placed on the structure. We begin by describing the restrictions on the regression coefficient matrices B and Γ. Consider the i^{th} structural equation. Suppose that β_i, the vector of unknown endogeneous parameters, consists of g_i elements and that γ_i, the vector of unknown exogenous parameters, consists of k_i elements so that δ_i consists of $r_i = g_i + k_i$ unknown parameters. It will be convenient to define the following matrices:

---

1 For a proof of this proposition see [14, Appendix A].

2 All of the results in this paper concern asymptotic covariance matrices and asymptotic efficiency. When there is no possibility for confusion, however, we shall occasionally drop the adjective "asymptotic" to simplify an already complex terminology.

C_i = the $g_i \times G$ matrix obtained by striking from a $G \times G$ identity matrix the rows corresponding to the known endogenous parameters of the i^{th} equation. (E.g., the p^{th} row of I_G is removed if β_{ip} is known _a priori_.)

D_i = The $k_i \times K$ matrix obtained by striking from a $K \times K$ identity matrix the rows corresponding to the known exogenous parameters of the i^{th} equation. (E.g., the p^{th} row of I_K is removed if γ_{ip} is known _a priori_.)

$\Pi_i = C_i\Pi$ = the matrix of reduced form regression coefficients corresponding to the unknown endogenous parameters of the i^{th} equation.

Finally, it is useful to define the $r_i \times K$ matrix

$$(4.1) \qquad W_i = \begin{bmatrix} C_i\Pi \\ D_i \end{bmatrix} = \begin{bmatrix} \Pi_i \\ D_i \end{bmatrix}$$

which summarizes all the prior information on the regression coefficients of the i^{th} structural equation.

The matrix H_{11} is obtained by differentiating the GK equations

$$(4.2) \qquad \pi_{rs} = - \sum_i \beta^{ri}\gamma_{is}$$

with respect to the elements of δ. Upon calculation one finds[1]

1 Cf. Goldberger [7, pp. 370-1].

$$(4.3) \qquad \frac{\partial \pi_{rs}}{\partial \beta_{pq}} = - \sum_i \frac{\partial \beta^{ri}}{\partial \beta_{pq}} \gamma_{is} = \sum_i \beta^{rp} \beta^{qi} \gamma_{is} = - \beta^{rp} \pi_{qs}$$

for all β_{pq} not known a priori; and

$$(4.4) \qquad \frac{\partial \pi_{rs}}{\partial \gamma_{pq}} = \begin{cases} - \beta^{rp} & \text{if } s = q \\ 0 & \text{if } s \neq q \end{cases}$$

for all γ_{pq} not known a priori. Thus

$$(4.5) \qquad \frac{\partial \pi_r}{\partial \beta_p'} = - \beta^{rp} \Pi' C_p' = - \beta^{rp} \Pi_p'$$

and

$$(4.6) \qquad \frac{\partial \pi_r}{\partial \gamma_p'} = - \beta^{rp} D_p' \quad .$$

The rp block of the partitioned matrix H_{11} is given by

$$(4.7) \qquad \frac{\partial \pi_r}{\partial \delta_p'} = - \beta^{rp} W_p' \; .$$

The complete matrix H_{11} is given by

$$H_{11} = \frac{\partial\pi}{\partial\delta'} = \begin{bmatrix} \frac{\partial\pi_1}{\partial\delta_1'} & \cdots & \frac{\partial\delta_1}{\partial\delta_G'} \\ \vdots & & \vdots \\ \frac{\partial\pi_G}{\partial\delta_1'} & \cdots & \frac{\partial\pi_G}{\partial\delta_G'} \end{bmatrix} = - \begin{bmatrix} \beta^{11}W_1' & \cdots & \beta^{1G}W_G' \\ \vdots & & \vdots \\ \beta^{G1}W_1' & \cdots & \beta^{GG}W_C' \end{bmatrix}$$

(4.8)

$$= - \begin{bmatrix} \beta^{11}I & \cdots & \beta^{1G}I \\ \vdots & & \vdots \\ \beta^{G1}I & \cdots & \beta^{GG}I \end{bmatrix} \begin{bmatrix} W_1' & 0 & \cdots & 0 \\ 0 & W_2' & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & W_G' \end{bmatrix}$$

$$= - (B^{-1} \otimes I)W'$$

where W is a block diagonal matrix with the W_i as diagonal blocks. The partial derivative matrix H_{11} is thus a $KG \times r$ matrix where

(4.9) $$r = \sum_{i=1}^{G} r_i$$

is the number of structural regression coefficients to be estimated.

If there are no other restrictions on the structural parameters, the possibility of structural estimation and the existence of restrictions on π depend on ρ, the rank of H_{11}. From (4.8) it is clear that the rank of H_{11} is equal to the rank of W since $(B^{-1} \otimes I)$ is nonsingular. Furthermore, because of the block diagonal form of W, the rank of W is the sum of the ranks of the W_i.

If each W_i has full row rank, then so will W and the structural parameters will be locally identified. In fact, since the constraint equations may be rewritten in the linear form

$$B\Pi^{o} = -\Gamma \quad (4.10)$$

local identification implies unique identification. For suppose that there exist two pairs (B_o, Γ_o) and (B_1, Γ_1) which both satisfy (4.10) and satisfy the a priori restrictions. Then for any real number λ, $\lambda B_o + (1 - \lambda)B_1$ and $\lambda\Gamma_o + (1 - \lambda)\Gamma_1$ will also satisfy (4.10) and the a priori restrictions. But this implies an infinite number of solutions in any neighborhood of δ_o and hence δ_o cannot be locally identified.

The block diagonal form for W also indicates that any δ_i is identified if W_i has full row rank. However, if any W_i has less than full row rank, then in general none of the coefficients in the i^{th} equation are identified. Hence, under the zero-order constraints considered here, each structural equation can be studied separately as far as identification is concerned. The crucial factor is the rank of the $r_i \times K$ matrix.[1]

$$W_{i\cdot} = \begin{bmatrix} \Pi_i \\ D_i \end{bmatrix} .$$

Since the rank of a matrix cannot exceed its smallest dimension, it is clear that the rank of W_i is no larger than K. Hence, a necessary condition for identifiability is that the number of parameters to be estimated for the i^{th} equation be less than or equal to K, the total number of exogenous variables in the system. Furthermore, since π_i is

---

[1] This rank condition for identifiability is precisely the one given by Koopmans and Hood [10, pp. 137-138].

homogeneous of degree zero in all of the elements of the i^{th} row of B and Γ, at least one of the zero-order restrictions must be nonhomogeneous. (That is, at least one of the restrictions $\beta_{ij} = k$, $\gamma_{ij} = k$ must have nonzero k). Otherwise π_i will also be homogeneous of degree zero in δ_i and, by Euler's theorem, W_i will necessarily have rank less than r_i. (For convenience we shall assume that the nonhomogeneous restriction is on an element of β_i.)[1]

We can summarize the results of this section as follows: The matrix H_{11} is given by (4.8) as the product of a nonsingular matrix and a block diagonal matrix W. If there are no other types of a priori restrictions, the i^{th} structural equation is identified if and only if W_i has full row rank. A necessary condition is that (1) r_i is no greater than K and (2) at least one of the a priori constraints involving the i^{th} structural equation is nonhomogeneous.

5. THE VARIANCE BOUND

We now may evaluate (3.14) and (3.15) to obtain the lower bounds on the covariance matrices of efficient estimates of δ and π. From (3.14), (4.8) and (3.5) we can write

$$
\begin{aligned}
M_{11} &= (H'_{11}R_{11}H_{11})^{-1} \\
&= [W(B^{-1} \otimes I)'(\Omega^{-1} \otimes \mathfrak{M})(B^{-1} \otimes I)W']^{-1} \qquad (5.1) \\
&= [W(\Sigma^{-1} \otimes \mathfrak{M})W']^{-1}
\end{aligned}
$$

1 This is the conventional normalization rule. Since each equation must have at least one nonzero β in order to be stochastic, normalization on β is always possible.

where use is made of the fact that $\Sigma = B\Omega B'$. Hence,

$$
\begin{aligned}
(5.2) \qquad N_{11} &= H_{11}(H_{11}'R_{11}H_{11})^{-1}H_{11}' \\
&= (B^{-1} \otimes I)W'[W(\Sigma^{-1} \otimes \mathcal{M})W']^{-1}W(B^{-1} \otimes I)' .
\end{aligned}
$$

The matrix M_{11} may be rewritten in a somewhat more meaningful way. Using (3.6) we can write

$$
(5.3) \qquad W(\Sigma^{-1} \otimes \mathcal{M})W' = \lim \frac{1}{n}E \begin{bmatrix} W_1X'XW_1'\sigma^{11} & \dots & W_1X'XW_G'\sigma^{1G} \\ \vdots & & \\ W_GX'XW_1'\sigma^{G1} & \dots & W_GX'XW_G'\sigma^{GG} \end{bmatrix} .
$$

But, by the definition of W_i ,

$$
\begin{aligned}
(5.4) \qquad XW_i' &= X[\Pi_i' \quad D_i'] \\
&= [Y_i - V_i \quad X_i] \equiv \overline{Z}_i ,
\end{aligned}
$$

where Y_i is the matrix of current endogeneous variables that appear in the i^{th} equation with unknown coefficients, V_i is the corresponding matrix of reduced form errors, and X_i is the matrix of exogenous variables that appear in the i^{th} equation with unknown coefficients. Hence $\overline{Z}_i$ is the "purified" matrix of variables which appear in the i^{th} structural equation with unknown coefficients. Then

$$
(5.5) \qquad M_{11} = \text{Plim } n \begin{bmatrix} \overline{Z}_1'\overline{Z}_1\sigma^{11} & \dots & \overline{Z}_1'\overline{Z}_G\sigma^{1G} \\ \vdots & & \vdots \\ \overline{Z}_G'\overline{Z}_1\sigma^{G1} & \dots & \overline{Z}_G'\overline{Z}_G\sigma^{GG} \end{bmatrix}^{-1} ,
$$

a form equivalent to the covariance matrix given by Zellner and Theil [15, p. 58] for the three-stage least-squares estimator.

Equations (5.1) and (5.2) express compactly the covariance matrices of asymptotically efficient estimators of the structural and reduced form regression coefficients. These expressions are derived under the assumption that (1) every structural equation is identified, and (2) there is no a priori information on Σ . It is of some interest to consider the question of relaxing these assumptions.

The problem of underidentification can easily be handled. Suppose that the first structural equation is underidentified; that is, the rank of W_1 is ρ_1 which is less than r_1 . Then δ_1 cannot be estimated consistently. Let W_1^* be the matrix consisting of the ρ_1 independent rows of W_1 . Then, if W_1^* replaces W_1 in W , equation (5.2) is still valid for the lower bound on the covariance matrix of an estimator of π . Furthermore, if the first ρ_1 rows and columns of M_{11} are ignored and W_1^* replaces W_1 , (5.1) remains valid for the lower bound on the covariance matrix for an estimator of $(\delta_2, \ldots, \delta_G)$. If W_1^* should be nonsingular (i.e., if $\rho_1 = K$) , then the $K \times K$ identity matrix will serve for W_1^* since both will span the same space. These observations follow from the discussion in Section 6.5 of [14].

The assumption that there is no a priori information on Σ is quite crucial to the derivation of (5.1) and (5.2). As was pointed out in Section 3, in the presence of Σ-restrictions the expressions for M_{11} and N_{11} become much more complicated. We turn to this case next.

6. COVARIANCE RESTRICTIONS

6.1 Introduction

Although a priori information concerning elements of the structural covariance matrix is probably rare in practice, it is still of some interest to examine the effects of such information on estimation efficiency. If it turns out that Σ-restrictions are very valuable in increasing the efficiency of estimating δ and π, then it would seem that more attention ought to be placed on learning about the variances and covariances of the structural disturbances. In any case, from a purely logical point of view, it is quite asymmetric to limit oneself to coefficient restrictions in a theoretical study of the simultaneous equations problem.

There are, of course, many ways to express a priori information about Σ. The most natural extension of the analysis in the previous sections is to consider the zero-order restrictions of knowing the numerical values of some of the σ_{ij}. The constrained Cramer-Rao bound for δ can be calculated from (3.12) after expressions for H_{21} and H_{22} are found. These expressions are relatively simple although the derivations are rather tedious and are given in Appendix A. Although there is no difficulty in evaluating H for the general case of zero-order restrictions, the resulting expressions are not very illuminating. For any given set of restrictions, equation (3.12) can be evaluated numerically; but general algebraic expressions for the Cramer-Rao bound are not interpretable. However, for two special cases of zero-order Σ-restrictions, an algebraic analysis is quite useful. The first case, which we will discuss in

Section 6.2, assumes that the statistician knows every element of Σ . The second case, which we will discuss in Section 6.3, assumes that the matrix Σ is known to be diagonal.

6.2. Σ Completely Known

The assumption that Σ is known a priori is, on the face of it, very implausible. It is difficult to imagine many real-world problems where the econometrician knows the true value of Σ but not the true values of B and Γ . Consider, however, the following case: A structure has been estimated in the past from a large sample and the very precise estimates of B, Γ, and Σ have been obtained. Because of certain technological changes, however, some elements of B and Γ have shifted. It is now desired to reestimate the model on a new (small) sample. Those elements of B, Γ, and Σ which have not shifted (and these might include every element of Σ) may be assumed to be known. The problem then is to efficiently estimate the remaining parameters under the assumption that Σ and certain other parameters are known.

Another justification for studying the case of a known Σ is to be able to compare the results with other problems involving covariances as nuisance parameters. It is well known that in the "traditional" normal linear regression model, prior information on the covariance matrix does not increase the efficiency of estimating the regression coefficients. This is a result of the block diagonal form of the information matrix. For example, in the unconstrained reduced form (3.3), the least-squares estimator is best regardless of whether Ω is known or unknown. This is not the case with simultaneous equations. Prior knowledge of Σ does improve the efficiency of estimating B, Γ, and Π. The best estimator of Π when Σ is unknown is not the best estimator when Σ is known. Thus one purpose of the present discussion is to shed light on this difference between the traditional regression model and the "simultaneous equations" case.

Our task is to find expressions for the Cramer-Rao bounds M_{11} and N_{11} when, in addition to the restrictions imposed on B and Γ, there is also the restriction that Σ is known. The Jacobian matrix for the transformation (2.3) is now

$$H = \begin{bmatrix} H_{11} \\ H_{21} \end{bmatrix} = \begin{bmatrix} \frac{\partial \pi}{\partial \delta'} \\ \frac{\partial \omega}{\partial \delta'} \end{bmatrix} \tag{6.1}$$

since Σ is known. If H has full column rank, the asymptotic variance bound for δ is now

$$\begin{aligned} M_{11} &= (H'RH)^{-1} = (H_{11}'R_{11}H_{11} + H_{21}'R_{22}H_{21})^{-1} \\ &= [M^{(1)} + M^{(2)}]^{-1} . \end{aligned} \tag{6.2}$$

The matrix $M^{(1)} = H_{11}'R_{11}H_{11}$ has been evalueated in Section 5 and is given by the inverse of (5.5). The matrix $M^{(2)} = H_{21}'R_{22}H_{21}$ is derived in Appendix A and takes the following form: $M^{(2)}$ is an $r \times r$ matrix consisting of zeros except for those elements corresponding to a pair of endogenous parameters. If $m_{ij}^{(2)}$ is the element of $M^{(2)}$ corresponding to the i^{th} and j^{th} elements of δ , $m_{ij}^{(2)}$ will be nonzero only if both δ_i and δ_j are elements of B . If δ_i is β_{pq} and δ_j is β_{rs} , then

$$(6.3) \qquad m_{ij}^{(2)} = \omega_{qs}\sigma^{pr} + \beta^{qr}\beta^{sp} .$$

From (5.5) we can obtain the corresponding element of $M^{(1)} = H_{11}'R_{11}H_{11}$:

$$(6.4) \qquad \begin{aligned} m_{ij}^{(1)} &= \text{Plim}\,\frac{1}{n}\,\bar{y}_q'\bar{y}_s\sigma^{pr} = \text{Plim}\,\frac{1}{n}\,(y_q - v_q)'(y_s - v_s)\,\sigma^{pr} \\ &= \text{Plim}\,\frac{1}{n}\,y_q'y_s\sigma^{pr} - \omega_{qs}\sigma^{pr} . \end{aligned}$$

Hence, the matrix M_{11} is obtained by inverting a matrix which is identical to $M^{(1)}$ except for those elements corresponding to a pair of elements of B . For those elements, $m_{ij}^{(1)}$ is replaced by

$$(6.5) \qquad m_{ij}^{(1)} + m_{ij}^{(2)} = \text{Plim}\,\frac{1}{n}\,y_q'y_s\sigma^{pr} + \beta^{qr}\beta^{sp} .$$

There still remains the question of how important the knowledge of Σ is in increasing estimation efficiency. If the structural parameters are identified by the coefficient restrictions alone, then the parameter δ can be estimated by ignoring the Σ-restrictions. The minimum asymptotic covariance matrix is then given by (5.5). Using the fact that Σ is known reduces this minimum covariance matrix by

$$(6.6) \qquad (H_{11}'R_{11}H_{11})^{-1} - (H_{11}'R_{11}H_{11} + H_{21}'R_{22}H_{21})^{-1} .$$

This matrix is necessarily positive semidefinite and will be nonzero if $M^{(2)} = H_{21}'R_{22}H_{21}$ is not zero. Examining (6.3) we see that for the diagonal element corresponding to $\delta_i = \beta_{pq}$, $m_{ii}^{(2)}$ is given by

$$(6.7) \qquad m_{ii}^{(2)} = \omega_{qq}\sigma^{pp} + \beta^{qp}\beta^{qp} ,$$

a number which is strictly positive. Hence knowledge of Σ necessarily increases the efficiency of the maximim-likelihood estimator of δ as long as there exists at least one unknown element of B . Similarly, the efficiency of the ML estimator of π is also increased since the covariance matrix is now

$$H_{11}[M^{(1)} + M^{(2)}]^{-1}H_{11}' ,$$

a matrix less positive definite than $H_{11}(H_{11}'R_{11}H_{11})^{-1}H_{11}'$.

If the coefficient restrictions taken by themselves are not sufficient to identify the structural parameters, then the knowledge of Σ may enable one to estimate parameters which otherwise would not be estimable. In this case, however, there is the possibility that the efficiency in estimating π is unaffected. If the structure is identified without the Σ-constraints then the addition of these constraints necessarily increases the efficiency of estimating both the structural and reduced form regression coefficients.

6.3. Σ Diagonal

A more realistic form of a priori information is the knowledge that Σ is a diagonal matrix. This assumption has been made in a number of econometric models (e.g., Basmann [3] and Klein [8]), although one suspects it is mathematical convenience rather than economic realism which has been the motivation. It is, however, the author's opinion that the zero-covariance assumption can often be justified in practice. We shall examine the effects of making this assumption without further justification.

The task is similar to the one of the previous subsection. We must find expressions for the Cramer-Rao bounds M_{11} and N_{11} when, in addition to the zero-order restrictions imposed on B and Γ, there is also the restriction that Σ is an unknown diagonal matrix. The asymptotic variance bound for δ is of the form (3.12). That is, assuming that H has full column rank,

$$
\begin{aligned}
M_{11} &= [H_{11}'R_{11}H_{11} + H_{21}'R_{22}H_{21} - H_{21}'R_{22}H_{22}(H_{22}'R_{22}H_{22})^{-1}H_{22}'R_{22}H_{21}]^{-1} \\
&= [M^{(1)} + M^{(2)} + M^{(3)}]^{-1} .
\end{aligned} \tag{6.8}
$$

Since $M^{(1)}$ and $M^{(2)}$ have been evaluated already, the remaining task is to evaluate $M^{(3)}$. Again, the algebraic derivation has been placed in Appendix A. The result is that $M^{(3)}$ is an $r \times r$ matrix of zeros except for those elements corresponding to a pair of elements of B. If the i^{th} element of δ is β_{pq} and the j^{th} element is β_{rs}, then

$$
m_{ij}^{(3)} = \begin{cases} 0 & \text{if } p \neq r \\ -2\beta^{qp}\beta^{sp} & \text{if } p = r \end{cases} \tag{6.9}
$$

To form the matrix M_{11} we add the three matrices $M^{(1)} + M^{(2)} + M^{(3)}$ and invert. Since $M^{(2)} + M^{(3)}$ is necessarily positive semidefinite, the *a priori* knowledge that Σ is diagonal increases estimation efficiency as long as $M^{(2)} + M^{(3)}$ is not zero.[1] Examining (6.3) and (6.9) for the diagonal element corresponding to β_{pq} ,

$$\begin{aligned} m_{ii}^{(2)} + m_{ii}^{(3)} &= \omega_{qq}\sigma^{pp} + \beta^{qp}\beta^{qp} - 2\beta^{qp}\beta^{qp} \\ &= \sum_j \beta^{qj}\sigma_{jj}\beta^{qj}\sigma^{pp} - \beta^{qp}\beta^{qp} \\ &= \sum_{j\neq p} \beta^{qj}\sigma_{jj}\beta^{qj} \ . \end{aligned} \tag{6.10}$$

The last expression is zero only if the q^{th} row of B^{-1} is zero except for β^{qp} (which must be nonzero since not all elements of a row of a nonsingular matrix can be zero). But in this case, the qr element of $B^{-1}B$ is

$$\sum_i \beta^{qi}\beta_{ir} = \beta^{qp}\beta_{pr} = 0$$

for all r not equal to q . Hence the expression in (6.10) is zero only if every element in the p^{th} row of B is zero except for the unknown β_{pq} . But this is not possible since by our normalization convention each row of B contains at least one nonzero element known *a priori*.[2] Thus we can conclude that (6.10) is not zero and that knowledge that Σ is diagonal

---

[1] Again we are assuming that $M^{(1)}$ is invertible so that estimation of the structure is possible even without the Σ-restrictions.

[2] Cf. Section 4, last paragraph.

makes M_{11} and N_{11} smaller than they would be if only the regression coefficient restrictions were used. Again, these conclusions are based on the assumption that the structure is identified even without the Σ-restrictions. If this is not the case, the addition of the Σ-restrictions may not affect efficiency at all.

6.4. Σ-Restrictions and Identification

We may now summarize some results on structural identification. When the zero-order restrictions involve only the elements of B and Γ, the identifiability of the elements of δ depends on the matrix H_{11}. When there are Σ-restrictions in addition to the restrictions on B and Γ, the entire H matrix must be examined. Recall that H is the $(GK + G^2) \times (r + r^*)$ matrix

$$H = \begin{bmatrix} H_{11} & 0 \\ & \\ H_{21} & H_{22} \end{bmatrix}.$$

The submatrix H_{11} is given in equation (4.8); typical elements of H_{21} and H_{22} are given in Appendix A. Applying the general theory developed in [14, Section 6.5], we can conclude that the complete set of structural parameters $\alpha = (\delta, \sigma)$ is locally identified if H has full column rank. The question of identification for simultaneous equation systems is thus one concerning the rank of the Jacobian matrix H.

The rank of H_{11} has been analyzed already in Section 4. In Appendix A it is shown that H_{22} always has full column rank under zero-order restrictions (given that B is assumed to be nonsingular). Hence, a sufficient condition for the identifiability of the complete set of structural parameters is that H_{11} have full column rank. However, depending on the number of Σ-restrictions, this condition is not always necessary.

When there are no restrictions on Σ, H_{22} is nonsingular. In that case H_{11} having full column rank is both necessary and sufficient for the local identifiability of α. With Σ-restrictions, however, the condition is no longer necessary. It is possible for H to have full rank $r + r^*$ while H_{11} has rank less than r. For this to be the case, it is clear that H must have no more rows than columns. Hence a necessary condition for identification is that $r + r^* \leq GK + G^2$. Or

$$(GK + G^2 - r) + (G^2 - r^*) \geq G^2 ;$$

the number of coefficient restrictions plus the number of covariance restrictions must be at least as great as the number of structural equations squared.

An analysis of sufficient conditions for identifiability is quite difficult when Σ-restrictions are present. Since identifiability is not the major topic of this study we shall not pursue the matter further. Fisher [6] using a different approach from ours, has studied the topic extensively. With perseverance, it is likely that his results could be reproven on the basis of the matrix H studied here.

7. EFFICIENT ESTIMATORS

7.1 Introduction

Up to now we have described the asymptotic covariance matrix of efficient estimators under various types of *a priori* information, but we have not discussed the problem of finding an asymptotically efficient estimator. It is not our purpose to present new estimators or to develop computational algorithms for old ones. All that will be done in this section is to determine which of the previously proposed estimators of δ and π are asymptotically efficient.

We shall restrict ourselves to zero-order structural parameter restrictions as we have throughout this paper. Returning to the general notation of Section 1, we shall write the likelihood (2.4) as

$$f(\pi, \omega) \tag{7.1}$$

and the constraints as

$$\pi = h_1(\delta) \tag{7.2}$$

$$\omega = h_2(\delta, \sigma) \ . \tag{7.3}$$

As before π and ω are vectors consisting of all the elements of Π and Ω ; δ and σ are vectors consisting of only the unknown elements of (B, Γ) and Σ . We define the unconstrained least-squares estimators

$$P = Y'X(X'X)^{-1} \quad \text{and} \quad S = \frac{1}{n} Y'[I - X(X'X)^{-1}X']Y \tag{7.4}$$

which, in vector form, are p and s :

(7.5) $$p = \text{vec } P \qquad s = \text{vec } S \ .$$

Finally we define $\hat{R}_{11}$ and $\hat{R}_{22}$ to be the asymptotic information matrices R_{11} and R_{22} evaluated at $\pi = p$ and $\omega = s$.

We shall assume that H has full column rank so that the structural parameters (δ, σ) are identified. Since reduced-form estimates can easily be obtained using (7.2) and (7.3), we shall discuss only structural estimation. Specifically, we shall consider the following estimators: (1) full-information maximum likelihood, (2) linearized maximum likelihood, (3) minimum chi square, and (4) three-stage least squares.

7.2. Full-Information Maximum Likelihood

The maximum-likelihood estimator of (δ, σ) is the solution to the extremal problem

(7.6) $$\max_{\delta,\sigma} \ f[h_,(\delta) \ , \ h_2(\delta, \sigma)]$$

which, under our assumptions, is equivalent to

(7.7) $$\max_{\delta,\sigma} \ 2n \log |\det B| - n \log \det \Sigma - \text{tr}[\Sigma^{-1}(BY' + \Gamma X')(YB' + X\Gamma')] \ .$$

If there are no Σ-restrictions, the maximum-likelihood estimator of δ can be expressed as the solution of the "concentrated" extremal problem [1]

(7.8) $$\min_{\delta} \frac{|(BY' + \Gamma X')(YB' + X\Gamma')|}{|B|^2} \ .$$

---

[1] See, for example, Koopmans and Hood [10, pp. 160-161].

The maximum-likelihood estimator of (δ, σ) given by (7.7) is consistent and asymptotically efficient as long as all of the structural restrictions are taken into account. That is, the function (7.7) is to be maximized only with respect to the unknown elements of (B, Γ, Σ). In particular, if Σ is restricted, the solution of (7.8) is not efficient. The maximum-likelihood estimator of (π, ω) can be obtained from the maximum-likelihood estimator of (δ, σ) by using (7.2) and (7.3). These reduced-form estimates are efficient as long as the structural estimates are.[1]

7.3. Linearized Maximum Likelihood

The maximum-likelihood estimators are difficult to compute since the normal equations for (7.7) and (7.8) are nonlinear. Suppose however some consistent, but inefficient, estimator of (δ, σ) is available. An example would be the two-stage least-squares estimator. Then one could linearize the log likelihood function (7.7) around the inefficient estimator and maximize it instead of the true likelihood function. Explicit formulas for the linearized maximum-likelihood estimator are given by Rothenberg and Leenders [13]. They also prove that the linearized estimator is asymptotically efficient.

---

[1] These results on the efficiency of maximum-likelihood estimators follow from the general theory of constrained maximum-likelihood estimation discussed in [1], [2], and [14].

7.4. Minimum Chi Square

The minimum-chi-square estimator of (δ, σ) is the solution of the extremal problem[1]

$$(7.9)\quad \min_{\delta,\sigma} \; [p - h_1(\delta)]'\hat{R}_{11}[p - h_1(\delta)] + [s - h_2(\delta, \sigma)]'\hat{R}_{22}[s - h_2(\delta, \sigma)]$$

which, upon substitution for $\hat{R}$ and h, is equivalent to

$$(7.10)\quad \min_{\delta,\sigma} \; tr[S^{-1}(P + B^{-1}T)X'X(P + B^{-1}T)' + (I - S^{-1}B^{-1}\Sigma B'^{-1})^2] \; .$$

If Σ is unrestricted, the second term can be made zero for any estimate $\hat{B}$ by setting

$$\hat{\Sigma} = \hat{B}S\hat{B}' \quad .$$

Thus, in the case of no Σ-restrictions, the minimum-chi-square estimator for δ is the solution of

$$(7.11)\qquad \min_{\delta} tr[S^{-1}(P + B^{-1}T)X'X(P + B^{-1}T)'] \; .$$

The minimum-chi-square estimator of (δ, σ) given by (7.9) is consistent and asymptotically efficient if all structural restrictions are taken into account. The reduced-form estimates found by using (7.2) and (7.3) are also efficient. These observations follow from the general theory of constrained estimation developed in [14]. The solution to (7.11) is called by Malinvaud [12, pp. 576-577] the minimum distance estimator of δ. It is clear from the preceding discussion that this estimator is efficient only if Σ is unrestricted.[2]

---

[1] The minimum-chi-square approach to constrained estimation is discussed in Section 6.4 of [14].

[2] See Appendix B for a discussion of Malinvaud's analysis.

7.5. Three-Stage Least Squares

The three-stage least-squares (3SLS) estimator of δ may be expressed as the solution to the problem[1]

$$\min_{\delta} \operatorname{tr}[\hat{\Sigma}(BP + \Gamma)X'X(BP + \Gamma)'] \tag{7.12}$$

where $\hat{\Sigma}$ is some consistent estimator of Σ (e.g., the two-stage least-squares estimator). Unlike the maximum-likelihood and minimum-chi-square methods, three-stage least squares is computationally convenient since (7.12) is quadratic in δ. It involves only solving a large system of linear equations.

When Σ is unconstrained, 3SLS is consistent and asymptotically efficient. This follows from the discussion in Section 5 where the asymptotic bound for δ was found to be (5.5), the same expression as derived by Zellner and Theil for the 3SLS estimator. However, when Σ is constrained, 3SLS is no longer efficient since its asymptotic variance remains unchanged although the asymptotic variance bound decreases.

8. A GENERALIZATION

By assuming that all *a priori* information takes the form of zero-order structural restrictions, we have been able to derive explicit expressions for the minimum variance bounds M_{11} and N_{11}. We shall now consider more general structural restrictions which include zero-order

---

1 This interpretation of three-stage least squares is due to Basmann [4]. The original presentation is given by Zellner and Theil [15].

restrictions as a special case. Unfortunately, it will not be possible to derive results as explicit as those in the previous sections.

Let α now be interpreted as the vector of all the elements of (B, Γ, Σ) . Then

(8.1) $$\theta = h(\alpha)$$

is the set of equations (2.3) relating the $G^2 + GK$ reduced-form parameters θ to the $2G^2 + GK$ structural parameters α . The matrix of partial derivatives

(8.2) $$H = \left[\frac{\partial\theta}{\partial\alpha'} \right]$$

cannot possibly have full column rank since it has more columns than rows. In fact, from the results of Section 4 and Appendix A, it follows that H has rank $G^2 + GK$.

Previously we have considered restrictions which set certain elements of α equal to known numbers. A more general assumption is that the restrictions can be represented by a set of equations

(8.3) $$\psi(\alpha) = 0$$

where ψ is a vector of k differentiable functions. It is clear that zero-order restrictions are a special case of (8.3). We shall assume that the k equations (8.3) are independent in a neighborhood of the true parameter α^0 . That is, we assume that ψ , the matrix of partial derivatives of ψ , has full row rank k for all α in some open neighborhood of α^0 .

If there are enough structural restrictions (8.3) then the reduced-form parameter space is restricted. In that case, the minimum variance bound for estimating θ is smaller than the unconstrained bound R^{-1} . Differentiating (8.1) and (8.3) we have in a neighborhood of α^0

$$d\theta = H d\alpha \tag{8.4}$$

and

$$0 = \psi d\alpha \tag{8.5}$$

where H and ψ are evaluated at the true α^{o}. We are interested in whether the vector space of elements $d\theta$ defined by (8.4) and (8.5) has dimension less than $G^2 + GK$, the dimension of the unconstrained reduced-form parameter space.

It is easily verified that the set of vectors $d\alpha$ which satisfy (8.5) is a vector space spanned by the columns of

$$I - \psi'(\psi\psi')^{-1}\psi \ , \tag{8.6}$$

an idempotent matrix of rank $2G^2 + GK - k$. Then it follows from (8.4) that the parameter space of $d\theta$ is spanned by the columns of the $(G^2 + GK) \times (2G^2 + GK)$ matrix

$$A \equiv H[I - \psi'(\psi\psi')^{-1}\psi] \ . \tag{8.7}$$

Since H has rank $G^2 + GK$, the rank of A is necessarily less than $G^2 + GK$ if k is greater than G^2. That is, the reduced form is restricted if there are more than G^2 independent restrictions of the form (8.3). This is not a necessary condition, however. A necessary and sufficient condition for the reduced form being restricted is that A have less than full row rank. The minimum variance bound for estimating θ is given by

$$\bar{A}(\bar{A}'R\bar{A})^{-1}\bar{A}' \tag{8.8}$$

where $\bar{A}$ is a matrix consisting of the independent columns of A. This reduces to R^{-1} if A has full row rank.

Besides increasing the efficiency of reduced-form estimation, the constraints (8.3) may identify the structure and permit structural estimation. By definition, α is identified if and only if there exists a unique α^o satisfying (8.1) and (8.3) when θ equals the true θ^o. Local identification of α requires that the only solution of

$$H d\alpha = 0$$

$$\psi d\alpha = 0$$

be the zero vector.[1] A necessary and sufficient condition for local identification is that the $(G^2 + GK + k) \times (2G^2 + GK)$ matrix

(8.9) $$\begin{bmatrix} H \\ \psi \end{bmatrix}$$

have full column rank. Since the rank of a matrix cannot exceed its smallest dimension, a necessary condition for identification is that k be no less than G^2.

If the matrix (8.9) has full column rank, then the matrix

(8.10) $$F \equiv (H'RH + \psi'\psi)^{-1}$$

---

[1] See, for example, Rothenberg [14, Section 6.5] and Fisher [6, Section 5.9].

exists. In that case it follows from the results given in Section 6.6 of [14] that the variance bound for estimating α is given by

(8.11) $$F - F\psi'(\psi F\psi')^{-1}\psi F .$$

The variance bound (8.8) for estimating θ may then be expressed as

(8.12) $$H[F - F\psi'(\psi F\psi')^{-1}\psi F]H' .$$

If there are no Σ-restrictions (i.e., none of the constraint equations ψ involve Σ), ψ may be partitioned according to (δ, σ) as:

$$\psi = [\psi_1 \ 0]$$

where ψ_1 is a $k \times (G + GK)$ matrix. Again some simplification of (8.9) - (8.12) occur. In particular, identification now depends on the rank of

(8.13) $$\begin{bmatrix} \psi_1 \\ H_{11} \end{bmatrix}$$

and the existence of restrictions on Π depends on the rank of

(8.14) $$H_{11}[I - \psi_1'(\psi_1\psi_1')^{-1}\psi_1] .$$

9. SUMMARY

In this paper the theory of efficient estimation with prior information in the form of constraint parameters has been applied to the Cowles Commission "simultaneous equations" model. An expression for the asymptotic minimum variance bound was found for both the reduced-form regression coefficients Π and the structural coefficients B and Γ. This was done for the case where all a priori restrictions were zero-order and involved either (B, Γ) alone or (B, Γ, Σ) together. It was found that overidentifying restrictions indeed do increase the efficiency of estimating Π and that Σ-restrictions are valuable in this respect. Finally, it has been shown that the three-stage least-squares estimator and Malinvaud's minimum-distance estimator are asymptotically efficient only if there are no restrictions on Σ. The maximum-likelihood estimator and the minimum-chi-square estimator which take into account all the restrictions are both asymptotically efficient.

The algebra produced in this paper yields only very modest qualitative results. The important question is whether the efficiency increases due to a priori restrictions are numerically important. The formulas derived here must be applied to some actual econometric problems in order to answer this question. Some interesting quantitative results already obtained suggest that the magnitude of the efficiency increase may be substantial. These results will be reported in a later paper.

APPENDIX A

We shall derive in this appendix the typical elements of the Jacobian matrices H_{21} and H_{22} and also of the matrices $H_{21}'R_{22}H_{21}$ and $H_{21}'R_{22}H_{22}$. Finally, for the case where Σ is diagonal, we shall derive an expression for the typical element of the matrix $H_{21}'R_{22}H_{22}(H_{22}'R_{22}H_{22})^{-1}H_{22}'R_{22}H_{21}$. The matrices H_{21} and H_{22} are obtained by differentiating the G^2 elements of Ω,

$$\omega_{rs} = \sum_i \sum_j \beta^{ri}\sigma_{ij}\beta^{sj}, \tag{A.1}$$

with respect to δ and σ.

Each element of the $G^2 \times r$ matrix H_{21} corresponds to an element of Ω and an element of δ. It is clear from (A.1) that the derivative of any ω_{rs} with respect to an element of Γ is zero. Hence H_{21} contains a column of zeros for each element of δ that comes from Γ. The elements of H_{21} corresponding to elements of B take the following form. If β_{pq} is not known a priori, then

$$\begin{aligned}\frac{\partial\omega_{rs}}{\partial\beta_{pq}} &= \sum_i \sum_j \left[\frac{\partial\beta^{ri}}{\partial\beta_{pq}}\sigma_{ij}\beta^{sj} + \beta^{ri}\sigma_{ij}\frac{\partial\beta^{sj}}{\partial\beta_{pq}}\right] \\ &= -\sum_i \sum_j \left[\beta^{rp}\beta^{qi}\sigma_{ij}\beta^{sj} + \beta^{ri}\sigma_{ij}\beta^{sp}\beta^{qj}\right] \qquad \text{(A.2)} \\ &= -(\omega_{sq}\beta^{rp} + \omega_{rq}\beta^{sp}) .\end{aligned}$$

The matrix H_{22} is of order $G^2 \times r^*$ (where r^* is the number of unknown elements of Σ) . If σ_{ab} is unknown, then

$$\text{(A.3)} \qquad \frac{\partial\omega_{rs}}{\partial\sigma_{ab}} = \beta^{ra}\beta^{sb} .$$

Hence H_{22} is obtained by striking from $B^{-1} \otimes B^{-1}$ the columns which refer to known elements of Σ . **Since $B^{-1} \otimes B^{-1}$ has full rank** (and therefore G^2 independent columns), H_{22} must necessarily have full column rank.

Each element of the $r \times r$ matrix $H_{21}'R_{22}H_{21}$ corresponds to a pair of structural regression coefficients. The ij element (corresponding to δ_i and δ_j) is zero if either δ_i or δ_j or both are elements of Γ . If $\delta_i = \beta_{pq}$ and $\delta_j = \beta_{p'q'}$, then, using (A.2) and (3.5), we have

$$(H_{21}'R_{22}H_{21})_{ij} =$$

$$= \frac{1}{2} \sum_r \sum_s \sum_{r'} \sum_{s'} (\omega_{sq}\beta^{rp} + \omega_{rq}\beta^{sp})\omega^{rr'}\omega^{ss'}(\omega_{s'q'}\beta^{r'p'} + \omega_{r'q'}\beta^{s'p'})$$

$$= \frac{1}{2}(\omega_{qq'}\sigma^{pp'} + \beta^{qp'}\beta^{q'p} + \beta^{qp'}\beta^{q'p} + \omega_{qq'}\sigma^{pp'})$$

$$= \omega_{qq'}\sigma^{pp'} + \beta^{qp'}\beta^{q'p} ,$$

where use has been made of the relation

$$\text{(A.5)} \qquad \sigma^{rs} = \sum_i \sum_j \beta^{ir}\omega^{ij}\beta^{js} .$$

Each element of the $r^* \times r^*$ matrix $H'_{22}R_{22}H_{22}$ corresponds to a pair of unknown elements of Σ. The ij element, where i represents σ_{ab} and j represents σ_{cd}, is given by

$$(H'_{22}R_{22}H_{22})_{ij} = \frac{1}{2} \sum_r \sum_s \sum_{r'} \sum_{s'} \beta^{ra}\beta^{sb}\omega^{rr'}\omega^{ss'}\beta^{r'c}\beta^{s'd}$$

(A.6)

$$= \frac{1}{2}\sigma^{ac}\sigma^{bd} .$$

Hence, $H'_{22}R_{22}H_{22}$ is obtained by striking from $\frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1}$ the rows and columns corresponding to the known elements of Σ.

Each element of the $r \times r^*$ matrix $H'_{21}R_{22}H_{22}$ corresponds to an element of δ and an unknown element of Σ. The rows corresponding to elements of Γ are zero. The other elements are of the following form (where the i^{th} row represents β_{pq} and the j^{th} column represents σ_{ab})

$$(H'_{21}R_{22}H_{22})_{ij} = -\frac{1}{2} \sum_r \sum_s \sum_{r'} \sum_{s'} (\omega_{sq}\beta^{rp} + \omega_{rq}\beta^{sp})\omega^{rr'}\omega^{ss'}\beta^{r'a}\beta^{s'b}$$

(A.7)

$$= -\frac{1}{2}(\beta^{qb}\sigma^{pa} + \beta^{qa}\sigma^{pb}) .$$

An important matrix for considering Σ-restrictions is

$$M^{(3)} = -H'_{21}R_{22}H_{22}(H'_{22}R_{22}H_{22})^{-1}H'_{22}R_{22}H_{21} .$$

Unfortunately, although we have an expression for the typical element of $H'_{22}R_{22}H_{22}$, we cannot find a simple expression for its inverse except for special cases. If the prior information is of the form that restricts Σ to be diagonal, then $H'_{22}R_{22}H_{22}$ is a $G \times G$ diagonal matrix with the ii element given by

$$(A.8) \qquad (H'_{22}R_{22}H_{22})_{ii} = \frac{1}{2}\sigma^{ii}\sigma^{ii} .$$

The inverse is a G x G diagonal matrix with the ii element equal to $2\sigma_{ii}^2$. Hence, if $\delta_i = \beta_{pq}$ and $\delta_j = \beta_{rs}$, we have

$$m_{ij}^{(3)} = -\frac{1}{2}\sum_i (\beta^{qi}\sigma^{pi} + \beta^{qi}\sigma^{pi})\sigma_{ii}^2(\beta^{si}\sigma^{ri} + \beta^{si}\sigma^{ri})$$

$$(A.9) \qquad = \begin{cases} 0 & \text{if } p \neq r \\ -2\beta^{sp}\beta^{qp} & \text{if } p = r \end{cases} .$$

APPENDIX B

In his recent book [12, Chapter 9 and 19], Professor Malinvaud attacks the problem of estimating simultaneous equation systems by an extension of his "minimum-distance" approach to general regression models. Unfortunately, there seems to be a minor flaw in his argument concerning the asymptotic efficiency of his proposed estimator. In this Appendix we shall briefly review his derivation and indicate where the difficulty lies. We shall convert Malinvaud's notation into the one we have been using.

Consider again the reduced form model (2.2):

$$y_t = \Pi x_t + v_t \quad . \tag{B.1}$$

Since Π is derived from the structure (2.1) we can consider Π to be a function of the structural regression coefficients δ :

$$\Pi = \Pi(\delta) \quad . \tag{B.2}$$

Malinvaud (p. 577) suggests that δ be estimated by δ^{**} , the solution of

$$\min_{\delta} \left[\sum_t (y_t - \Pi x_t)' S^{-1} (y_t - \Pi x_t)\right] \tag{B.3}$$

where S is defined in our equation (7.4). As Malinvaud shows on page 308, this problem is equivalent to the one given in (7.11) of our chapter.

In Theorem 19.2 (p. 577) Malinvaud states that δ^{**} is asymptotically efficient if the errors are normally distributed. His proof is based ultimately on Theorem 9.5 (p. 302) which concerns the general case of nonlinear regression with additive errors. However, for part of the proof of Theorem 9.5 it is implicitly assumed that the true error covariance matrix Ω is <u>not</u> a function of δ.[1] In the simultaneous equations case, Ω is necessarily a function of δ and hence Theorem 9.5 is not directly applicable. Thus, that part of Theorem 19.2 which concerns efficiency is not proved by Malinvaud.

In fact Theorem 19.2 is correct as long as there are no Σ-restrictions. Since that is the case which Malinvaud seems to have in mind,[2] none of his results are affected. Yet Malinvaud's "proof" of the theorem nowhere makes use of the fact that Σ-restrictions are excluded. This, indeed, is the reason for suspecting the derivation in the first place.

A heuristic argument for the applicability of Theorem 9.5 to the simultaneous equations problem is as follows: If Σ is unconstrained, then the coefficient restrictions put no constraints on Ω. Thus the dependence of Ω on δ is not an effective constraint and can be ignored. In that case Theorem 19.2 is indeed a special case of Theorem 9.5. This heuristic argument is of course not the only way of getting around the difficulty. In Section 7 of the present chapter we have proven what is essentially Malinvaud's Theorem 19.2 using a quite different approach.

---

1 Otherwise the argument in the footnote on page 303 is not valid.

2 Cf. the line below equation (4) on page 568 and the line below equation (11) on page 578.

Since Malinvaud's Theorem 19.2 is correct as long as there are no Σ-constraints, the flaw in the proof has no important consequences. However, by ignoring the possibility that Ω may be a function of δ, the results in Chapter 9 are weaker than they need be. Furthermore, the analysis of Σ-restrictions, which can be handled by the minimum distance approach quite easily, is excluded as a result of this lapse.

APPENDIX C

The results in this paper are based on the assumption that Σ , the matrix of structural error covariances, is nonsingular. This means that all G structural equations are stochastic with nondegenerate random error terms. In most economic models, however, there appear linear identities -- nonstochastic equations which contain no unknown parameters. The purpose of this appendix is to show that, after a trivial redefinition of certain matrices, the results of the paper apply to the case where there are linear identities.

One way to handle the case of linear identities is simply to "solve them out." That is, the identities can be used to reduce the number of structural equations and the number of endogenous variables so that the reduced system has a nonsingular covariance matrix of structural errors. If the system is linear and B is nonsingular, this can always be done. The trouble with this method is that the *a priori* constraints will become more complicated. If the original system had only zero-order coefficient restrictions, the process of solving out the identities will introduce restrictions of a higher order. Furthermore, the parameters of the original model are usually more easy to interpret than the parameters of the reduced model. For these reasons, therefore, it is desirable to work with models which contain the identities.

Suppose the system contains G' stochastic equations and $G-G'$ identities. We assume that there are no unknown parameters in the identities. The complete system (2.1) can be partitioned as follows:

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} x = \begin{bmatrix} u_1 \\ 0 \end{bmatrix} \tag{C.1}$$

where y_1 is a vector of G' of the dependent variables, y_2 is a vector of the remaining dependent variables, x is the vector of all K endogenous variables, and u is the G'-dimensional vector of disturbances. The matrices B_{21}, B_{22}, and Γ_2 are known a priori. If the matrix B is nonsingular, then there exists some ordering of the endogenous variables for which B_{22} is nonsingular. We assume y_1 and y_2 are chosen so that B_{22} is nonsingular.

The system (C.1) is equivalent (as far as the stochastic part is concerned) to

$$B_{11}y_1 + B_{12}[-B_{22}^{-1} B_{21}y_1 - B_{22}^{-1} \Gamma_2 x] + \Gamma_1 x = u_1 \tag{C.2}$$

or

$$\bar{B} y_1 + \bar{\Gamma} x = u_1 \tag{C.3}$$

where

$$\begin{aligned} \bar{B} &= B_{11} - B_{12}B_{22}^{-1}B_{21} \\ \bar{\Gamma} &= \Gamma_1 - B_{12}B_{22}^{-1} \Gamma_2 . \end{aligned} \tag{C.4}$$

It is easy to verify that $\bar{B}$ is simply the inverse of B^{11}, the $G' \times G'$ upper left submatrix of B^{-1} :

$$B^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}^{-1} = \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix} . \tag{C.5}$$

The reduced form for (C.3) is

$$y_1 = \Pi_1 x + v_1 \tag{C.6}$$

where

$$\Pi_1 = - \bar{B}^{-1} \bar{\Gamma} = - [B^{11} \Gamma_1 + B^{12} \Gamma_2] \tag{C.7}$$

and

$$v_1 = \bar{B}^{-1} u_1 = B^{11} u_1 \ . \tag{C.8}$$

All of the relevant stochastic information is given by the probability distribution of v_1 since v_1 is uniquely related to the basic error term u_1. The constraint equations relating the reduced form parameters Π_1 and Ω_{11} to the structural parameters B_{11}, B_{12}, Γ_1, and Σ_{11} are given by

$$\begin{aligned} \Pi_1 &= - [B^{11} \Gamma_1 + B^{12} \Gamma_2] \\ \Omega_{11} &= B^{11} \Sigma_{11} B^{11'} \end{aligned} \tag{C.9}$$

But equations (C.9) are simply a subset of the equations (1.3):

$$\begin{aligned} \Pi &= \begin{bmatrix} \Pi_1 \\ \Pi_2 \end{bmatrix} = \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \\ \Omega &= \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix}^{-1} \end{aligned} \tag{C.10}$$

Hence the derivatives of (C.9) can be obtained by taking a subset of the derivative of (C.10). But these latter derivatives have already been calculated in Sections 4 and Appendix A. Since these calculations do not depend on the invertibility of Σ, they are valid in the present context. The only change needed is that ω^{ij} and δ^{ij} should be interpreted as being typical elements of Ω_{11}^{-1} and Σ_{11}^{-1}. Elements of the form β^{ij} and ω_{ij} should be interpreted as typical elements of the <u>full</u> G x G matrice B^{-1} and Ω.

The relevant information matrix for (Π_1, Ω_{11}) is

(C.11)
$$\begin{bmatrix} \Omega_{11}^{-1} \otimes \mathcal{M} & 0 \\ 0 & \frac{1}{2}(\Omega_{11}^{-1} \otimes \Omega_{11}^{-1}) \end{bmatrix} .$$

Equation (4.8) becomes

(C.12)
$$H_{11} = -(B^{11} \otimes I)W'$$

where W is unchanged. Equation (5.1) becomes

$$\begin{aligned} M_{11} &= (H_{11}' R_{11} H_{11})^{-1} \\ &= [W(B^{11} \otimes I)'(\Omega_{11}^{-1} \otimes \mathcal{M})(B^{11} \times I)W']^{-1} . \\ &= [W(\Sigma_{11}^{-1} \otimes \mathcal{M})W']^{-1} \end{aligned}$$

REFERENCES

[1] Aitchison, J. and S. D. Silvey, "Maximum-likelihood estimation of parameters subject to restraints," Annals of Mathematical Statistics, Vol. 29 (1958), pp. 813-828.

[2] ____________________, "Maximum-likelihood estimation procedures and associated tests of Significance," Journal of the Royal Statistical Society, Series B, Vol. 22 (1960), pp. 154-171.

[3] Basmann, R. L., " Remarks concerning the application of exact finite sample distribution functions of GCL estimators in econometric statistical inference," Journal of the American Statistical Association, Vol. 58 (1963), pp. 943-976.

[4] __________, "On the application of the identifiability test statistic and its exact finite sample distribution function in predictive testing of explanatory economic models," mimeographed paper, (1965).

[5] Chipman, J. S., "On least squares with insufficient observations," Journal of the American Statistical Association, Vol. 59 (1964), pp. 1078-1111.

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

[7] Goldberger, Arthur S., Econometric Theory, New York: John Wiley and Sons, 1964.

[8] Klein, L. R., Economic Fluctuations in the United States, 1921-1941, Cowles Commission Monograph II, New York: John Wiley and Sons, 1950.

[9] __________, "The efficiency of estimation in econometric models," in Essays in Economics and Econometrics, Chapel Hill: University of North Carolina, 1960, pp. 216-232.

[10] Koopmans, T. C. and W. C. Hood, "The estimation of simultaneous linear economic relationships," Chapter 6 in Studies in Econometric Method. Cowles Commission Monograph 14, W. C. Hood and T. C. Koopmans, editors, New York: John Wiley and Sons, 1953, pp. 112-199.

[11] Koopmans, T. C., H. Rubin, and R. B. Leipnik, "Measuring the equation systems of dynamic economics," Chapter 2 in Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10, T. C. Koopmans, editor, New York: John Wiley and Sons, 1950, pp. 53-237.

[12] Malinvaud, E., Methodes Statistiques de l'Econometrie, Paris: Dunod, 1964.

[13] Rothenberg, T. J. and C. T. Leenders, "Efficient estimation of simultaneous equation systems," Econometrica, Vol. 32 (1964), pp. 57-76.

[14] Rothenberg, T. J., "Efficient estimation with a priori information: a classical approach," Cowles Foundation Discussion Paper No. 205, 1966.

[15] Zellner, A. and H. Theil, "Three-stage least squares: simultaneous estimation of simultaneous equations," Econometrica, Vol. 30 (1962), pp. 54-78.