The Identification of Structural Characteristics (continued)

by T. C. Koopmans

An Econometric Example

The following general specification S_0 has been used extensively in econometric work*:

$$F\{u(1),\dots,u(T)\} = \prod_{t=1}^{T} F\{u(t)\}, \qquad \mathcal{E}u(t)=0, \quad \operatorname{tr}\mathcal{E}u'(t)u(t)<\infty,$$

(1) $$\phi\{y(t), y(t-1),\dots,y(t-\tau^0); u(t)\} \equiv$$
$$\equiv \mathrm{B}_0 y'(t) + \mathrm{B}_1 y'(t-1) + \dots + \mathrm{B}_{\tau^0} y'(t-\tau^0) + \Gamma z'(t) - u(t) = 0, \quad \det \mathrm{B}_0 \neq 0.$$

where the elements of the vector $z(t)$ are numerically given functions of time. Particular specifications S_1, S_2 ... have mostly consisted in prescribed values of elements of the matrix

(2) $$\mathrm{A} = (\mathrm{B}_0 \quad \mathrm{B}_1 \quad \dots \quad \mathrm{B}_{\tau^0} \quad \Gamma)$$

sometimes prescribed values of linear functions of the elements in one row of A. It can always be arranged that of the linear restrictions on any one row of A, at most one is non-homogeneous (normalization rule), the others homogeneous.

It has been shown that in the model S_0, a necessary and sufficient condition for the equivalence of two structures $S\{F(u), \mathrm{A}\}$ and $S^*\{F^*(u^*), \mathrm{A}^*\}$ is that they are connected by a linear transformation

(3) $$\mathrm{A}^* = \Upsilon \mathrm{A}, \qquad u^* = \Upsilon u$$

with non-singular matrix Υ. The model

(4) $\mathcal{S} = S_0 \times S_1 \times S_2 \times \dots$

identifies a parameter α_{gk} if, whenever both A and A^* belong to S, (3) implies that

(5) $$\alpha^*_{gk} = \alpha_{gk}$$

In order to attain such identifiability, it is necessary that one non-homogeneous restriction (normalization rule) on the g-th row of A be specified in S. If G represents the number of rows of A, (and the rank), it can be proved that

*The vector function z is from now on denoted by a small letter instead of a capital.

it is necessary for the simultaneous identifiability of all elements α_{gk}, $k = 1, \ldots, K$ in the g-th row α_g of $\mathcal{A}$, that at least G-1 additional non-homogeneous restrictions be imposed on that row, say

(6) $$\alpha_g \Phi'^{(g)} = 0, \qquad \rho(\Phi'^{(g)}) \geq G-1$$

where $\alpha_g \equiv (\alpha_{g1} \ \ldots \ \alpha_{gK})$, and $\rho(X)$ denotes the rank of X. These restrictions (6) are also sufficient (in addition to the normalization rule) if

(7) $$\rho(\mathcal{A}\Phi'^{(g)}) = G-1$$

The g-th row of the "rank criterion matrix" $\mathcal{A}\Phi'^{(g)}$ in (7) consists of zeros only, because of (6). Therefore, (7) requires the other rows to be linearly independent.

As stated before in general terms, the identifiability of the α_{gk} for given g is a property of the joint distribution function $\mathcal{G}\{y \mid z, \mathcal{A}, F(u)\}$ of the observations y, and as such subject to statistical test. In the present case, this is also seen from the fact that the rank of $\mathcal{A}\Phi'^{(g)}$ is preserved by the transformations (3).

<u>Comparative Remarks on the Three Examples Studied</u>

In each of the three examples considered, the model contains a general specification prescribing a parametric form of the structural relationships. Further particular specifications therefore take the form of parameter specifications in the function $\phi(y,u)$ and possibly in $F(u)$. A comparison of the three examples shows a striking formal similarity of the identification problem, which justifies our speaking of identification problems as a separate group of problems preparatory to statistical inference, of quite widespread occurrence. The same definitions of structure, model, parameter, identifiability are applicable and useful in each example.

The conclusion of formal similarity of the identification problem in the three cases deserves the main emphasis in the present context. In the following discussion we also discern differences of substance, connected largely with the more or less predominant role assigned to "theory" in the specifications of the model in each case. Nevertheless, this discussion is believed not to detract from the importance and generality of the formal concept of identifiability, and from the need for an analysis of identifiability and its criteria in specific applications of statistical methods.

To discuss further similarities and differences, I wish to distinguish in both the psychometric and the biometric case an aprioristic model and an empirical model, and to classify the econometric model as aprioristic. The aprioristic model in the psychometric example is obtained by postulating the existence of a "true rank" ρ, not necessarily equal to the minimum rank of Π, compatible with $\mu_{yy} = \Pi \mu_{\zeta\zeta} \Pi' + \Delta$, ($\Delta$ diagonal), but possibly bounded in the model by an a priori constant ρ_0. In the biometric example the aprioristic model is one in which merely normality of the ν's is postulated, but not "smallness."

In all three aprioristic models, non-identifiable parameters may or do occur, and the identifiability of some parameters depends on the value of identifiable functions of other parameters. In the econometric case, the identifiability of elements of α depends on the rank of matrices formed from other elements of α. In the aprioristic model of the psychometric example, the identifiability of ρ depends on whether the particular (uniformly identifiable) matrix $\mu_{yy} = \Pi \mu_{\zeta\zeta} \Pi' - \Delta$ does or does not permit more than one value of ρ below its a priori bound ρ_0.

In the a prioristic biometric model the covariance matrix Φ of the ν's is uniformly non-identifiable (calling the ζ's "parameters" instead of "random variables" gets around this fact only in an artificial and formal way, which creates more trouble than it avoids). With respect to ρ the situation seems completely analogous to the psychometric model. If an a priori bound ρ_0 on ρ permits only one value ρ such that there exists a Θ satisfying

$$\xi(\Theta) = G - \zeta , \qquad \Theta y' = w' \text{ is normally distributed,}$$

then ρ is identified and equal to that value. Whenever the distribution function $G(y)$ becomes such that more than one value is possible, ρ lacks identifiability. If ρ is identifiable, then Θ is identifiable by the orthagonalization imposed.

Another interesting similarity between the three situations, which could well have been expected, is the following. In each case where not all parameters are identifiable in all structures, uniformly identifiable functions of these parameters can be constructed, knowledge of which constitutes scientific information of a more limited usefulness. In the econometric example, the parameters

$$(I \quad \Pi) \equiv -B_0^{-1}\alpha , \qquad \Phi \equiv B_0^{-1} \Sigma B_0^{-1\prime}$$

of the reduced form

$$y' = \Pi z' + v' , \qquad \mathcal{E} v'v = \Phi$$

are still useful for prediction of y from z under unchanged structure, but insufficient for prediction of the effects of known changes in structure. In the psychometric example, re-definition of ρ as the minimum rank compatible with a given covariance matrix M_{yy} of test scores provides a "description" of the test battery and population at hand, which may or may not stand up as an "explanation" when more tests are added.

With ρ so re-defined as a uniformly identifiable parameter, we may speak of the empirical model in the psychometric example. Similarly, an empirical model in the biometric example is obtained, either by defining ρ as the smallest value which permits $\Theta y' = w'$ to be normally distributed, or, as suggested by Rasch, by some criterion of "smallness" of the elements of w.

In the first alternative ς is uniformly identifiable. The second alternative needs further elaboration before a similar conclusion can be drawn.

In the empirical models so defined, the principle of "simplicity" often stressed by Milton Friedman is utilized for the definition of ς. In the aprioristic models, on the other hand, this principle operates only to select the (linear) form of the relationships $\phi(y,u) = 0$, and perhaps properties of the distribution function $F(u)$, but <u>not</u> to determine the number of relations relevant to the study.

Considering now only the empirical models in the cases of the psychometric and biometric examples, we note a few more formal analogies. In all three cases, the degree of overidentification of the structure, allowed or supported by the data, is a matter of great interest. In the econometric and psychometric examples, a substantial degree of overidentification, of α and π respectively, is looked upon as empirical support for preconceived "theory". In the biometric example, in which theory plays a smaller role, overidentification of Θ adds to the practical usefulness of the relationships found.

An attempt at classification of the parameter specifications incorporated in the three models, with some reference to their motivations, will throw further light on similarities and differences between the three statistical situations.

The first, most trivial, group of parameter specifications consists of <u>rules of normalization.</u> The object of these specifications is to remove an inconvenient indeterminacy of which the presence and the trivial character are known to author and reader alike.

Orthogonalization (specifications of diagonality of covariance matrices of latent variables) may likewise be intended solely to produce determinacy in a situation of recognized indeterminacy. This is the case in the biometric example as long as no overidentifying parameter specifications are introduced, and in the case of factor analysis as long as only the space spanned by the common factors is sought, with no attempt to recognize simple structure on the basis of specific abilities. However, as soon as overidentifying parameter specifications are introduced elsewhere, the orthagonalizing specifications themselves usually also become overidentifying, in which case they become statements regarding the nature of the true structure, which may or may not be intended, and should be avoided or reconsidered in the latter case.

Still in the category motivated by convenience are the dummy parameter specifications the use of which has been advocated in the econometric problem whenever identifiability of some of the parameters has been found through prior analysis to be absent without such specifications. The purpose is again to facilitate the process of estimation of identifiable parameters in the face of the non-identifiability-in-principle of some other parameters: Formal identifiability of the latter parameters is produced by dummy specifications maintained only during a certain phase of the analysis.

The remaining (possibly over-) identifying restrictions on α in the econometric, on Π in the psychometric, and on Θ in the biometric case are differently motivated. In the econometric case, the motivation is primarily a priori, and hints from the data as to overidentifying parameter specifications will be accepted only where the evidence is strong and does not contradict "theory." In the psychometric case, mild a priori expectations as to where the zeros in Π fall are not in the least "imposed" by the model. The only thing imposed is that there should be many zeros, in places not preassigned, but duly scattered, and in their totality producing considerable overidentification. In the biometric example, the situation is still slightly different. Many zeros involving overidentification are welcomed, but if they can be arranged in nice blocks in the corners of [illegible], so much the better for the usefulness of the result.

<u>Note on the History of the Identification Problem, With References</u>

The identification problem has been discussed, in various terminologies and formulations, by quantitative thinkers in several fields. It is interesting to note that many important contributions have come from researchers so close to particular fields of application that they are not thought of <u>primarily</u> as mathematical statisticians by their purer colleagues.

In economics, contributions of increasing clarity and generality were made by Pigou, Henry Schultz, Frisch, Marschak and Haavelmo. The main contributions to the formalization and explicit mathematical analysis of the problem were made so far by Haavelmo, Koopmans and Rubin, and Hurwicz.

In factor analysis, Thurstone's discussion abounds with problems of identifiability. In biometrics, the "method of path coefficients" of Sewall Wright is essentially a method where a structure is postulated behind the observable distributions, and the identifiability of that structure discussed.

<u>References</u>:

Pigou, "A method of determining the numerical values of elasticities of demand."
<u>Economic Journal</u>, 1910, reprinted in "Economics of Welfare."

H. Schultz, "Theory and measurement of demand," Chicago, 1938, pp. 72-80.

Frisch, "Pitfalls in the statistical construction of demand and supply curves,"
Leipzig, 1933.
"Statistical versus theoretical relations in economic macrodynamics,"
Mimeographed, 1938.

Marschak, "Economic interdependence and statistical analysis," in "Studies in Mathematical Economics and Econometrics," Chicago, 1942.

Haavelmo, "The probability approach in econometrics," supplement to Econometrica July, 1944.

Koopmans and Rubin, "Measuring the equation systems of dynamic economics,"* Chapter V, section 2.

Hurwicz, "Generalization of the concept of identification,"*

Thurstone, see Reiersol paper No. 303.

Sewall Wright, "The method of path coefficients," Annals of Mathematical Statistics, September, 1934, pp. 161-215.

---

*In "Statistical Inference in Dynamic Economic Models," Cowles Commission Monograph No. 10, forthcoming.