Reprinted from THE ANNALS OF MATHEMATICAL STATISTICS
Vol. XXI, No. 2, June, 1950

THE IDENTIFICATION OF STRUCTURAL CHARACTERISTICS[1]

BY T. C. KOOPMANS AND O. REIERSØL

Cowles Commission for Research in Economics

1. Introduction.

1.1. *"Population" versus "structure."* In a fundamental paper (Fisher, [1]) R. A. Fisher distinguished as the first group of problems in mathematical statistics the "specification of the mathematical form of the population from which the data are regarded as a sample." It is the purpose of this article to suggest a reformulation of the specification problem, appropriate to many applications of statistical methods, and to point out the consequent emergence of a new group of problems, to be called identification problems.

In many fields the objective of the investigator's inquisitiveness is not just a "population" in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated. The word "physical" is used merely to convey that the structure concept is based on the investigator's ideas as to the "explanation" or "formation" of the phenomena studied, briefly, on his theory of these phenomena, whether they are classified as physical in the literal sense, biological, psychological, sociological, economic or otherwise. Examples of such structures, drawn from the fields of economic fluctuations and of psychological factor analysis, are given in sections 3 and 4. More detailed discussions of these examples can be found in other publications by the present authors and by others [15], [19]. In this article, we are therefore not concerned with the merits of particular assumptions entering into the specifications considered. Our examples are used only as the basis for a generalizing formulation (Section 2) and a comparative discussion (Section 5) of the identification problem, i.e., the problem of drawing inferences from the probability distribution of the observed variables to the underlying structure. The belief is here expressed that this 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.

The general formulation of the identification problem in Section 2 is, therefore, held abstract. Some readers may prefer to give substance to the various concepts by reading Sections 3–4 alongside Section 2. In addition, we insert here a simple example showing the main features of the identification problem.

---

[1] To be included in Cowles Commission Papers, New Series, No. 39. The authors reported on this study in papers before the Berkeley meeting of the Institute of Mathematical Statistics in June 1948. We are indebted to Dr. G. Rasch of the University of Copenhagen and to Professor L. L. Thurstone of the University of Chicago for many fruitful discussions on the subject matter of this article, for which the responsibility lies exclusively with the authors.

1.2. *A simple example of the identification problem.* This example is concerned with the problem of estimating the parameters α, β, of a linear relationship

$$\eta_2 = \alpha + \beta\eta_1 \tag{1.1}$$

between two variables η_1 and η_2 both of which are observed only subject to errors of observation u_1 and u_2. Thus, observations are available only for the variables

$$y_i = \eta_i + u_i \quad \text{where} \quad E(u_i) = 0, \quad i = 1, 2. \tag{1.2}$$

The question under what conditions a consistent estimate of β exists has repeatedly attracted attention. To discuss this question, we shall consider a model in which η_1 is independent of (u_1, u_2) and in which the joint distribution of u_1 and u_2 is normal.

If also the distribution of η_1 is normal, it is easy to see that β cannot be determined from a knowledge of the joint probability distribution of the observed variables y_1 and y_2.[2] In this case the joint distribution of y_1 and y_2 is also normal and the distribution is completely characterized by five parameters, $E(y_1)$, $E(y_2)$, var (y_1), var (y_2), and cov (y_1, y_2). The parameters β and var (η_1) may now be chosen in any way such that the second term in the right hand member of

$$\begin{bmatrix} \text{var } (y_1) & \text{cov } (y_1, y_2) \\ \text{cov } (y_1, y_2) & \text{var } (y_2) \end{bmatrix} = \begin{bmatrix} 1 & \beta \\ \beta & \beta^2 \end{bmatrix} \text{var } (\eta_1) + \begin{bmatrix} \text{var } (u_1) & \text{cov } (u_1, u_2) \\ \text{cov } (u_1, u_2) & \text{var } (u_2) \end{bmatrix}$$

is a positive definite matrix. It is clear that if the left hand member is non-singular, this condition can be met for any arbitrary value of β combined with a sufficiently small value of var (η_1).

It can be shown that β is uniquely determined by the joint probability distribution of y_1 and y_2 if this distribution is not normal. We shall prove this in the case that certain semi-invariants exist.[3]

Let $\phi_{y_1y_2}(t_1, t_2)$ denote the characteristic function of the joint distribution of y_1 and y_2

$$\phi_{y_1y_2}(t_1, t_2) = E(e^{y_1it_1 + y_2it_2}), \tag{1.3}$$

and let

$$\psi_{y_1y_2}(t_1, t_2) = \log \phi_{y_1y_2}(t_1, t_2). \tag{1.4}$$

Similar notations will be used for the characteristic functions of other random variables, and the logarithms of these functions.

Since (u_1, u_2) and (η_1, η_2) are independent, we obtain

$$\psi_{y_1y_2}(t_1, t_2) = \psi_{\eta_1\eta_2}(t_1, t_2) + \psi_{u_1u_2}(t_1, t_2), \tag{1.5}$$

[2] See [13], middle of page 70.

[3] The following proof is analogous to that given by Geary [8] in the case when the u's are not supposed to be normally distributed, but independent.

and from equations (1.1) and (1.3) we obtain

$$\phi_{\eta_1\eta_2}(t_1, t_2) = E(e^{\eta_1 it_1 + (\alpha+\beta\eta_1)it_2}) = e^{\alpha i t_2}\phi_{\eta_1}(t_1 + \beta t_2),$$

or

$$\psi_{\eta_1\eta_2} = \alpha i t_2 + \psi_{\eta_1}(t_1 + \beta t_2). \tag{1.6}$$

Combining (1.5) and (1.6), we have

$$\psi_{y_1y_2}(t_1, t_2) = \alpha i t_2 + \psi_{\eta_1}(t_1 + \beta t_2) + \psi_{u_1u_2}(t_1, t_2), \tag{1.7}$$

where $\psi_{u_1u_2}(t_1, t_2)$ is a polynomial of second degree, since the joint distribution u_1 and u_2 is normal. Let κ_{rs} be the semi-invariants of the distribution of (y_1, y_2) and let κ_r be the semi-invariants of the distribution of η_1. Comparing coefficients in equation (1.7), we obtain

$$\kappa_{rs} = \beta^s \kappa_{r+s} \qquad (r + s \geq 3) \tag{1.8}$$

and from this equation again

$$\kappa_{rs} = \beta\kappa_{r+1,s-1} \qquad (r + s \geq 3,\ s \geq 1). \tag{1.9}$$

If at least one κ_{rs} with $r + s \geq 3$, is finite and different from zero (which implies that the joint distribution of y_1 and y_2 is not normal), β may be determined from one such equation given the joint distribution function of y_1 and y_2.

1.3. *Remarks on the history of the identification problem.* The identification problem has been discussed, in various terminologies and formulations, by quantitative thinkers in several fields. It is interesting to note that most of the contributions have come from researchers whose main attention was directed to particular fields of application. For this reason, perhaps, its general formulation was not attempted until recently.

In economics, contributions of increasing explicitness and generality were made by Pigou [18], Henry Schultz [20], Frisch [3], [4], [5], [6], [7], Marschak [17]. The main contributions to the formalization and explicit mathematical analysis of the problem were made so far by Haavelmo [9], Koopmans and Rubin [15], Wald [24], and Hurwicz [10].

In his books on factor analysis [21], [22], Thurstone discusses in several places questions of identifiability. Previously the lack of identifiability in a certain factor analysis model had been demonstrated by numerical examples by G. H. Thomson [27]. Models used in the analysis of latent structure in attitude and opinion research by Lazarsfeld [16] give rise to similar identification problems. In biometrics, the "method of path coefficients" of Sewall Wright [25], is essentially a method where a structure is postulated behind the observable distribution, and the identifiability of that structure discussed. The identification problem is also met with in the theory of the design of experiments, particularly in the method of confounding (Fisher [2], Chapter 7, Yates [26]). When con-

founding is used, the identifiability of certain parameters (second order interactions, say) is sacrificed in order to gain certain advantages in the testing of hypotheses concerning (and in the estimation of) the parameters that remain identifiable (main effects and first order interactions, say).

2. General formulation of the identification problem.

2.1. *Latent variables, observed variables, and structure.* In each of the examples considered in this article, the distributional specification applies directly to certain non-observable or in any case non-observed variables, variously referred to as errors of observation (like u_1 and u_2 above), disturbances, "true" variables (like η_1 above), specific factors, etc. We shall refer to these as *latent variables*, denoted by a vector u. In addition, certain *structural relationships*—like (1.1) and (1.2)—are specified which connect the latent variables with the *observed variables*, denoted by a vector y. The specification is therefore concerned with the mathematical forms of both the distribution of the latent variables and the relationships connecting observed and latent variables.

The term "mathematical form" carries a suggestion of parametric specification which obviously is not the only possible type. We shall therefore employ terms and concepts introduced by Hurwicz [10] which cover both parametric and non-parametric specifications. By a *structure* $S = (F, \phi)$ we understand a particular probability distribution function

$$F(u) \tag{2.1}$$

of the latent variables—thought of, if you wish, as given numerically to a desired degree of accuracy, either by a cumulative distribution surface or curve or table, or parametrically by numerical values of the parameters—combined with a particular structural relationship (or set of simultaneously valid relationships)

$$\phi(y, u) = 0 \tag{2.2}$$

between observed and latent variables—again given numerically by curves, surfaces or parameters—which permits unique determination of the observed variables y from the values of the latent variables u (except possibly for a set of u-values occurring with probability zero). The corresponding probability distribution

$$H(y \mid S) \tag{2.3}$$

of the apparent variables is therefore uniquely determined by the structure S, and is said to be *generated* by S.

2.2. *Specification of a model.* We shall use the term *model* to signify a set of structures. We can thus say that the specification problem is concerned with specifying a model[4] $\mathfrak{S}$ which by hypothesis contains the structure S generating the distribution H of the observed variables.

[4] A set will be denoted by a German character corresponding to the Latin character denoting its representative element.

As a result of this reformulation of the specification problem, a new problem of inference arises, which logically precedes all problems of estimation or of testing hypotheses. It has already been deduced from the definition of structure that a given structure S generates one and only one probability distribution $H(y \mid S)$ of the apparent variables. However, statistical inference from any number of observations can relate only to characteristics of the distribution of the observed variables. The limit of statistical inference is an exact knowledge of this distribution function, a limit not attainable but approachable if very large samples can be taken. Anything not implied in this distribution is not a possible object of statistical inference.

2.3. *Identifiability of structural characteristics by a model.* It is therefore a question of great practical importance whether a statement converse to the one just made is valid: can the distribution H of apparent variables, generated by a given structure S contained in a model $\mathfrak{S}$, be generated by only one structure in that model? This is by no means implied in the definitions given, and it is not generally true. Whether or not it is true in a particular instance depends—as illustrated in our examples—always on the model $\mathfrak{S}$, and often on the given structure S besides. If it is true, we shall say that the model $\mathfrak{S}$ *identifies* the given structure S, or that the structure S is *identifiable* by the model.[5]

If a structure S is not identifiable by a model $\mathfrak{S}$, some of its characteristics may still be uniquely determinable. By a *structural parameter* $\theta(S)$ we understand a functional of the structure S. (This definition applies, of course, equally to the case of non-parametric specification of the functions F, ϕ defining the structure.) We further define that two structures S and S^* are (observationally) *equivalent* if they generate the same distribution of observed variables,

$$H(y \mid S) = H(y \mid S^*) \qquad \text{for all } y. \tag{2.4}$$

We then say that a model $\mathfrak{S}$ identifies a parameter $\theta(S)$ in a structure S_0, if that parameter has the same value in all structures S_0^*, contained in $\mathfrak{S}$ and equivalent to S_0. This definition can obviously be extended to characteristics $\chi(S)$ of a structure S, other than parameters, such as the functional form of a relationship represented by a component of the vector ϕ, etc.

2.4. *The identification problem.* It has now become clear that our reformulation of the specification problem has given rise to a new group of *identification* problems: to determine which of the parameters or other characteristics of a given structure are identifiable by (or "within") a given model.

It is perhaps premature to attempt assigning to identification problems a definite place in a classification of statistical problems such as was undertaken by Fisher. One might regard problems of identifiability as a necessary part of the specification problem. We would consider such a classification acceptable, provided the temptation to specify models in such a way as to produce identifiability of relevant characteristics is resisted. Scientific honesty demands that

[5] The concept here designated briefly as "identifiability" has been called "unique identifiability" in another context (Koopmans and Rubin [15], also Hurwicz [10]) in contrast with "multiple" or "incomplete" identifiability.

the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested.

Identification problems are not problems of *statistical* inference in a strict sense, since the study of identifiability proceeds from a hypothetical exact knowledge of the probability distribution of observed variables rather than from a finite sample of observations. However, it is clear that the study of identifiability is undertaken in order to explore the limitations of statistical inference.

2.5. *Identifiability is subject to statistical test.* Further interpenetration of the pre-statistical analysis of identifiability with problems of statistical inference proper arises from the fact, amply illustrated by our examples, that the identifiability of a structural characteristic $\chi(S)$ often depends not only on the model, but also on the given structure S. Thus, each structural characteristic χ divides the model $\mathfrak{S}$ exhaustively into two mutually exclusive subsets of structures

$$\mathfrak{S} = \mathfrak{S}_\chi + \mathfrak{S}_{\bar{\chi}} \tag{2.5}$$

(of which one may be empty), such that $\chi(S)$ is uniquely identifiable in S_0 by the model if S_0 belongs to $\mathfrak{S}_\chi$, and not uniquely identifiable if S_0 belongs to $\mathfrak{S}_{\bar{\chi}}$. We shall call $\chi(S)$ *uniformly identifiable* by $\mathfrak{S}$ if $\mathfrak{S}_{\bar{\chi}}$ coincides with $\mathfrak{S}$.

The subdivision of $\mathfrak{S}$ into $\mathfrak{S}_\chi$ and $\mathfrak{S}_{\bar{\chi}}$ has an important property: If S_0 belongs to $\mathfrak{S}_\chi$, then all structures S_0^* equivalent to S_0 also belong to $\mathfrak{S}_\chi$, and a similar statement holds for $\mathfrak{S}_{\bar{\chi}}$. This property follows directly from the definition of identifiability of $\chi(S)$ given above. Its meaning is that the identifiability of $\chi(S)$ in S_0 depends only on the distribution of $H(y) = H(y \mid S_0)$ of observed variables generated by S_0. To the subdivision of the model corresponds an exhaustive subdivision

$$\mathfrak{H} = \mathfrak{H}_\chi + \mathfrak{H}_{\bar{\chi}} \tag{2.6}$$

of the set

$$\mathfrak{H} = \mathfrak{H}(\mathfrak{S}) \tag{2.7}$$

of all distribution functions $H(y \mid S)$ generated by the structures S of $\mathfrak{S}$, into the subset $\mathfrak{H}_\chi$ containing those distribution functions $H(y \mid S)$ generated by structures S in which $\chi(S)$ is uniquely identifiable, and the subset $\mathfrak{H}_{\bar{\chi}}$ containing functions $H(y \mid S)$ generated by structures for which the opposite is true.

Hence, whenever the identifiability of $\chi(S)$ cannot be decided in the same sense (affirmatively or negatively) for all structures S of $\mathfrak{S}$ as a result of either $\mathfrak{S}_\chi$ or $\mathfrak{S}_{\bar{\chi}}$ being empty, then the identifiability of the characteristic $\chi(S)$ of the structure S generating the observations is a property of the distribution $H(y \mid S)$ of the observations. This identifiability is equivalent to the hypothesis

$$H(y \mid S) \quad \text{belongs to} \quad \mathfrak{H}_\chi, \tag{2.8}$$

which is in principle[6] subject to statistical test under the maintained hypothesis

$$H(y \mid S) \quad \text{belongs to} \quad \mathfrak{H}. \tag{2.9}$$

2.6. *Testing particular specifications.* Often the model is defined by one general specification supplemented with a number of particular specifications which are "detachable pieces" in the sense that they can be removed, added or replaced by alternatives to construct alternative models. We may define the *general specification* as a set $\mathfrak{S}$ of structures which is postulated to contain the model $\mathfrak{S}'$ in question as a subset. *Particular specifications* can then be defined as subsets $\mathfrak{S}_1, \mathfrak{S}_2, \cdots$ of $\mathfrak{S}$ of which the model $\mathfrak{S}'$ is the intersection

$$\mathfrak{S}' \equiv \mathfrak{S} \cap \mathfrak{S}_1 \cap \mathfrak{S}_2 \cap \cdots . \tag{2.10}$$

An example is that of parametric specification of the "form" of the functions $\phi(y, u)$ defining the structural relationships and of the distribution function $F(u)$ of latent variables as the general specification, and specifications of the values of certain parameters of ϕ and F as particular specifications.

In such situations, it is an important question whether a given particular specification is—again in principle—subject to statistical test. Whenever the answer depends on the other particular specifications, we may ask further which minimum set of other particular specifications must (together with the general specification) be entered into the "maintained hypothesis" in order that that given particular specification be subject to statistical test. A formal answer to this question, facilitating specific answers in each concrete case, can be given as follows.

Let a model $\mathfrak{S}$ be narrowed down to an alternative model

$$\mathfrak{S}' = \mathfrak{S} \cap \mathfrak{S}_1 \tag{2.11}$$

by a particular specification $\mathfrak{S}_1$. This particular specification will be called *observationally restrictive* if the set $\mathfrak{H}(\mathfrak{S}')$ of all distribution functions $H(y \mid S')$ of observed variables generated by the structures S' of $\mathfrak{S}'$ is a proper subset of the set $\mathfrak{H}(\mathfrak{S}')$ of all distribution functions $H(y \mid S)$ generated by the structures S of $\mathfrak{S}$. A statistical test of the particular specification $\mathfrak{S}_1$ can then be constructed by choosing as the hypothesis subject to test

$$H(y) \quad \text{belongs to} \quad \mathfrak{H}(\mathfrak{S}'), \tag{2.12}$$

and as the maintained hypothesis

$$H(y) \quad \text{belongs to} \quad \mathfrak{H}(\mathfrak{S}). \tag{2.13}$$

The particular specification $\mathfrak{S}_1$ remains subject to test if the model $\mathfrak{S}$ is stripped of such other particular specifications which are not necessary for the observationally restrictive character of $\mathfrak{S}_1$, although of course the outcome of the test may become either less or more certain as a result.

[6] See sub-section 2.7 below.

A frequent case of an observationally restrictive specification is that where a parameter $\theta(S)$ already identifiable in almost all structures S of $\mathfrak{S}$, is restricted by $\mathfrak{S}_1$ to a prescribed value (or to a prescribed point set not containing all points of its domain for all S of $\mathfrak{S}$). In this case, the specification in question has been called *overidentifying*.

2.7. *Remarks on the testing of hypotheses.* In subsections 2.5 and 2.6 we have without further inquiry applied the expression "hypothesis in principle subject to test" to any hypothesis which narrows down the set $\mathfrak{H}$ of distribution functions H generated by structures of the model to a proper subset $\mathfrak{H}'$. It will be clear that, to make a test actually possible, $\mathfrak{H}'$ cannot be allowed to be everywhere dense in $\mathfrak{H}$. For instance, if $\mathfrak{H}$ is defined parametrically, a hypothesis restricting $\mathfrak{H}'$ to rational values of the parameters is clearly not subject to statistical test. Just what set-theoretical requirements on $\mathfrak{H}'$ are needed to make a test possible is a separate problem which we shall not attempt to discuss.

We have also in another sense oversimplified the problem of testing particular specifications. In practice this problem presents itself as the choice of one out of many possible combinations of several particular specifications, rather than a number of separate and unconnected choices between the rejection and the adoption of each particular specification under consideration. Present theory of choice between two alternatives does not meet this situation.

3. An econometric example.[7]

In econometric studies[8] economic fluctuations have been described by a system of difference equations in (observed) economic variables y, subject to two kinds of outside influences, emanating respectively from (observed) exogenous—i.e., non-economic—variables z, and from (latent) random disturbances u. Each of these equations is given a definite meaning in terms of economic behavior. There may for instance be equations explaining respectively consumption expenditure (from incomes of various groups, price changes, etc.), the supply of consumers' goods (from price margins between such goods and their raw materials and labor, productive capacity, etc.), investment expenditure, the supply of capital goods, etc. The purpose of the identification discussion is to investigate whether, on the basis of given a priori knowledge as to the form of these equations, and in particular as to what variables occur in any designated equation, procedures of estimation or testing of hypotheses can be directed to the parameters of the equations of economic behavior themselves, rather than to the parameters of "secondary" equations dependent on (derivable from) two or more of the behavior equations.

In the case of linear systems of equations, a possible form for the general specification (the model $\mathfrak{S}$) is as follows.

$$(3.1)\quad \mathrm{B}_0 y'(t) + \mathrm{B}_1 y'(t-1) + \cdots + \mathrm{B}_{\tau_{\max}} y'(t-\tau_{\max}) + \Gamma z'(t) = u'(t)$$

[7] For an expository discussion of identification problems in econometric models see [14].

[8] See, for instance, J. Tinbergen [23] and L. R. Klein [12].

represents the structural relationships. Here $y'(t)$, $z'(t)$, $u'(t)$ are column vectors (the transposes of row vectors) of G, K and G elements, respectively, for each discrete time point or period $t = 1, 2, \cdots, T$, also $t = 0, -1, \cdots, 1 - \tau_{\max}$, for $y'(t)$. $B_0, B_1, \cdots, B_{\tau_{\max}}$ are square matrices of order G, and Γ is a matrix of G rows and K columns.

(3.2) B_0 is non-singular.

(3.3) The observed values $z(t)$, $t = 1, \cdots, T$, are held constant in repeated samples, and the components of $z(t)$ are linearly independent.

(3.4) The components of $u(t)$ have a joint distribution function $F(u)$ (with zero means and finite variances) which is independent of t and of $z(t)$.

(3.5) $u(t)$ and $u(t')$ are independently distributed if $t \neq t'$.

Particular specifications $\mathfrak{S}_1, \mathfrak{S}_2, \cdots$, that have been most frequently employed indicate prescribed values (usually zero) of specified elements of the matrix

$$A \equiv [B_0 \; B_1 \cdots B_{\tau_{\max}} \; \Gamma] \tag{3.6}$$

or of given linear functions of the elements of the g^{th} row $\alpha(g)$ of A, for each value $g = 1, \cdots, G$ of g. 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. The homogeneous restrictions state which variables enter into each equation, and possibly with which ratios between some of their coefficients.

It has been shown [15] that in the model $\mathfrak{S}$, a necessary and sufficient condition for the equivalence of two structures $S \equiv \{F(u), A\}$ and $S^* \equiv \{F^*(u^*), A^*\}$ is that they are connected by a linear transformation

$$A^* = \Upsilon A, \qquad u'^* = \Upsilon u', \tag{3.7}$$

with non-singular matrix Υ. By definition, the model

$$\mathfrak{S}' = \mathfrak{S} \cap \mathfrak{S}_1 \cap \mathfrak{S}_2 \cap \cdots \tag{3.8}$$

identifies a parameter α_{gk} if, whenever A and A* belong to equivalent structures S and S^*, respectively, of $\mathfrak{S}'$, we have

$$\alpha^*_{gk} = \alpha_{gk}. \tag{3.9}$$

In order to attain such identifiability by linear restrictions on the g^{th} row of A it is necessary that one non-homogeneous restriction (normalization rule) on the g^{th} row of A be specified in $\mathfrak{S}'$. Recalling that G represents the number of rows (and the rank) of A, it can be proved that it is further necessary for the simultaneous identifiability of all elements α_{gk}, $k = 1, \cdots, K$, in the g^{th} row $\alpha(g)$ of A, that at least $G - 1$ additional non-homogeneous restrictions be imposed on that row, say

$$\alpha(g)\Phi'(g) = 0, \qquad \rho\{\Phi'(g)\} \geqq G - 1, \tag{3.10}$$

where $\alpha(g) \equiv [\alpha_{g1} \cdots \alpha_{gK}]$, the $\Phi(g)$ are given matrices (often with elements 0 or 1 only), and $\rho(X)$ denotes the rank of X. These restrictions (3.10) are also sufficient (in addition to the normalization rule) if

$$\rho\{A\Phi'(g)\} = G - 1. \tag{3.11}$$

The g^{th} row of the "rank criterion matrix" $A\Phi'(g)$ in (3.11) consists of zeros only, because of (3.10). Therefore, (3.11) requires the other rows of that matrix to be linearly independent.[9]

Thus, even if the model $\mathfrak{Z}'$ includes, besides a normalization rule, the necessary condition (3.10) for the identifiability of the g^{th} behavior equation, such identifiability is still absent in certain structures, corresponding to a point set (generally of measure zero) in the space of the coefficients of the remaining equations, viz., the point set in which (3.11) is not satisfied. Whether or not A actually falls within this point set is, as was stated before in more general terms, a property of the joint distribution function $H(y \mid z)$ of the observations y, and is therefore subject to statistical test. In the present case, this is also seen from the fact that the rank of $A\Phi_0'$ is preserved by the transformation (3.7), and is therefore itself an identifiable parameter.

For certain scientific purposes explicit knowledge of A is unnecessary. One such purpose is "prediction without change in structure," i.e., prediction of a value of $y(t)$ for a future time t from a hypothetical value of $z(t)$ on the assumption that A and $F(u)$ have not changed between the observation period and the time point to which the prediction applies. Such prediction can be based on the knowledge of (*a*) the population regressions

$$y'(t) = \Pi_1 y'(t-1) + \cdots + \Pi_{\tau_{\max}} y'(t-\tau_{\max}) + \Pi_z z'(t) + v'(t) \tag{3.12}$$

of the "jointly dependent" variables $y(t)$ on the "predetermined" variables $y(t-1), \cdots, y(t-\tau_{\max}), z(t)$ and of (*b*) the distribution function $K(v)$ of the population residuals

$$v(t) = y(t) - E\{y(t) \mid y(t-1), \cdots, y(t-\tau_{\max}), z(t)\} \tag{3.13}$$

from these regressions. Of course, the matrices "Π" are functions of the structural parameters (3.6) through

$$[-I\ \Pi] \equiv [-I\ \Pi_1 \cdots \Pi_{\tau_{\max}}\ \Pi_z] = -B_0^{-1}A \tag{3.14}$$

and $K(v)$ can be derived from $F(u)$ through the transformation

$$v' = B_0^{-1}u'. \tag{3.15}$$

The important fact is that Π and $K(v)$, by their definitions, depend only on the distribution function $H(y \mid z)$ of the observations, and are therefore uniformly identifiable. This is also reflected in the fact that the right hand members of (3.14) and (3.15) are invariant for the transformation (3.7).

[9] In that case, overidentification of $\alpha(g)$ will result if the inequality sign in (3.10) holds.

However, the most relevant economic problems are those in which a change in A or $F(u)$ is actually or hypothetically present, and in which therefore the identifiability of the relevant parts or functions of A and of the characteristics of $F(u)$ requires separate inquiry.[10]

4. An example from factor analysis.[11] Factor analysis has been presented in different forms by different authors. We shall here consider the multiple factor analysis of Thurstone only [21], [22].

The factor analysis methods were developed primarily for the purpose of analyzing intelligence tests, but they have also been used for other psychological problems and in other sciences.

Suppose that a person is given a battery of G tests. Let his score in test i be y_i. The fundamental assumption in factor analysis is that these scores can be explained in terms of a relatively small number of hypothetical primary factors. Let $z_1, z_2, \cdots, z_\rho$ denote the hypothetical scores of the person in the common factors, i.e., those primary factors which are common to at least two tests in the battery. We assume that y_i is a homogeneous linear function of the scores z_k plus a unique part v_i, which may be thought of as consisting of an error term plus the contribution of a specific factor. The coefficients π_{ik} in the linear function just mentioned are called factor loadings. The factor loading π_{ik} expresses the relative importance of the common factor k in the answering of test i.

We shall introduce the row vectors $y = [y_i]$, $z = [z_i]$, $v = [v_i]$ and the matrix $\Pi = [\pi_{ik}]$. The covariance matrices of the sets of variables y, z, and v will be denoted by M_{yy}, M_{zz}, and Δ, respectively.

In contrast with the preceding example, the variables y are the only observed variables. The variables v and z are latent variables.

Our model will be given by the following specifications:

(4.1) $y' = \Pi z' + v'$.

(4.2) $E(z) = 0$ and $E(v) = 0$.

(4.3) The set of variables z is stochastically independent of the set of variables v.

---

[10] See Hurwicz [11].

[11] Proofs of the statements in this section will be found in a separate paper by one of the authors (Reiersøl [19]). It should be noted that the notation is different in the two papers. In the separate paper the notation is close to that of Thurstone. In the present paper the notation has been chosen to correspond in some way to the notation in the econometric example. A list of corresponding symbols in the present paper and in Thurstone's books follows:

Present paper:	y_i	z_k	π_{ik}	G	ρ	M_{yy}	M_{zz}	Δ
Thurstone:	s_j	x_m	a_{jm}	n	r	R_1	R_{pq}	$R_1 - R$

It should be noted that M_{yy}, M_{zz}, and Δ are covariance matrices of the original variables, while R_1, R_{pq}, and R are covariance matrices of standardized variables.

(4.4) Δ is diagonal and different from 0.

(4.5) The elements of z and v are jointly normally distributed.

(4.6) Each y_i is correlated with at least one of the other y's.

(4.7) The rank of Π equals the number ρ of its columns.

(4.8) M_{zz} is nonsingular.

(4.9) ρ is the smallest number of variables z which is compatible with the joint probability distribution of the observed variables y and specifications (4.1)–(4.8).

(4.10) Each column of Π contains at least ρ zeros (in unspecified places).

(4.11) A normalization rule fixing the units of the variables x and a rule fixing the order of the columns of Π.

Denote by Π_k the matrix consisting of all the rows of Π which have a zero in the k^{th} column. Let the number of rows in the matrix Π_k be p_k. Let Π_{ki} denote the submatrix of Π_k which we get when deleting the i^{th} row of Π_k. Using these notations we shall formulate the final specification of our model.

(4.12) The rank of each of the matrices Π_{ki} ($k = 1, 2, \cdots, \rho$; $i = 1, 2, \cdots, p_k$) is $\rho - 1$.

Specification (4.1) represents the structural relationships.

Specification (4.10) means that the experimenter thinks he can construct a sufficient number of tests where at least one of the common primary factors is absent.

We shall first consider a model $\mathfrak{S}$ containing Specifications (4.1)–(4.9) only. From (4.9) follows that ρ is uniformly identifiable.

Let $\rho_G = \frac{1}{2}(2G + 1 - \sqrt{8G + 1})$. If $\rho > \rho_G$, the matrix Δ is generally not identifiable. If $\rho < \rho_G$, Δ generally is identifiable. When $\rho = \rho_G$, the number of values of Δ, which correspond to a given covariance matrix M_{yy}, is usually finite, and may be equal to one or greater than one. The matrices Π and M_{zz} are never identifiable in the model $\mathfrak{S}$. If Δ is identifiable, the set of all structures $\{\Pi^*, M^*_{zz}, \Delta\}$ equivalent to the structure $\{\Pi, M_{zz}, \Delta\}$ is given by the set of all matrices

$$\Pi^* = \Pi\Psi \tag{4.13}$$

and

$$M^*_{zz} = \Psi^{-1}M_{zz}(\Psi')^{-1}, \tag{4.14}$$

where Ψ is any square, ρ-rowed and nonsingular matrix.

In the following we shall confine our discussion to the case $\rho < \rho_G$, and to structures in which the matrix M_{yy} is such that Δ is identifiable in $\mathfrak{S}$.

We shall now consider the model $\mathfrak{S}'$ defined by Specifications (4.1)–(4.11). In this model a necessary and sufficient condition for the identifiability of Π is that any square ρ-rowed minor of Π which is of rank $\rho - 1$ is contained in one of the matrices Π_k. This condition excludes the possibility that all elements belonging to the intersection of $\rho - 1$ rows and two columns of Π are all equal to zero. In order to be able to use this result, the experimenter would have to be able to construct tests where one, but not more than one, common factor would be absent. Therefore the result is not particularly useful. In order not to exclude the case where two common factors occur in more than $\rho - 2$ tests, we have introduced Specification (4.12).

We shall finally consider the model $\mathfrak{S}''$ defined by Specifications (4.1)–(4.12). Assuming M_{yy} known, we can determine some value Π^* of Π which satisfies Specifications (4.1)–(4.9). Since, by assumption, Δ is identifiable in $\mathfrak{S}$, Π^* must be of the form $\Pi\Psi$, where Π is the true factor loadings matrix and Ψ is non-singular. Let Π_k^* be a submatrix of Π^* containing all the columns of Π^* and satisfying the following conditions

(4.15) The rank of Π_k^* is $\rho - 1$.

(4.16) The addition to Π_k^* of a row contained in Π^* but not in Π_k^* increases the rank to ρ.

(4.17) Each submatrix of Π_k^* obtained by deleting one row of Π_k^* has rank $\rho - 1$.

A necessary and sufficient condition for the identifiability of Π in the complete model $\mathfrak{S}''$ is that there exist exactly ρ submatrices Π_k^* of Π^* which satisfy conditions (4.15)–(4.17), and that the ρ vectors q_k, satisfying the equations $\Pi_k^* q_k = 0$ when $k = 1, 2, \cdots, \rho$, are linearly independent.

It should be noted that Specifications (4.10) and (4.12) are observationally restrictive, i.e., they are in principle subject to statistical test.

5. A comparative discussion of the examples given. Some comparative remarks on the three examples given in sections 1.2, 3 and 4 may illustrate our general discussion of the identification problem, given in section 2.

In each of the three examples considered, the model contains a general specification prescribing a parametric form of the structural relationships (2.2). Further particular specifications therefore take the form of parameter specifications in the function $\phi(y, u)$ in (2.2) and possibly in the distribution function (2.1) of latent variables. A comparison of the three examples shows a striking formal similarity of the identification problems to which they give rise. This similarity 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. In all three cases, parameters occur, the identifiability of which depends on other identifiable structural characteristics (the normality of a distribution function in one case, the ranks of parameter matrices in the other two cases).

Our remaining remarks will be drawn from the econometric and factor analysis examples only, partly because these illustrate the identification problem in greater elaboration, partly because the closer similarity of these examples permits us to notice interesting differences in greater detail.

Let us consider the particular case of the econometric example when there are no time lags between the y's in the structural relationships, i.e., when $\tau_{\max} = 0$. In this case the *reduced form* (3.12) in the econometric example is of the same form as equation (4.1), which defines the structural relationships in the factor analysis example. The notation in the factor analysis example has been chosen with this similarity in mind. However, it should be emphasized that, while the variables y are observed in both examples and the variables v are latent in both examples, the variables z are observed in the econometric example and latent in the factor analysis example, and even the number of variables z is an unknown parameter ρ in the latter example. For this reason, the discussion of the identifiability of Δ in factor analysis has no counterpart in the econometric model. Furthermore, the identifiability of the matrix Π, which is automatic and uniform in the econometric model $\mathfrak{S}_e$, say, requires detailed specifications in the factor analysis model $\mathfrak{S}_f$, say, including the diagonality of Δ and prescriptions about the number of zero elements in each column.

The observability of z in the econometric case is exploited to postulate, behind the reduced form (3.12), a structure $(F(u), \Lambda)$ to be identified (where possible) from further specifications based on economic theory. Here we meet with another analogy, with differences, between the identification problem of Λ in $\mathfrak{S}_e$ and that of Π (given Δ) in $\mathfrak{S}_f$. In the latter problem, the set of matrices Π^*, belonging to a set of equivalent structures, is given by equation (4.13). This equation is analogous to the first of the equations (3.7) in the econometric case, with Π in $\mathfrak{S}_f$ now corresponding to Λ' in $\mathfrak{S}_e$.

If we were to specify zeros in assigned places in the factor loadings matrix Π, and to introduce a normalization rule for each column of Π, the results quoted in the econometric example would immediately be applicable to the factor analysis case. A necessary condition for the identifiability of Π, given that of Δ, would be that the number of specified zeros in each column of Π be at least $\rho - 1$. Necessary and sufficient for identifiability would be that the matrix consisting of all rows of Π which have specified zeros in the k^{th} column, be of the rank $\rho - 1$, for each value of k.

However, instead of specifying that given elements of Π be equal to zero, Thurstone assumes that we know that there is a certain minimum number of zeros in each column, but that we do not know which particular elements are zero. The specification of a certain number of zeros in undesignated places obviously represents a weaker assumption than the specification of the same number of zeros in designated places. It is therefore not surprising that the specification of $\rho - 1$ zeros in undesignated places in each column is never sufficient for identifiability of Π. Thus, in the model $\mathfrak{S}_f$, we have introduced the stronger specification (4.10). We have seen that even this specification is too weak to be practically

useful, and have introduced the additional Specification (4.12), which makes the factor analysis model still more different from the econometric model.

Continuing the analogy in which A' in $\mathfrak{S}_e$ corresponds to Π in $\mathfrak{S}_f$, we note an important feature common to both examples, and present in other situations as well. Even if specifications sufficient, in number and variety of "points of application," for the identifiability of all structural parameters cannot be derived from a priori considerations, it remains possible to construct uniformly identifiable functions of these parameters, knowledge of which constitutes scientific information of more limited usefulness.

In the econometric example we have already seen that for certain purposes a knowledge of the uniformly identifiable matrix Π of the reduced form is sufficient, while for other purposes we need to know the matrix A. As a further illustration, suppose that we want to test for persistence of the structure by comparing the equation systems which we estimate from data for two different periods. Disregarding errors of estimation (which are not our present topic), if A is the same in both cases, Π will also be the same in both cases. It is therefore possible to arrive at a rejection of the persistence hypothesis by determining Π in both cases. Suppose next that one row (or several rows) of A are different in the two periods, while the other rows of A are identical in the two cases. If B_0 changes from one period to the other, we may expect each element of Π to change. If we can determine A for each period, the equality (as between periods) of some of the rows of A will indicate precisely the extent of validity of the persistence hypothesis. If we cannot determine A but only Π in each case, this verification will be lost.

Similarly, it may in factor analysis be sufficient for some purposes to consider what we may call the reduced form of Π. Let Π_I be the upper square part of Π which we shall assume to be nonsingular. The matrix $\Lambda = \Pi\,\Pi_I^{-1}$ will be called the reduced form of Π. It will be of the form $\begin{bmatrix} I \\ \Lambda_{II} \end{bmatrix}$. Λ is always identifiable when Δ is identifiable.

Suppose now that the same battery of tests is given to two different populations. Suppose that some of the factor loadings are different in the two populations, while other factor loadings are the same. If at least one of the different factor loadings occurs in the matrix Π_I, then each element of Λ_{II} may be expected to change, and the partial identity of the two structures cannot be discovered if we determine Λ only and not Π. On the other hand, if Π is the same in both cases, also Λ will be the same in both populations.

Let us next consider two different batteries given to the same population. We shall suppose that the two batteries have some tests in common. For each test which is common to the two batteries we ought to find the same factor loadings in both batteries. In other words, the matrices Π in the two cases ought to be partly identical. On the other hand, if Π_I contains rows corresponding to tests which are not common to the two batteries, the matrices Λ_{II} will be entirely different in the two cases. Therefore, again, identification of Π will be necessary to verify the equality of the factor loadings of tests common to both batteries.

A final remark relates to observationally restrictive specifications. Particularly where the model is to a large degree speculative, empirical confirmation of the validity or usefulness of the model is obtained only to the extent that observationally restrictive specifications are upheld by the data. Thus, Thurstone emphasizes that the number of factors ρ should be well below the value ρ_G found above to be necessary in general for the identifiability of Δ, before a factor analysis can be regarded as successful (Thurstone [22], p. 294).

In econometric work, greater reliance is sometimes placed on a priori specification of the form of a behavior equation, particularly the variables occurring in it. If the linear restrictions on an equation in a linear system are just sufficient for its identifiability, estimation of the parameters of that equation is possible, but none of the identifying restrictions are themselves subject to test. Again, dependence on a priori information is diminished (but not eliminated) to the extent that a greater number of overidentifying restrictions are imposed and are upheld by the data.

REFERENCES

[1] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc.*, London, Ser. A, Vol. 222 (1922), p. 309.

[2] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh and London, 1935.

[3] R. Frisch, "Correlation and scatter in statistical variables," *Nordic Stat. Jour.*, Vol. 1 (1928), p. 36.

[4] R. Frisch and B. D. Mudgett, "Statistical correlation and the theory of cluster types," *Jour. Am. Stat. Assn.*, Vol. 26 (1931), p. 375.

[5] R. Frisch, "Pitfalls in the statistical construction of demand and supply curves," *Veröffentlichungen der Frankfurter Ges. für Konjunkturforschung*, Neue Folge, Heft 5, Leipzig 1935.

[6] R. Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publ. No. 5, Universitetets Økonomiske Institutt, Oslo, 1934.

[7] R. Frisch, *Statistical Versus Theoretical Relations in Economic Macrodynamics*, mimeographed document for League of Nations conference, 1938.

[8] R. C. Geary, "Inherent relations between random variables," *Proc. Roy. Irish Acad.*, Vol. 47 (1942), Sect. A, No. 6.

[9] T. Haavelmo, "The probability approach in econometrics," *Econometrica*, Vol. 12 (1944), Suppl.

[10] L. Hurwicz, "Generalization of the concept of identification," *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York, John Wiley and Sons, 1950.

[11] L. Hurwicz, "Prediction and least-squares," ibid.

[12] L. R. Klein, "Economic fluctuations in the United States, 1921–1941," *Cowles Commission Monograph 11*, New York, John Wiley and Sons, 1950.

[13] T. C. Koopmans, *Linear Regression Analysis of Economic Time Series*, Netherlands Economic Institute, Publ. No. 20, Haarlem, 1937.

[14] T. C. Koopmans, "Identification problems in economic model construction," *Econometrica*, Vol. 17 (1949), p. 125.

[15] T. C. Koopmans, H. Rubin, and R. B. Leipnik, "Measuring the equation systems of dynamic economics," *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York, John Wiley and Sons, 1950.

[16] P. F. LAZARSFELD, "The logical and mathematical foundation of latent structure analysis." "The interpretation of some latent structures." *Measurement and prediction*, Vol. 4, Studies in Psychology of World War II, 1950.

[17] J. MARSCHAK, "Economic interdependence and statistical analysis," *Studies in Mathematical Economics and Econometrics* (1942), University of Chicago Press, 1942, pp. 135–150.

[18] A. C. PIGOU, "A method of determining the numerical values of elasticities of demand," *Economic Journal*, Vol. 20 (1910), pp. 636–640.

[19] O. REIERSØL, "On the identifiability of parameters in Thurstone's multiple factor analysis," *Psychometrika*, Vol. 15 (1950).

[20] H. SCHULTZ, *Theory and Measurement of Demand*, University of Chicago Press, 1938.

[21] L. L. THURSTONE, *The Vectors of Mind*, University of Chicago Press, 1935.

[22] L. L. THURSTONE, *Multiple-factor Analysis*, University of Chicago Press, 1947.

[23] J. TINBERGEN, *Business cycles in the United States of America, 1919–1932*, League of Nations, Geneva, 1939.

[24] A. WALD, "Note on the identification of economic relations," *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York, John Wiley and Sons, 1950.

[25] S. WRIGHT, "The method of path coefficients," *Annals of Math. Stat.*, Vol. 5 (1934), p. 161.

[26] F. YATES, "The design and analysis of factorial experiments," *Imp. Bur. Soil Sci. Tech. Comm.*, No. 35, 1937.

[27] G. H. THOMSON, "The proof or disproof of the existence of general ability," *Brit. Jour. Psych.*, Vol. 9 (1919), p. 323.