ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Identification in Parametric Models
Author(s): Thomas J. Rothenberg
Reviewed work(s):
Source: *Econometrica*, Vol. 39, No. 3 (May, 1971), pp. 577-591
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/1913267
Accessed: 22/12/2011 12:31

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The Econometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Econometrica.*

http://www.jstor.org

Econometrica, Vol. 39, No. 3 (May, 1971)

IDENTIFICATION IN PARAMETRIC MODELS

BY THOMAS J. ROTHENBERG[1]

A theory of identification is developed for a general stochastic model whose probability law is determined by a finite number of parameters. It is shown under weak regularity conditions that local identifiability of the unknown parameter vector is equivalent to nonsingularity of the information matrix. The use of "reduced-form" parameters to establish identifiability is also analyzed. The general results are applied to the familiar problem of determining whether the coefficients of a system of linear simultaneous equations are identifiable.

1. INTRODUCTION

THE IDENTIFICATION PROBLEM concerns the possibility of drawing inferences from observed samples to an underlying theoretical structure. An important part of econometric theory involves the derivation of conditions under which a given structure will be identifiable. The basic results for linear simultaneous equation systems under linear parameter constraints were given by Koopmans and Rubin [**10**] in 1950. Extensions to nonlinear systems and nonlinear constraints were made by Wald [**15**], Fisher [**4**, **5**, **6**], and others. A summary of these results can be found in Fisher's comprehensive study [**7**]. The identification problem has also been thoroughly analyzed in the context of the classical single-equation errors-in-variables model. The basic papers here are by Neyman [**12**] and Reiersøl [**13**].

Most of this previous work on the identification problem has emphasized the special features of the particular model being examined. This has tended to obscure the fact that the problem of structural identification is a very general one. It is not restricted to simultaneous-equation or errors-in-variables models. As Koopmans and Reiersøl [**9**] emphasize, the identification problem is "a general and fundamental problem arising, in many fields of inquiry, as a concomitant of the scientific procedure that postulates the existence of a structure." In their important paper Koopmans and Reiersøl define the basic characteristics of the general identification problem. In the present paper we shall, in the case of a general parametric model, derive some identifiability criteria. These criteria include the standard rank conditions for linear models as special cases.

Our approach is based in part on the information matrix of classical mathematical statistics. Since this matrix is a measure of the amount of information about the unknown parameters available in the sample, it is not surprising that it should be related to identification. For lack of identification is simply the lack of sufficient information to distinguish between alternative structures. The following results make this relationship more precise.[2]

[1] This research was supported in part by a grant from the Ford Foundation administered through the Center for Research in Management Science, University of California, Berkeley. The research was completed while the author was visiting Associate Professor at the Catholic University of Louvain. I am grateful to Theodore Groves for many helpful discussions and to the referees for their valuable comments.

[2] The approach taken here is inspired by the work of Aitchison and Silvey [**1**, **2**] on constrained maximum-likelihood estimation.

2. BASIC CONCEPTS

Let Y be a vector-valued random variable in R^n representing the outcome of some random experiment. The probability distribution function for Y is known to belong to a family $\mathscr{F}$ of distribution functions on R^n. A structure S is a set of hypotheses which implies a unique distribution function $F(S) \in \mathscr{F}$. The set of a priori possible structures to be investigated is called a model and is denoted by $\mathscr{S}$. By definition there is a unique distribution function associated with each structure in $\mathscr{S}$. The identification problem concerns the existence of a unique inverse association. Following Koopmans and Reiersøl [9], we have the following definitions.

DEFINITION 1′: Two structures in $\mathscr{S}$ are said to be *observationally equivalent* if they imply the same probability distribution for the observable random variable Y.

DEFINITION 2′: A structure S in $\mathscr{S}$ is said to be *identifiable* if there is no other structure in $\mathscr{S}$ which is observationally equivalent.

In order to say more about the identification problem, we must be more specific about the sets $\mathscr{S}$ and $\mathscr{F}$. This is usually done by assuming Y is generated by a set of linear equations with an additive latent error term. We shall take another approach and merely assume that the distribution of Y has a parametric representation. Specifically, we assume that every structure S is described by an m-dimensional real vector α and that the model is described by a set $A \subset R^m$. We associate with each α in A a continuous probability density function[3] $f(y, \alpha)$ which, except for the parameter α, is known to the statistician. Thus the problem of distinguishing between structures is reduced to the problem of distinguishing between parameter points. In this framework we have the following definitions.

DEFINITION 1: Two parameter points (structures) α^1 and α^2 are said to be *observationally equivalent* if $f(y, \alpha^1) = f(y, \alpha^2)$ for all y in R^n.

DEFINITION 2: A parameter point α^0 in A is said to be *identifiable* if there is no other α in A which is observationally equivalent.

Since the set of structures is simply a subset of R^m, it is possible to speak of two structures being close to one another. It is natural then to consider the concept of local identification. This occurs when there may be a number of observationally equivalent structures but they are isolated from each other. We state this more formally in Definition 3.

DEFINITION 3: A parameter point α^0 is said to be *locally identifiable* if there exists an open neighborhood of α^0 containing no other α in A which is observationally equivalent.

[3] The continuity assumption is made solely for ease of exposition. By adding the phrase "except for sets of sample points having zero probability" at the appropriate places, all of our results hold for general probability functions. We shall use the words density function and likelihood function interchangeably when referring to f.

To emphasize the distinction between Definition 2 and Definition 3, we shall speak of *global identification* in the former case and *local identification* in the latter. Needless to say, global identification implies local identification.

The identification problem is to find conditions on $f(y, \alpha)$ and A that are necessary and sufficient for the identification of the parameters in A. We shall give some partial answers to this problem in the so-called "regular" parametric case.[4] This case is described by the following assumptions.

ASSUMPTION I: *The structural parameter space A is an open set in R^m.*

ASSUMPTION II: *The function f is a proper density function for every α in A. In particular, f is nonnegative and the equation $\int f(y, \alpha)\, dy = 1$ holds for all α in A.*

ASSUMPTION III: *The set B of y values for which f is strictly positive is the same for all α in A. We shall refer to the set B as the sample space of Y.*

ASSUMPTION IV: *The function f is smooth in α. Specifically, we assume that for all α in a convex set containing A and for all y in the sample space B the functions $f(y, \alpha)$ and $\log f(y, \alpha)$ are continuously differentiable with respect to α.*

ASSUMPTION V: *The elements of the information matrix*

$$R(\alpha) = [r_{ij}(\alpha)] = E\left[\frac{\partial \log f}{\partial \alpha_i} \cdot \frac{\partial \log f}{\partial \alpha_j}\right]$$

exist and are continuous functions of α everywhere in A.

We shall also need the following definition.

DEFINITION 4: Let $M(\alpha)$ be a matrix whose elements are continuous functions of α everywhere in A. The point $\alpha^0 \in A$ is said to be a *regular point* of the matrix if there exists an open neighborhood of α^0 in which $M(\alpha)$ has constant rank.

3. CRITERIA FOR LOCAL IDENTIFICATION

Under the above assumptions and definitions we have the following result.

THEOREM 1: *Let α^0 be a regular point of the matrix $R(\alpha)$. Then α^0 is locally identifiable if and only if $R(\alpha^0)$ is nonsingular.*

PROOF: Using the logarithmic density function $g(y, \alpha) \equiv \log f(y, \alpha)$ and its partial derivatives

$$g_i(y, \alpha) \equiv \frac{\partial \log f(y, \alpha)}{\partial \alpha_i},$$

[4] See, for example, Cramér [3, p. 479].

we have by the mean value theorem, for all y in B and for all α in a neighborhood of α^0,

$$g(y,\alpha) - g(y,\alpha^0) = \sum_{i=1}^{m} g_i(y,\alpha^*)(\alpha_i - \alpha_i^0) \tag{3.1}$$

where α^* lies between α and α^0. Suppose that α^0 is not locally identifiable. Then there exists an infinite sequence of vectors $\{\alpha^1, \alpha^2, \ldots, \alpha^k, \ldots\}$ approaching α^0 such that $g(y,\alpha^k) = g(y,\alpha^0)$ for all y and each k. This implies, for all y, $\Sigma_i g_i(y,\alpha^*)d_i^k = 0$ where

$$d_i^k = \frac{\alpha_i^k - \alpha_i^0}{|\alpha^k - \alpha^0|}.$$

The sequence $\{d^k\}$ is an infinite sequence on the unit sphere and therefore there exists a limit point d. As $\alpha^k \to \alpha^0$, d^k approaches d and in the limit we have

$$\sum_i g_i(y,\alpha^0)d_i = 0 \tag{3.2}$$

for all y. But this implies

$$d'R(\alpha^0)d = E\left[\sum_i g_i(y,\alpha^0)d_i\right]^2 = 0 \tag{3.3}$$

and, hence, $R(\alpha^0)$ must be singular.

To show the converse, suppose that $R(\alpha)$ has constant rank $\rho < m$ in a neighborhood of α^0. Then consider the characteristic vector $c(\alpha)$ associated with one of the zero roots of $R(\alpha)$. Since

$$0 = c'Rc = E\left[\sum_i g_i(y,\alpha)c_i(\alpha)\right]^2 \tag{3.4}$$

we have, for all y in B and all α near α^0,

$$\sum_i g_i(y,\alpha)c_i(\alpha) \equiv 0. \tag{3.5}$$

Since $R(\alpha)$ is continuous and has constant rank, the vector function $c(\alpha)$ is continuous in a neighborhood of α^0. Consider then the curve γ defined by the function $\alpha(t)$ which solves for $0 \leqslant t \leqslant t^*$ the differential equation

$$\frac{\partial \alpha_i(t)}{\partial t} = c_i(\alpha) \qquad (i = 1, \ldots, m),$$

$$\alpha_i(0) = \alpha_i^0.$$

The log density function is differentiable in t with

$$\frac{\partial g[y,\alpha(t)]}{\partial t} = \sum_i g_i[y,\alpha(t)]c_i[\alpha(t)]. \tag{3.6}$$

But by (3.5) this is zero for all y in B and $0 \leqslant t \leqslant t^*$. Thus g is constant on the curve γ and α^0 is unidentifiable. This completes the proof of Theorem 1.

As an example of Theorem 1 consider the nonlinear regression model

$$y_t = h(\alpha, x_t) + \varepsilon_t \qquad (t = 1, \ldots, n),$$

where α is an unknown m-dimensional parameter vector and h is a known function twice differentiable in α. The ε_t are independent normal random variables with zero mean and unit variance. The x_t are nonrandom observed numbers. The logarithmic density function for the sample $y_1, \ldots, y_n$ is

$$\log f = -\tfrac{1}{2}\sum_t [y_t - h(\alpha, x_t)]^2$$

and the information matrix has typical element

$$r_{ij}(\alpha) = \sum_t h_i(\alpha, x_t)h_j(\alpha, x_t)$$

where h_i is the partial derivative of h with respect to α_i. Forming the $m \times n$ matrix $H(\alpha) = [h_{it}] = [h_i(\alpha, x_t)]$, we can write the information matrix as $R(\alpha) = HH'$. Thus a vector α^0 is locally identifiable if $H(\alpha^0)$ has rank m.

In many statistical problems the m-dimensional vector α is restricted to a lower ordered subspace. In these cases the restricted parameter space is not an open set in R^m and Theorem 1 is not applicable. In order to handle this common case we add to Assumptions I–V the following assumption.

ASSUMPTION VI: *The vector α is known to satisfy a set of constraint equations*

$$\psi_i(\alpha) = 0 \qquad (i = 1, \ldots, k), \tag{3.7}$$

where each ψ_i is a known function possessing continuous partial derivatives.

The new parameter space A' is then the intersection of the set A and the solution set of (3.7). We denote the Jacobian matrix for the ψ_i as

$$\Psi(\alpha) = [\psi_{ij}(\alpha)] = \left[\frac{\partial \psi_i}{\partial \alpha_j}\right]$$

and define the $(m + k) \times m$ partitioned matrix

$$V(\alpha) = \begin{bmatrix} R(\alpha) \\ \Psi(\alpha) \end{bmatrix}.$$

Under Assumptions I–VI we have the following theorem.

THEOREM 2: *Suppose $\alpha^0 \in A'$ is a regular point of both $\Psi(\alpha)$ and $V(\alpha)$. Then α^0 is locally identifiable if and only if $V(\alpha^0)$ has rank m.*

PROOF: Suppose $\Psi(\alpha)$ has rank s for all α in a neighborhood of α^0. Then the constrained parameter space is, in a neighborhood of α^0, a manifold of $m - s$ dimensions. Thus A' is locally homeomorphic to an open set of R^{m-s}. By examining the information matrix with respect to a new parameter vector which describes the lower order space, we can then use Theorem 1 to establish local identifiability.

Suppose we partition α into two components, α_1 having s elements and α_2 having $m - s$ elements. By the implicit function theorem, there exists a partition and a differentiable mapping q such that the vector equation

$$\alpha_1 = q(\alpha_2) \tag{3.8}$$

is satisfied for all solutions of (3.7) in a neighborhood of α^0. Then, on the manifold, the likelihood function is

$$f^*(y, \alpha_2) = f[y, q(\alpha_2), \alpha_2]$$

and the information matrix with respect to α_2 is

$$R^*(\alpha) = [Q'I]R\begin{bmatrix} Q \\ I \end{bmatrix} \tag{3.9}$$

where $R = R(\alpha)$ and Q is the Jacobian matrix

$$Q(\alpha_2) = \left[\frac{\partial q_i(\alpha_2)}{\partial \alpha_{2j}}\right].$$

The matrix $R^*(\alpha)$ will be singular if and only if there exists a nonzero vector z such that

$$R\begin{bmatrix} Q \\ I \end{bmatrix} z = 0. \tag{3.10}$$

But from (3.7) and (3.8) one sees that a vector x has the form

$$x = \begin{bmatrix} Q \\ I \end{bmatrix} z$$

if and only if it is a solution of $\Psi x = 0$. That is, z is a solution of (3.10) if and only if there exists a nonzero vector x satisfying

$$\begin{bmatrix} R \\ \Psi \end{bmatrix} x \equiv Vx = 0. \tag{3.11}$$

Thus the rank of $R^*(\alpha)$ is $m - s$ if and only if the rank of $V(\alpha)$ is m. By applying Theorem 1 we get our result.

4. CRITERIA FOR GLOBAL IDENTIFICATION

For most problems we are interested in global identification rather than simply local identification. Unfortunately it is more difficult to obtain global results. Of course local identification is a necessary condition for global identification. Therefore an examination of the information matrix would appear to be a first step in an actual investigation. Nevertheless, one would like to have some general conditions which are sufficient to guarantee global identification. In this section we shall present some results in this direction. For simplicity we shall drop Assumption VI and return to the unconstrained parameter space A.

Going through the proof of Theorem 1, one is tempted to conclude that a sufficient condition for the global identifiability of α^0 is that $R(\alpha)$ be invertible for all α. Consider again equation (3.1) which expresses for arbitrary α^0 and α^1 in A the difference in log densities:

$$g(y, \alpha^1) - g(y, \alpha^0) = \sum_i g_i(y, \alpha^*)(\alpha_i^1 - \alpha_i^0).$$

If α^0 and α^1 were observationally equivalent this difference would be zero for all y. Letting $d_i = \alpha_i^1 - \alpha_i^0$ we would have

$$E\left[\sum_i g_i(y, \alpha^*)d_i\right]^2 = 0. \tag{4.1}$$

If α^* did not depend on y, then (4.1) would equal $d'R(\alpha^*)d$. Thus if $R(\alpha)$ were nonsingular everywhere, then there could exist no observationally equivalent points. But, of course, α^* *will* in general depend on y and the above reasoning is invalid. There are, however, cases where α^* is independent of y and then we can prove a global result. Suppose, for example, the logarithmic likelihood function has the form

$$g(y, \alpha) = A(y) + B(\alpha) + \sum_{i=1}^{m} \alpha_i D_i(y) \tag{4.2}$$

with B differentiable in α. This is a special case of the multivariate exponential family of densities.[5] It is easy to verify that the multivariate normal density with known covariance matrix and unknown mean vector α leads to this form. Using (4.2) we have

$$g(y, \alpha^1) - g(y, \alpha^0) = B(\alpha^1) - B(\alpha^0) + \sum_i D_i(y)(\alpha_i^1 - \alpha_i^0). \tag{4.3}$$

Suppose (4.3) equals zero for all y. Then we can apply the mean value theorem to $B(\alpha)$, obtaining

$$\begin{aligned} 0 &= \sum_i [B_i(\alpha^*) + D_i(y)](\alpha_i^1 - \alpha_i^0) \\ &= \sum_i g_i(y, \alpha^*)d_i \end{aligned}$$

where B_i is the derivative of B with respect to α_i and where α^* is independent of y. Thus we have the next theorem.

THEOREM 3: *Let $f(y, \alpha)$ be a member of the exponential family defined by* (4.2). *If $R(\alpha)$ is nonsingular in a convex set containing A, then every α in A is globally identifiable.*

Outside the exponential family it does not seem possible to prove conditions for global identifiability using only the information matrix. Other approaches

[5] See, for example, Lehmann [**11**, p. 51].

must be used. The following situation appears often in practice. Suppose that the parameters α_i are in fact directly interpretable as characteristics of the density function $f(y, \alpha)$. For example the α_i might be moments of the probability distribution. Then distinct values of α necessarily imply distinct distributions of Y and therefore are identifiable. One simple result of this approach is given in the next theorem.

THEOREM 4: *Suppose there exist m known functions* $\phi_1(Y), \ldots, \phi_m(Y)$ *such that, for all* α *in A,*

$$\alpha_i = E[\phi_i(Y)] \qquad (i = 1, \ldots, m).$$

Then every α *in A is identifiable.*

The proof is trivial. If $f(y, \alpha^1) = f(y, \alpha^0)$ for all y, then for all i

$$\int \phi_i(y) f(y, \alpha^1)\, dy = \int \phi_i(y) f(y, \alpha^0)\, dy$$

and $\alpha^1 = \alpha^0$.

Theorems 3 and 4 cover only a limited number of cases. Yet they are crucial since they are the basis of most results in identification theory. Theorem 4, for example, is the starting point of Reiersøl's proof [**13**] that the structural errors-in-variables model is identifiable if the observations are not normally distributed. Furthermore, the traditional approach used in the econometric literature begins with a proof that the reduced-form parameters are identifiable because they are population moments.

5. STRUCTURE AND REDUCED FORM

In this section we shall analyze an important special case that arises in many actual identification problems. This occurs when there exist reduced-form parameters that can be used to help establish the identification of structural parameters. The standard econometric literature on identification as summarized by Fisher [7] deals exclusively with this case.

Again we are concerned with the identifiability of an unknown m-dimensional structural parameter vector α^0 which lies in the open set $A \subset R^m$. We do not need the regularity Assumptions II–V given in Section 2 but we do reintroduce Assumption VI (which for convenience we rewrite).

ASSUMPTION VI: *The vector* α *is known to satisfy a set of continuously differentiable constraint equations* $\psi_i(\alpha) = 0$ $(i = 1, \ldots, k)$, *with Jacobian matrix* $\Psi(\alpha)$. *The constrained structural parameter space is denoted by* A'.

In addition we add two other assumptions.

ASSUMPTION VII: *The probability density for Y depends on the structural parameter* α *only through an r-dimensional "reduced-form" parameter* θ. *That is, there*

exist r known continuously differentiable functions $\theta_i = h_i(\alpha)$ $(i = 1, \ldots, r)$ *mapping A into* R^r *and a function* $f^*(y, \theta)$ *such that*

$$f(y, \alpha) = f^*[y, h(\alpha)] = f^*(y, \theta)$$

for all y in B and for all α *in A.*

ASSUMPTION VIII: *Let* $A^* \subset R^r$ *be the image of* A' *under the mapping h. Then every* θ *in* A^* *is assumed to be (globally) identifiable.*

Thus the parameters α constitute the basic structures of the model A'. Every structure, however, implies a reduced-form parameter θ which completely characterizes the probability distribution of Y. Furthermore these reduced-form parameters are globally identifiable. Hence the identification of a vector $\alpha^0 \in A'$ depends solely on the properties of the mappings h and ψ. If θ^0 is the image of α^0, then α^0 is (globally) identifiable if and only if the equations

$$\begin{aligned} \theta_i^0 &= h_i(\alpha) && (i = 1, \ldots, r), \\ 0 &= \psi_j(\alpha) && (j = 1, \ldots, k), \end{aligned} \tag{5.1}$$

have the unique solution α^0. The identification problem then becomes simply a question of the uniqueness of solutions to systems of equations. Here we may use the classical results of analysis. Defining the Jacobian matrices

$$H(\alpha) = \left[\frac{\partial h_i}{\partial \alpha_j}\right], \qquad \Psi(\alpha) = \left[\frac{\partial \psi_i}{\partial \alpha_j}\right],$$

and the partitioned matrix of order $(r + k) \times m$

$$W(\alpha) = \begin{bmatrix} H(\alpha) \\ \Psi(\alpha) \end{bmatrix}, \tag{5.2}$$

we have the next theorems.

THEOREM 5: *If the functions* h_i *and* ψ_i *are linear, then* α^0 *is (globally) identifiable if and only if the (constant) matrix W has rank m.*

THEOREM 6: *If* α^0 *is a regular point of* $W(\alpha)$*, then* α^0 *is locally identifiable if and only if* $W(\alpha^0)$ *has rank m.*

The case of global identification when the h_i and ψ_i are nonlinear is more difficult. All we can offer is an overly strong sufficiency condition using a result due to Gale and Nikaido [**8**].[6]

THEOREM 7: *If* A' *is a convex set, then a sufficient condition for the global identification of* α^0 *is that there exists an* $m \times m$ *submatrix* $\overline{W}$ *of W such that the determinant of* $\overline{W}$ *is positive and* $\overline{W} + \overline{W}'$ *is positive semidefinite throughout* A'.

[6] See Fisher [7, p. 159] for a similar result.

Finally, we may state a result due to Wald [**15**] concerning the local identifiability of a given element of α. Consider, say, the first element of α^0. We say that α_1^0 is locally identifiable if there exists an open neighborhood of α^0 in which all vectors observationally equivalent to α^0 have the same value for α_1.[7] Then suppose we add to (5.1) the fact that α_1 is known. We have

$$\begin{aligned} \theta_i^0 &= h_i(\alpha) && (i = 1, \ldots, r), \\ 0 &= \psi_j(\alpha) && (j = 1, \ldots, k), \\ \alpha_1^0 &= \alpha_1, \end{aligned} \tag{5.3}$$

which has Jacobian matrix $W_1(\alpha)$. It is clear that W_1 is simply W with an added row consisting of one in the first column and zeros elsewhere. If α_1^0 is identifiable then there is no new information in (5.3) that is not already in (5.1). Conversely, if there is no new information in (5.3) then α_1^0 must be identifiable. Locally the Jacobian matrix summarizes the information in the equation system. This leads us to another theorem.

THEOREM 8: *If α^0 is a regular point of both $W(\alpha)$ and $W_1(\alpha)$, then α_1^0 is locally identifiable if and only if $W(\alpha)$ and $W_1(\alpha)$ have the same rank.*

The proof can be found in Wald [**15**] and Fisher [**7**, p. 181]. If should be noted that the theorem is globally valid if the h_i and ψ_i are linear.

6. THE SIMULTANEOUS EQUATIONS MODEL

The results of the previous sections can be applied to the econometric model developed by Koopmans, Rubin, and Leipnik in [**10**]. The model contains a set of G linear equations

$$By_t + \Gamma x_t = u_t \tag{6.1}$$

where y_t and u_t are G-dimensional vectors of random variables and x_t is a K-dimensional vector of nonrandom exogenous variables; the $G \times G$ matrix B and the $G \times K$ matrix Γ are parameters. The vector u_t is assumed to be normally distributed with mean zero and covariance matrix Σ. The parameter space A consists of the $(2G + K)G$ elements of (B, Γ, Σ) such that B and Σ are nonsingular. In addition, however, there are the constraints

$$\psi_i(B, \Gamma, \Sigma) = 0 \qquad (i = 1, \ldots, k). \tag{6.2}$$

After premultiplying (6.1) by B^{-1} we obtain

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

[7] Note that this definition does not exclude there being two observationally equivalent points having the first components arbitrarily close.

where v_t has mean vector zero and covariance matrix

$$\Omega = B^{-1}\Sigma B'^{-1}. \tag{6.3}$$

If a random sample of size n is taken from this process, one obtains the $n \times K$ matrix X of observations on x_t and the $n \times G$ matrix Y of observations on y_t. If X has full rank K, it is easy to verify that everywhere in A

$$E[(X'X)^{-1}X'Y] = \Pi'$$

$$E\frac{Y'[I - X(X'X)^{-1}X']Y}{T - K} = \Omega.$$

Thus by Theorem 4, Π and Ω are identifiable. Furthermore, since u_t is normally distributed, the density for Y depends only on Π and Ω. Hence Assumption VII is satisfied. We may thus consider (B, Γ, Σ) to be the structural parameter α and (Π, Ω) to be the reduced form parameter θ, and then use the results of Section 5.[8]

Let $(B^0, \Gamma^0, \Sigma^0)$ be some structure satisfying the constraints (6.2). Let (Π^0, Ω^0) be the implied reduced form. Then the identifiability of the structure depends on the uniqueness of solutions of [9]

$$B\Pi^0 + \Gamma = 0, \tag{6.4a}$$

$$B\Omega^0 B' - \Sigma = 0, \tag{6.4b}$$

$$\psi_i(B, \Gamma, \Sigma) = 0 \qquad (i = 1, \ldots, k). \tag{6.4c}$$

Let $\beta = \operatorname{vec} B$ be the G^2-dimensional vector formed from the elements of B taken a row at a time. Similarly, we define the GK-dimensional vector $\gamma = \operatorname{vec} \Gamma$ and the G^2-dimensional vector $\sigma = \operatorname{vec} \Sigma$. Then the vector α may be written as

$$\alpha = \begin{bmatrix} \beta \\ \gamma \\ \sigma \end{bmatrix} = \begin{bmatrix} \operatorname{vec} B \\ \operatorname{vec} \Gamma \\ \operatorname{vec} \Sigma \end{bmatrix}.$$

We also define the three partial derivative matrices

$$\Psi_\beta = \left[\frac{\partial \psi_i}{\partial \beta_j}\right], \qquad \Psi_\gamma = \left[\frac{\partial \psi_i}{\partial \gamma_j}\right], \qquad \Psi_\sigma = \left[\frac{\partial \psi_i}{\partial \sigma_j}\right].$$

Differentiating, we can write the Jacobian matrix for the functions in (6.4) in the partitioned form

$$W = \begin{bmatrix} (I_G \otimes \Pi^0)' & I_{GK} & 0 \\ \Delta & 0 & -I_{GG} \\ \Psi_\beta & \Psi_\gamma & \Psi_\sigma \end{bmatrix} \tag{6.5}$$

[8] Alternatively, if we do not assume normality but ignore all information not contained in the first two moments of the distribution, then again Π and Ω characterize the process and Section 5 is relevant.

[9] Since Ω^0 is necessarily symmetric, equation (6.4b) imposes the same constraint on Σ. Hence the symmetry condition on Σ need not be included in (6.4c). Of course the ψ functions must be consistent with the symmetry. In the following we assume that the constraints have been written in such a way that $\psi_i(B, \Gamma, \Sigma) = \psi_i(B, \Gamma, \Sigma')$ for all i; e.g., the constraint $\sigma_{12} = 1$ would be written $\sigma_{12} + \sigma_{21} = 2$.

where $\otimes$ represents the Kronecker product and where Δ is a $G^2 \times G^2$ matrix having the property that $\Psi_\sigma \Delta = 2\Psi_\sigma(I \otimes B\Omega^0)$. But, when evaluated at α^0, W may be rewritten as

$$W(\alpha^0) = \begin{bmatrix} 0 & I_{GK} & 0 \\ 0 & 0 & -I_{GG} \\ W^* & \Psi_\gamma & \Psi_\sigma \end{bmatrix} \begin{bmatrix} (I_G \otimes B^{-1})' & 0 & 0 \\ (I_G \otimes \Pi)' & I_{GK} & 0 \\ -\Delta & 0 & I_{GG} \end{bmatrix} \tag{6.6}$$

where W^* is the $k \times G^2$ matrix given by[10]

$$W^* = \Psi_\beta(I_G \otimes B') + \Psi_\gamma(I_G \otimes \Gamma') + \Psi_\sigma(I_G \otimes 2\Sigma) = \Psi \begin{bmatrix} I_G \otimes B' \\ I_G \otimes \Gamma' \\ I_G \otimes 2\Sigma \end{bmatrix}. \tag{6.7}$$

The second partitioned matrix in (6.6) is nonsingular for all α in the unrestricted parameter space A. Hence the rank of W equals $G(G + K)$ plus the rank of W^*. Using Theorem 6 we have the following theorem.

THEOREM 9: *Suppose the $G(2G + K)$-element structural parameter α^0 is a regular point of W^*. Then the parameter is locally identifiable if and only if W^* has rank G^2.*

This result was first found by Wegge [**16**] using a different method of proof. Some implications of Theorem 9 are developed in [**14**] for the special case where all the restrictions are of the exclusion type. The following corollaries are immediate consequences of the arguments leading to the theorem.

COROLLARY 1: *A necessary condition for the identification of α^0 is that there be at least G^2 independent restrictions ψ_i.*

COROLLARY 2: *If each function ψ_i is linear in B and Γ and independent of Σ, then Theorem 9 is also globally true.*

It is also possible to apply Theorem 8 to this model. Let Ψ_i be the $(k + 1) \times (2G^2 + GK)$ matrix obtained by appending to Ψ a row with one in the ith column and zeros elsewhere. We then define

$$W_i^* = \Psi_i \begin{bmatrix} I_G \otimes B' \\ I_G \otimes \Gamma' \\ I_G \otimes 2\Sigma \end{bmatrix}.$$

THEOREM 10: *If α^0 is a regular point of both W^* and W_i^*, its ith component is locally identifiable if and only if W^* and W_i^* have the same rank. If each function ψ_i is linear in B and Γ and independent of Σ, then the result is global.*

[10] All the matrices in (6.7) are evaluated at α^0.

Using Theorem 10 we can now examine the familiar case where each constraint ψ_i is a linear function of the coefficients of the *same* equation. Let δ_p be the vector of coefficients in the pth structural equation. That is δ_p' is the pth row of $D = (B\Gamma)$. Then we assume that the constraints are of the form

$$\Psi_p \delta_p = c_p \qquad (p = 1, \ldots, G), \tag{6.8}$$

where Ψ_p is a $k_p \times (G + K)$ matrix of constants and c_p is a vector constant. There are then $k = \Sigma k_p$ constraints in total.

Substituting into (6.7) we find that W^* is given by the $k \times G^2$ block-diagonal matrix

$$W^* = \begin{bmatrix} W_1^* & & 0 \\ & \ddots & \\ 0 & & W_G^* \end{bmatrix} \tag{6.9}$$

where the pth diagonal block is given by the $k_p \times G$ matrix

$$W_p^* = \Psi_p (B^0 \Gamma^0)' \equiv \Psi_p D^{0\prime}. \tag{6.10}$$

To investigate the identifiability of the ith component of δ_p^0 (i.e., the pi element of D^0), we must examine

$$W_{pi}^* = \Psi_{pi} (B^0 \Gamma^0)' \tag{6.11}$$

where Ψ_{pi} is obtained by appending to Ψ_p a row with one in the ith column and zeros elsewhere. Given the block diagonal form of (6.9), Theorem 10 leads us to the following (apparently new) result.

COROLLARY 3: *Under linear constraints on the coefficients of the same equation, the ith element of δ_p^0 is identifiable if and only if W_p^* and W_{pi}^* have the same rank.*

If W_p^* has full rank G, then adding a new row to Ψ_p cannot change its rank. Hence every element of δ_p is identifiable. Conversely, if W_p^* has rank less than G, at least one element of δ_p must be unidentifiable. Thus we have the classical result given by Koopmans and Rubin [**10**].

COROLLARY 4: *Under linear constraints on the coefficients of the same equation, the vector of coefficients δ_p^0 is identifiable if and only if W_p^* has rank G.*[11]

These last two corollaries may be illustrated by the three-equation model

$$\beta_{11} y_1 + \beta_{12} y_2 + \beta_{13} y_3 + \gamma_{11} x_1 + \gamma_{12} x_2 = u_1,$$
$$\beta_{21} y_1 + \beta_{22} y_2 + \beta_{23} y_3 + \gamma_{21} x_1 + \gamma_{22} x_2 = u_2,$$
$$\beta_{31} y_1 + \beta_{32} y_2 + \beta_{33} y_3 + \gamma_{31} x_1 + \gamma_{32} x_2 = u_3,$$

[11] We incorporate the normalization rule in (6.8). Hence our rank condition is G instead of the usual $G - 1$.

where the constraint functions are

$$\beta_{11} = \beta_{22} = \beta_{33} = 1,$$
$$\gamma_{21} = \gamma_{22} = \gamma_{31} = \gamma_{32} = 0,$$
$$\gamma_{11} + \gamma_{12} = 1.$$

In addition to the normalization rule we know that x_1 and x_2 do not appear in the last two equations and have coefficients in the first equation summing to one. In the notation of equations (6.8) and (6.10) we have

$$\Psi_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$\Psi_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \Psi_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and

$$W_1^* = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} \\ 1 & 0 & 0 \end{bmatrix},$$

$$W_2^* = \begin{bmatrix} \beta_{12} & 1 & \beta_{32} \\ \gamma_{11} & 0 & 0 \\ \gamma_{12} & 0 & 0 \end{bmatrix}, \qquad W_3^* = \begin{bmatrix} \beta_{13} & \beta_{23} & 1 \\ \gamma_{11} & 0 & 0 \\ \gamma_{12} & 0 & 0 \end{bmatrix}.$$

Each of the W_p^* has rank less than three and hence none of the equations is identifiable. However, simple calculation yields from equation (6.11)

$$W_{14}^* = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} \\ 1 & 0 & 0 \\ \gamma_{11} & 0 & 0 \end{bmatrix}, \qquad W_{15}^* = \begin{bmatrix} 1 & \beta_{21} & \beta_{31} \\ 1 & 0 & 0 \\ \gamma_{12} & 0 & 0 \end{bmatrix}.$$

Since W_{14}^* and W_{15}^* have the same rank as W_1^*, both γ_{11} and γ_{12} are identifiable. The remaining unknown parameters ($\beta_{12}, \beta_{13}, \beta_{21}, \beta_{23}, \beta_{31}, \beta_{33}$) are not identifiable.

In closing we may note that the analysis here of the linear simultaneous equations model does not require any use of the information matrix. This is due to the fact that Π and Ω are assumed to completely characterize the distribution of Y. If, however, the u_t were known to have some nonnormal density or if the equation system (6.1) were nonlinear, this would no longer be the case. Identifiability typically does require the investigation of the probability distribution of the

observations. In these cases, the information matrix is often a convenient starting point.

University of California, Berkeley

Manuscript received March, 1969; revision received November, 1969.

REFERENCES

[1] AITCHISON, J., AND S. D. SILVEY: "Maximum-Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29 (1958).
[2] ———: "Maximum-Likelihood Estimation Procedures and Associated Tests of Significance," *Journal of the Royal Statistical Society*, Series B, 22 (1960).
[3] CRAMÉR, H.: *Mathematical Methods of Statistics*. Princeton, Princeton University Press, 1946.
[4] FISHER, FRANKLIN M.: "Generalization of the Rank and Order Conditions for Identifiability," *Econometrica*, 27 (1959).
[5] ———: "Identifiability Criteria in Nonlinear Systems," *Econometrica*, 29 (1961).
[6] ———: "Uncorrelated Disturbances and Identifiability Criteria," *International Economic Review*, 4 (1963).
[7] ———: *The Identification Problem in Economics*. New York, McGraw-Hill, 1966.
[8] GALE, DAVID, AND H. NIKAIDO: "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159 (1965).
[9] KOOPMANS, T. C., AND O. REIERSØL: "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21 (1950).
[10] KOOPMANS, T. C., H. RUBIN, AND R. B. LEIPNIK: "Measuring the Equation Systems of Dynamic Economics," in *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph 10, New York, John Wiley, 1950.
[11] LEHMANN, E. L.: *Testing Statistical Hypotheses*. New York, John Wiley, 1959.
[12] NEYMAN, J.: "Remarks on a Paper by E. C. Rhodes," *Journal of the Royal Statistical Society*, 100 (1937).
[13] REIERSØL, OLAV: "Identifiability of a Linear Relation between Variables which are Subject to Error," *Econometrica*, 18 (1950).
[14] ROTHENBERG, THOMAS J.: "Structural Restrictions and Estimation Efficiency in Linear Econometric Models," Cowles Foundation Discussion Paper 213, 1966.
[15] WALD, A.: "Note on the Identification of Economic Relations," in *Statistical Inference in Dynamic Economic Models*. New York, John Wiley, 1950.
[16] WEGGE, L.: "Identifiability Criteria for a System of Equations as a Whole," *Australian Journal of Statistics*, 7 (1965).