A GENERAL APPROACH TO IDENTIFICATION

by

B.B. van der Genugten

Research memorandum

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Hogeschoollaan 225 Tilburg (Holland)
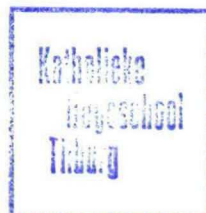
FEW
61.

A GENERAL APPROACH TO IDENTIFICATION

by

B.B. van der Genugten

Research Memorandum

R 41

v identification

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Hogeschoollaan 225, Tilburg (Holland)

Summary.

The paper deals with the concept of identification in inferential statistics. At first a general concept of identification is defined and developed. Thereafter, the general theory is applied to univariate linear regression and simultaneous equation systems. Finally, attention is paid to models with lagged variables and some new related problems are stated.

Contents.

1. <u>Introduction.</u>

The problem of identification arises in various fields of inferential statistics. In univariate linear regression it has led to the theory of estimable or identifiable functions (e.g. see Bose [2] and Scheffé [2, section 1.4]). In simultaneous equation systems it culminates in the question of the identification of structural equations (Koopmans [7], Fisher [4], Malinvaud [6, chapter 18], Schönfeld [11, chapter 15]). Other fields that should be mentioned are time-series (Hannan [7] and [8]), factor analysis (Lawley [8, chapter 2] and Anderson [1]) and statistical decision theory (Ferguson [3, chapter 4]).

In literature as indicated above the formulation and the treatment of the identification problem varies with the field of application. The aim of this paper is:
- to develop a concept of identification for the general model of (non-sequential) inferential statistics (sections 2 and 3),
- to apply the general theory to some useful particular models in univariate linear regression and simultaneous equation systems and to show that in these cases the well-known results on identification still hold under very weak conditions with respect to the parameterspace (sections 4 and 5),
- to pay attention to some new identification problems in models with lagged variables and autocorrelation, suggested by the general approach (section 6).

2. Basic ideas.

In (non-sequential) inferential statistics the set of data is considered to be an outcome of some random vector $\underline{y}$ (the sample), which has possible values in a set Y (the sample space) and is measurable with respect to some σ-field $\mathcal{B}$ of subsets of Y. Of the true probability distribution of $\underline{y}$ it is only known that it belongs to some given (non-empty) class $\mathcal{P}$ of probability distributions on (Y, $\mathcal{B}$).

A statistical conclusion indicates the true probability distribution or, more general , the true function value of some given function α on the class $\mathcal{P}$ (e.g. mean, variance). The statistician should reach his conclusion by means of a statistical procedure which is based on the properties of the class $\mathcal{P}$ and not on something else, because only the elements of $\mathcal{P}$ influence the outcome of the sample.

In most statistical problems the class $\mathcal{P}$ of all possible sample distributions is not primarily given but is thought to be generated by some parameter, chosen in some natural and simple way. More precisely, there is given a mapping P from some known space θ (the parameter space) into some given class of probability distributions on (Y, $\mathcal{B}$). The range of this mapping gives the class $\mathcal{P}$ of all possible sample distributions. If P_θ denotes the image of $\theta \in \theta$ under P then we can shortly write $\mathcal{P} = \{P_\theta, \theta \in \theta\}$. Now, a statistical conclusion can be considered to be an indication of the true parameter value, or more general, of the true function value of some given function on the space θ.

The parametric formulation is very attractive in practice because of the direct interpretation of the parameter. However, it can introduce the problem of identification. Let ψ be a function defined on θ. If ψ coincides with some function α on $\mathcal{P}$ ($\psi(\theta) = \alpha(P_\theta)$ for all $\theta \in \theta$ for some α) then there is no problem at all. In this case we call ψ identifiable. However, if ψ is not identifiable there exist $\theta_1, \theta_2 \in \theta$ with $\psi(\theta_1) \neq \psi(\theta_2)$ and $P_{\theta_1} = P_{\theta_2}$. Now the statistician should never indicate the function value of ψ. If he does so, he discriminates between two values θ_1 en θ_2 with the same sample distribution. Then his conclusion cannot be based upon a procedure that only depends on the properties of the elements of $\mathcal{P}$. In particular, the identity

($\psi(\Theta) = \Theta$ for all $\Theta \in \theta$) is identifiable iff (if and only if) there exists an 1-1 correspondence between θ and $\mathcal{P}$(i.e. the inverse of P on $\mathcal{P}$exists).

Herewith, the relevance of the following definition of identification in inferential statistics is justified:

<u>Definition 1.1</u>. Let $\mathcal{P} = \{P_\Theta, \Theta \in \theta\}$ be a parametric class of probability distributions on a space $(Y, \mathcal{B})$. A function ψ defined on θ is called identifiable (with respect to $\mathcal{P}$) if for all $\Theta_1, \Theta_2 \in \theta$ we have:

$$\psi(\Theta_1) \neq \psi(\Theta_2) \Rightarrow P_{\Theta_1} \neq P_{\Theta_2}$$

In particular, the class $\mathcal{P}$ is called identifiable if the identity is identifiable.

Every function of an identifiable function is identifiable. A vectorial function is identifiable iff all its components are identifiable. If the parametric class $\mathcal{P}$ is identifiable then any function of its parameter is identifiable.

In testing problems the usual concept of identification is a particular case of definition 1.1. Consider the testing problem (H_0, H_1) with $H_0 \cup H_1 = \theta$ and $H_0 \cap H_1 = \emptyset$. If $\Theta_0 \in H_0$ and $\Theta_1 \in H_1$ implies $P_{\Theta_0} \neq P_{\Theta_1}$ then the testing problem is called identifiable. If (H_0, H_1) is not identifiable then there exist $\Theta_0 \in H_0$, $\Theta_1 \in H_1$ with $P_{\Theta_0} = P_{\Theta_1}$ and now the statistician should refuse any choice between H_0 and H_1. Identifiability of testing problems in this sense is equivalent to the identifiability of the indicator function of the set H_0 (or H_1).

In point-estimating problems any function, for which an unbiased estimator exists, is identifiable. This immediately follows from the fact that different expectations of an estimator imply different sample distributions. Of course, in general the converse statement is wrong.

We terminate this section with an example for illustrative purposes.

<u>Example</u>. Let $\underline{y} = (\underline{y}_1, \ldots, \underline{y}_n)$ be a random sample from the normal distribution $N(\mu, \sigma^2)$ with unknown mean μ and unknown variance σ^2. By specifying

$\theta = (\mu,\sigma)$ the distribution P_θ of $\underline{y}$ on $(\mathbb{R}_n, \mathcal{B}_n)$ is determined. If we choose $\theta = \{(\mu,\sigma): -\infty < \mu < \infty, \sigma \geq 0\}$ then μ and σ (more precisely the functions $\psi_1(\mu,\sigma) = \mu$ and $\psi_2(\mu,\sigma) = \sigma$) are identifiable. However, if we choose $\theta = \{(\mu,\sigma): -\infty < \mu, \sigma < \infty\}$ then μ is identifiable but σ is not.

3. Extensions.

In statistical analysis the data set is almost always finite. This leads to models containing a finite number of random variables. However, in studying asymptotic properties of statistical procedures or time series it is often desirable to enlarge those finite models in such a way that they contain infinitely many random variables. We give a general framework for this and discuss the concept of identification within this context.

Let $\{\underline{y}_t, t \in T\}$ be a random vector process with one-sided discrete time domain $T = \{1,2,\ldots\}$. Here $\underline{y}_t$ is interpreted as the observable random vector at "time" t. Measurability of events in terms of the process refers to the σ-field generated by all finite products of some fixed σ-field of subsets of the state space with itself. The set of data is supposed to be an outcome of the part of the process corresponding to the time epochs t = 1,....,n. In this way a sample $\underline{y}^{(n)} = (\underline{y}_1,\ldots,\underline{y}_n)$, a sample space Y_n and a σ-field of subsets $\mathcal{B}_n$ of Y_n are generated for each n = 1,2,...,

It is supposed that the unknown true probability distribution of the process $\{\underline{y}_t, t \in T\}$ belongs to some given parametric class $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. Then the probability distribution of the sample $\underline{y}^{(n)}$ belongs to the class $\mathcal{P}_n = \{P_\theta^{(n)}, \theta \in \Theta\}$, where $P^{(n)}$ is the marginal distribution of P_θ corresponding to the first n components.

Identifiability of a function ψ can be considered with respect to $\mathcal{P}$ as well as to $\mathcal{P}_n$, n = 1,2,.... . identification with respect to $\mathcal{P}_N$ implies that with respect to $\mathcal{P}$ and $\mathcal{P}_n$ for all n = N, N+1,... . From Kolmogorov's law it follows that identification with respect to $\mathcal{P}$ implies that with respect to $\mathcal{P}_N$ for some N sufficiently large.

From a statistical point of vieuw only identification with respect to the class $\mathcal{P}_n$, where n is the sample size, is relevant. It follows that results on identification with respect to $\mathcal{P}$ have not much value. This point is often ignored in literature dealing with identification in time series. (We will make this more precisely in section 6 in a special case).

In particular, consistent estimation does not imply identification. Let ψ be consistently estimated by the sequence $\underline{t}_n = t_n(\underline{y}^{(n)})$. Then $\psi(\theta_1) \neq \psi(\theta_2)$ implies $P_{\theta_1} \neq P_{\theta_2}$ and therefore ψ is identifiable with respect

to $\mathcal{P}$. However, without further information nothing can be said about identification with respect to $\mathcal{P}_n$ for some fixed sample size n.

In the foregoing the assumption that T is one-sided is not essential and extensions can be made for two-sides $T = \{\ldots-1, 0, 1,\ldots\}$ in an obvious way.

4. Univariate linear regression

In univariate linear regression with n observations (n = 1,2,...)
we start with

$$\underline{y} = X\beta + \underline{\epsilon} = \mu + \underline{\epsilon},$$

where $\underline{y}$ is the (observable) random n-vector with values taken by the dependent
variable, X is the (observable) non-random n×k-matrix taken by k explanatory
variables (k = 1,2,...), β an k-vector of regression coefficients and $\underline{\epsilon}$
a (non-observable) random n-vector of disturbances with expectation zero.
Therefore $\mu = X\beta$ is the expectation of $\underline{y}$.

We suppose that the distribution is not fully known and is characte-
rized by some parameter ζ with values in a given set Z (more formally:
there is given a mapping from Z into some given class of n-variate probability
distributions, all having expectation zero).

Furthermore, we assume that it is only known that β belongs to a
given subset K of $\mathbb{R}_k$. Then the distribution of the sample $\underline{y}$ can be characte-
rized by $\theta = (\beta,\zeta)$.

Finally, we assume that the distribution of $\underline{\epsilon}$ does not depend on
the regression coefficients, formally expressed by the choice $\theta = K \times Z$. This
leads to a parametric class $\mathcal{P} = \{P_\theta, \theta \in \theta\}$ of probability distributions
of the sample $\underline{y}$.

In this linear model we are interested in the identifiability
of the vector β of regression coefficients (i.e. the function $\beta(\theta) = \beta$)
or, more general, of some function ψ of this vector (i.e. the composed
function $\psi(\beta(\theta))$ where $\psi(\beta)$ is defined for $\beta \in K$).

The identifiability of the expectation vector μ (i.e. the function
$\mu(\beta) = X\beta$) is clear. Therefore any function ν of this vector (the composed
function $\nu(\mu(\beta(\theta)))$ where $\nu(\mu)$ is defined for $\nu \in L = \mu(K)$) is identifiable.
The following theorem states that only functions of this particular type
are identifiable.

Theorem 4.1. The function $\psi(\beta(\theta))$ is identifiable iff there exists a function ν from $L = \mu(K)$ onto $\Psi = \psi(K)$ such that $\nu(\mu(\beta)) = \psi(\beta)$ for all $\beta \in K$.

Proof. ("If"). Let $\theta_1, \theta_2 \in \theta$ with $\psi(\beta(\theta_1)) \neq \psi(\beta(\theta_2))$. Then $\nu(\mu(\beta(\theta_1))) \neq \nu(\mu(\beta(\theta_2)))$ and therefore $P_{\theta_1} \neq P_{\theta_2}$ since $\nu(\mu(\beta(\theta)))$ is identifiable. Then $\psi(\beta(\theta))$ is identifiable by definition.

("Only if"). It suffices to show that $\psi(\beta_1) \neq \psi(\beta_2)$ implies $\mu(\beta_1) \neq \mu(\beta_2)$ for all $\beta_1, \beta_2 \in K$. Therefore, let $\beta_1, \beta_2 \in K$ with $\psi(\beta_1) \neq \psi(\beta_2)$. Choose some $\zeta \in Z$ and let $\theta_1 = (\beta_1, \zeta)$, $\theta_2 = (\beta_2, \zeta)$. Then $\theta_1, \theta_2 \in K \times Z = \theta$ with $\beta_1 = \beta(\theta_1)$, $\beta_2 = \beta(\theta_2)$. Thus, $\psi(\beta(\theta_1)) \neq \psi(\beta(\theta_2))$ and $\zeta(\theta_1) = \zeta(\theta_2)$. Since $\psi(\beta(\theta))$ is identifiable it follows that $P_{\theta_1} \neq P_{\theta_2}$. Therefore $\mu(\beta(\theta_1)) \neq \mu(\beta(\theta_2))$ since the distribution P_θ is completely determined by $(\mu(\beta(\theta)), \zeta(\theta))$. This gives $\mu(\beta_1) \neq \mu(\beta_2)$.

Of course, results of the foregoing type are well-known in literature but are always stated with superfluous conditions (e.g. linear space K, linear function ψ, some specific covariancestructure of the disturbances).

Theorem 4.2. The function $\beta(\theta)$ is identifiable iff for every $\mu \in L$ the equation $\mu = X\beta$ has at most one (and therefore a unique) solution $\beta \in K$.

Proof. Follows from theorem 4.1 by taking $\psi(\beta) = \beta$.

Theorem 4.3. Sufficient for the identifiability of the function $\beta(\theta)$ is the non-collinearity condition $r(X) = k$. If K is open in $\mathbb{R}_k$ then this condition is also necessary.

Proof. We use theorem 4.2.

("Sufficient"). Let $\mu = X\beta_1 = X\beta_2 \in K$. Then $\beta_1 = \beta_2$ since $r(X) = k$.

("Necessary"). Let $\mu = X\beta_1$ with $\beta_1 \in K$. If $r(X) < k$ then there exists an $\beta_0 \neq 0$ with $X\beta_0 = 0$. Then $\mu = X(\beta_1 + \epsilon\beta_0) = X\beta_2$ with $\beta_2 = \beta_1 + \epsilon\beta_0$. For $\epsilon > 0$ sufficiently small we have $\beta_2 \in K$ since K is open.

From theorem 4.1 some more special results can be derived.

For arbitrary K it follows that a testing problem in terms of $\beta \in K$ is identifiable iff it is equivalent to a testing problem in terms

of $\mu \in L$.

Furthermore, let K be a linear subspace of $\mathbb{R}_k$. Then L is a linear subspace of $\mathbb{R}_n$. For a linear function $\psi(\beta) = D\beta$ a corresponding function ν of theorem 4.1 is necessarily linear. Therefore we can write $\nu(\mu) = A\mu$. It follows that the function $D\beta$ is identifiable iff for some A we have that $AX\beta = D\beta$ for all $\beta \in K$. In particular, for $K = \mathbb{R}_k$ this leads to the condition $AX = D$ for some A or, equivalently, to the classic condition that the rows of D are linear combinations of the rows of X (compare Scheffé [10, section 1.4]).

5. Simultaneous equation systems

Many models in linear regression can be reduced to the univariate model of section 4 as far as it concerns identification. As a start for simultaneous equation systems we mention the Zellner-model (see e.g. Zellner [15]), which contains n observations on m dependent and k explanatory variables. It has m equations of the form

$$\underline{y}_i = X_i \, \beta_i + \underline{\varepsilon}_i \qquad , \qquad i = 1,\ldots,m.$$

where in the ith equation $\underline{y}_i$ is the (observable) random n-vector with values taken by the ith dependent variable, X_i is an (observable) non-random $n \times k_i$-matrix by a subset of k_i explanatory variables, β_i an k_i-vector of regression coefficients, and $\underline{\varepsilon}_i$ a (non-observable) random n-vector of disturbances with expectation zero ($1 \le k \le \Sigma k_i$). This model can easily be written in the form of the linear model of section 4. The identification problem is reduced to that of the univariate linear model.

Such a reduction to the univariate case is no longer possible and new difficulties arise if on the right side of the ith equation other dependent variables appear. Then we get a simultaneous equation system of the form

$$(5.1) \qquad \underline{y}_i = X_i \, \beta_i + \underline{Y}_i \, \alpha_i + \underline{\varepsilon}_i = [\, X_i \ \underline{Y}_i\,]\begin{bmatrix}\beta_i \\ \alpha_i\end{bmatrix} + \underline{\varepsilon}_i = \underline{W}_i \, \delta_i + \underline{\varepsilon}_i, \ i = 1,\ldots,m$$

where $\underline{Y}_i$ is the (observable) random $n \times m_i$-matrix with values taken by a subset of m_i dependent variables and α_i is an m_i-vector of regression coefficients ($1 \le m \le \Sigma m_i$). We set $n_i = k_i + m_i$, $N = \Sigma n_i$ (with $n_i \ge 1$) and $\delta = (\delta_1',\ldots,\delta_m')'$, the N-vector of all regression coefficients.

The structural equations (5.1) can be written as

$$(5.2) \qquad \underline{Y}A + XB + \underline{E} = 0 \qquad ,$$

where $\underline{Y} = [\,\underline{y}_1,\ldots,\underline{y}_m\,]$, $\underline{E} = [\,\underline{\varepsilon}_1,\ldots,\underline{\varepsilon}_m\,]$, $X = [\,x_1,\ldots,x_k\,]$, A an $m \times m$-matrix containing diagonal elements -1; the regression coefficients of the α_i

and zero's otherwise, B an k×m-matrix containing the regression coefficients of the β_i and zero's otherwise. Since the dependent variables should be functions of the explanatory variables and the disturbances we have to assume $\det(A) \neq 0$ (completeness of the system).

In the complete system of (5.2) the reduced form becomes

$$(5.3) \qquad \underline{Y} = X\Pi + \underline{E}^* = M + \underline{E}^*$$

where $\Pi = - BA^{-1}$ and $\underline{E}^* = - \underline{E}A^{-1}$. Therefore $M = X\Pi$ is the expectation of $\underline{Y}$. Note that A, B and Π are completely determined by δ.

We proceed as in the univariate linear model of section 4. We suppose that the distribution of $\underline{E}$ is characterized by some parameter ζ with values in a given set Z.

Furthermore, we assume that it is only known that δ belongs to a given subset Δ of $\mathbb{R}_N$ such that every δ-value generates a complete system $(\det(A(\delta)) \neq 0$ for all $\delta \in \Delta)$. Then the distribution of the samply $\underline{Y}$ can be characterized by $\Theta = (\delta, \zeta)$.

Finally, we assume that the distribution of $\underline{E}$ does not depend on the regression coefficients ($\theta = \Delta \times Z$). This leads to a parametric class $\mathcal{P} = \{P_\Theta, \Theta \in \theta\}$ of probability distributions of the sample $\underline{Y}$.

In this model we are interested in the identifiability of some function ψ of the vector δ (i.e. the composed function $\psi(\delta(\Theta))$). As in the univariate case the identifiability of a function ν of the expectation M (i.e. the function $M(\delta) = X\Pi(\delta)$) is clear. However, the converse in the sense of theorem 4.1 no longer holds in general. The main point is that the distribution of $\underline{E}^*$ in (5.3) depends on δ and therefore can help with identification. This does not take place in cases where the system is closed under non-singular linear transformations. We proceed with such cases.

Definition 5.1. The system (5.1) or (5.2) is closed (under non-singular linear transformations) if the class of all distributions of $\underline{E}$ equals the class of all distributions of $\underline{E}T$ for every non-singular m×m-matrix T.

Theorem 5.1. Sufficient for the identifiability of the function $\psi(\delta(\Theta))$ is that there exists a function ν from $L = M(\Delta)$ onto $\Psi = \psi(\Delta)$ such that $\nu(M(\delta)) = \psi(\delta)$ for all $\delta \in \Delta$. If the system is closed then this condition is also necessary.

Proof. ("Sufficient"). Compare the "if" part of the proof of theorem 4.1. ("Necessary"). It suffices to show that $\psi(\delta_1) \neq \psi(\delta_2)$ implies that $M(\delta_1) \neq M(\delta_2)$ for all $\delta_1, \delta_2 \in \Delta$. Therefore let $\delta_1, \delta_2 \in \Delta$ with $\psi(\delta_1) \neq \psi(\delta_2)$. Choose some distribution of $\underline{E}$ corresponding to $\zeta_1 \in Z$. Since the system is closed there exists an $\zeta_2 \in Z$ such that

$$\text{Distr.}(\underline{E}|\zeta_2) = \text{Distr.}(\underline{E}.A^{-1}(\delta_1).A(\delta_2)|\zeta_1)$$

or, equivalently,

$$\text{Distr.}(\underline{E}^*|\Theta_1) = \text{Distr.}(-\underline{E}\,A^{-1}(\delta_1)|\zeta_1) = \text{Distr.}(-\underline{E}\,A^{-1}(\delta_2)|\zeta_2) =$$

$$= \text{Distr.}(\underline{E}^*|\Theta_2)$$

where $\Theta_1 = (\delta_1, \zeta_1)$, $\Theta_2 = (\delta_2, \zeta_2) \in \Delta \times Z = \Theta$. Furthermore, since $\psi(\delta(\Theta))$ is identifiable it follows from

$$\psi(\delta(\Theta_1)) = \psi(\delta_1) \neq \psi(\delta_2) = \psi(\delta(\Theta_2))$$

that $P_{\Theta_1} \neq P_{\Theta_2}$. Therefore, $M(\delta(\Theta_1)) \neq M(\delta(\Theta_2))$ since the distribution P_Θ of $\underline{Y}$ is completely determined by $M(\delta(\Theta))$ and $\text{Distr.}(\underline{E}^*|\Theta)$. This gives $M(\delta_1) \neq M(\delta_2)$.

From theorem 5.1 some special cases can be derived. We consider the identifiablility of $\delta(\Theta)$ or, equivalently, $A(\delta(\Theta))$ and $B(\delta(\Theta))$ (the identifiability of the structural form).

Theorem 5.2. (identifiability of the structural form).
Sufficient for the identifiability of $\delta(\Theta)$ is that for every $M \in L$ the equation

$$M = X\Pi(\delta) = - XB(\delta)A^{-1}(\delta)$$

has at most one (and therefore á unique) solution $\delta \in \Delta$. If the system is closed then this condition is also necessary.

Proof. The result immediately follows from $M = X\Pi$, $\Pi = - BA^{-1}$ and theorem 5.1 by taking $\psi(\delta) = \delta$.

It is not easy to verify the condition in theorem 5.2. Therefore we give another one that can be verified straightforwards from the structural form. We have to introduce some new notations (compare Theil [13, section 10.2]).

For fixed i we think (merely for sake of convenience) the variables in (5.2) and (5.3) ordered in such a way that the variables in the ith equation of (5.1) are preceeding the other ones. Furthermore, we think the dependent variable on the left side of this equation to be the first one. Then we can split up (5.2) and (5.3) in a corresponding way (not interesting submatrices are denoted by stars)

$$(5.2)' \quad [\underline{y}_i \quad \underline{Y}_i \quad *] \begin{bmatrix} -1 & * \\ \alpha_i & * \\ 0 & A_i \end{bmatrix} + [X_i \quad X_i^*] \begin{bmatrix} \beta_i & * \\ 0 & B_i \end{bmatrix} + \underline{E} = 0$$

$$(5.3)' \quad [\underline{y}_i \quad \underline{Y}_i \quad *] = [X_i \quad X_i^*] \begin{bmatrix} \pi_i & \Pi_i & * \\ \pi_i^* & \Pi_i^* & \Pi_i^{**} \end{bmatrix} + \underline{E}^* \quad ,$$

with the obvious interpretations if $m_i = 0$, $m_i = m-1$, $k_i = 0$ or $k_i = k$.

Comparing (5.2)' and (5.3)' it follows that the relationship $\Pi A = -B$ is equivalent to

$$\begin{bmatrix} \pi_i & \Pi_i \\ \pi_i^* & \Pi_i^* \end{bmatrix} \begin{bmatrix} -1 \\ \alpha_i \end{bmatrix} = \begin{bmatrix} -\beta_i \\ 0 \end{bmatrix} \quad \text{for all } i = 1,\ldots,m,$$

or

$$(5.4) \quad \left. \begin{array}{l} \pi_i^* = \Pi_i^* \, \alpha_i \\[2ex] \beta_i = \pi_i - \Pi_i \, \alpha_i \end{array} \right\} \quad i = 1, \ldots, m$$

Furthermore, it follows from $\Pi = -BA^{-1}$ and $-I = -AA^{-1}$ that

$$\begin{bmatrix} \pi_i^* & \Pi_i^* & \Pi_i^{**} \\[2ex] 0 & 0 & -I \end{bmatrix} = \begin{bmatrix} 0 & -B_i \\[2ex] 0 & -A_i \end{bmatrix} \cdot A^{-1}$$

and with the first equation of (5.4) this leads to

$$(5.5) \quad r(C_i) = r(\Pi_i^*) + (m - m_i - 1),$$

where $C_i' = [\, A_i' \quad B_i' \,]$.

      With the relationships (5.4) and (5.5) we are in the position to prove the following theorem:

Theorem 5.3. (rank condition for the structural form)

Sufficient for the identifiability of the function $\delta(\theta)$ are the conditions $r(X) = k$ and $r(C_i(\delta)) = m-1$ for all $i = 1, \ldots, m$ and all $\delta \in \Delta$. If the system is closed and Δ is open in $\mathbb{R}_N$ then these conditions are also necessary.

Proof. According to (5.5) we can replace the conditions of the theorem by $r(X) = k$ and $r(\Pi_i^*(\delta)) = m_i$ for all $i = 1, \ldots, m$ and $\delta \in \Delta$. ("Sufficient"). From the sufficient part of theorem 5.2 we see that we have to prove that for every $M \in L$ the equation $M = X\Pi(\delta)$ has at most one solution $\delta \in \Delta$. Let $M = X\Pi(\delta_1) = X\Pi(\delta_2)$ with $\delta_1, \delta_2 \in \Delta$. Since $r(X) = k$ this gives $\Pi(\delta_1) = \Pi(\delta_2)$. Denote by δ_{i1} or α_{i1}, β_{i1} the ith part of δ_1 (the part corresponding to the ith equation). Similarly for δ_2. Since $r(\Pi_i^*(\delta)) = m_i$ for all $\delta \in \Delta$ we get from (5.4) that $\alpha_{i1} = \alpha_{i2}$, $\beta_{i1} = \beta_{i2}$ for all i, or $\delta_1 = \delta_2$.

("Neccessary"). From the necessary part of theorem 5.2 we see that we have to prove that for some $M \in L$ the equation $M = X\Pi(\delta)$ has at least two solutions $\delta \in \Delta$ if the conditions fail.

At first, suppose $r(\Pi_i^*(\delta_1)) < m_i$ for some i and some $\delta_1 \in \Delta$. Then there exists an $s \neq 0$ with $\Pi_i^*(\delta_1).s = 0$. So, for fixed $\Pi(\delta_1)$ the equations (5.4) are satisfied for the components α_{i1}, β_{i1} of δ_1 as well as for

$$\alpha_{i2} = \alpha_{i1} + \varepsilon.s \quad , \quad \beta_{i2} = \beta_{i1} - \varepsilon. \Pi_i(\delta_1).s$$

Replace in δ_1 the ith part α_{i1}, β_{i1} by α_{i2}, β_{i2} and denote this vector by δ_2. Then $\Pi(\delta_1) = \Pi(\delta_2)$ and $\delta_2 \in \Delta$ for $\varepsilon > 0$ sufficiently small since Δ is open. This implies $M = X\Pi(\delta_1) = X\Pi(\delta_2)$ with $\delta_1 \neq \delta_2$ and $\delta_1, \delta_2 \in \Delta$. Finally, suppose $r(X) < k$ and $r(\Pi_i^*(\delta)) = m_i$ for all $i = 1,\ldots,m$ and all $\delta \in \Delta$. Take some $M \in L$. Then we can write $M = X\Pi(\delta_1)$ with $\delta_1 \in \Delta$. Since $r(X) < k$ there exists an $c \neq 0$ with $0 = Xc = [X_i \quad X_i^*][c_i' \quad c_i'^*]'$. Take some fixed i and replace in δ_1 the ith part α_{i1}, β_{i1} by

$$\alpha_{i2} = \alpha_{i1} + \varepsilon\{\Pi_i^{'*}(\delta_1)\Pi_i^*(\delta_1)\}^{-1} c_i^*, \quad \beta_{i2} = \beta_{i1} + \varepsilon[c_i - \Pi_i(\delta_1).\{\Pi_i^{'*}(\delta_1)\Pi_i^*(\delta_1)\}^{-1} c_i^*]$$

Denote this vector by δ_2. Then $\delta_2 \in \Delta$ for $\varepsilon > 0$ sufficiently small since Δ is open. Furthermore, $\Pi(\delta_2) = \Pi(\delta_1) + \varepsilon\Pi_0$ where Π_0 is a matrix with i^{th} column equal to c and with elements zero elsewhere. This gives $M = X\Pi(\delta_1) = X(\Pi(\delta_1) + \varepsilon\Pi_0) = X\Pi(\delta_2)$ with $\delta_1 \neq \delta_2$ and $\delta_1, \delta_2 \in \Delta$.

In practice without further knowlegde of the regression coefficients one would like to take $\Delta = \mathbb{R}_N$. However, due to the completeness and identifiability conditions almost always closed subsets of lower dimensions of $\mathbb{R}_N$ have to be excluded. The condition in theorem 5.3 that Δ is open in $\mathbb{R}_N$ matches perfectly to this situation.

In literature in this field as mentioned in the introduction a somewhat different approach is followed. In the foregoing situation one takes for Δ some set such that $\mathbb{R}_N - \Delta$ has Lebesgue (or some other) measure zero. Then with the general definition 1.1 of identification the theorem 5.3 needs no longer to be true for some particular choice of Δ. Therefore if one wishes that this theorem maintains to hold the

definition of identification has to be changed. From an axiomatic point
of view this seems to be unsatisfactory. However, approaches of that kind
has been made in literature. For an overview see e.g. Schönfeld [12] .

By the way, another argument that makes open sets Δ attractive
comes from point estimation. Usual estimators for δ (e.g. 2SLS, 3SLS) can
take values outside Δ and therefore, at least formally, should be defined
there in some other way. However, if Δ is open then such changes have no
influence on asymptotic properties (e.g. consistency, asymptotic normality)
and therefore in such cases one can be a little careless in the definitions.

6. Linear models with lagged variables.

In univariate linear regression with lagged variables we start with a random process $\{\underline{y}_t,\ t \in T\}$ with $T = \{\dots-1,0,1,\dots\}$ which satisfies for all $t \in T$ the difference equation (convergence in mean square):

$$(6.1) \qquad \sum_{g=0}^{\infty} \alpha_g \underline{y}_{t-g} = \beta' x_t + \underline{\varepsilon}_t \qquad , \qquad \alpha_0 = 1$$

with

$$\underline{\varepsilon}_t = \sum_{h=0}^{\infty} \gamma_h \underline{n}_{t-h} \qquad , \qquad \gamma_0 = 1,$$

where for the tth-period $\underline{y}_t$ is the (observable) random value of the dependent variable, x_t is an (observable) non-random k-vector taken by the explanatory variables and $\underline{\varepsilon}_t$ is a (non-observable) disturbance generated by a (weakly) white noise process $\{\underline{n}_t,\ t \in T\}$ with expectation zero.

We want to define $\{\underline{y}_t\}$ as the (a.s) uniquely determined solution of (6.1). Therefore we restrict ourselves to covariance stationary processes $\{\underline{y}_t\}$ with non-exponentially increasing expectation μ_t for $t \to -\infty$. If the power series $A(z)=\sum_0^{\infty} \alpha_g z^g$, $C(z) = \sum_0^{\infty} \gamma_h z^h$ have convergence radii larger than 1 with $A(z) \neq 0$ for all $|z| \leq 1$ and if x_t does not increase exponentially fast for $t \to -\infty$ then this uniqueness is guaranteed. Furthermore, the solution has the one-sided representation

$$\underline{y}_t = \beta' \sum_0^{\infty} \psi_r x_{t-r} + \sum_0^{\infty} \varphi_r \underline{n}_{t-r}$$

where

$$\phi(z) = \sum_0^{\infty} \varphi_r z^r = C(z)/A(z), \Psi(r) = \sum_0^{\infty} \psi_r z^r = 1/A(z),\ |z| \leq 1.$$

The foregoing conditions are rather strong but always satisfied in practice. Since our aim is identification we feel no need to weaken them.

We suppose that the distribution of $\{\underline{n}_t\}$ is not fully known and is characterized by some parameter ζ with values in Z. Furthermore, we assume that $\alpha = (\alpha_1, \alpha_2, \dots) \in A$, $\beta \in K$, $\gamma = (\gamma_1, \gamma_2, \dots) \in C$ such that for

every $\alpha \in A$ and $\gamma \in C$ the corresponding $A(z)$ and $C(z)$ have convergence radii larger than 1 with $A(z) \neq 0$ for all $|z| \leq 1$. Then with $\theta = (\alpha, \beta, \gamma, \zeta)$ there corresponds a distribution P_θ of $\{\underline{y}_t, t \in T\}$, uniquely determined by (2.1), in the class of all covariance stationary distributions with non-exponentially increasing expectation for $t \to -\infty$. This leads to a parametric class $\mathcal{P} = \{P_\theta, \theta \in \theta\}$ of probability distributions of the process $\{\underline{y}_t, t \in T\}$ if we take $\theta = A \times B \times C \times Z$.

Finally, we suppose that the process $\{\underline{y}_t, t \in T\}$ is actually observed in the time period $\{1, \ldots, n\}$. Then $\underline{y}^{(n)} = (\underline{y}_1, \ldots, \underline{y}_n)$ has to be considered as the sample. This leads to the parametric class $\mathcal{P}_n = \{P_\theta^{(n)}, \theta \in \theta\}$ of distributions of $\underline{y}^{(n)}$, where $P_\theta^{(n)}$ is the marginal distribution of P_θ with respect to the time period $\{1, \ldots, n\}$.

Questions concerning the identifiability of functions of the parameter with respect to the class $\mathcal{P}$ have been discussed by Hannan [5,6] for some related cases. As pointed out in section 3, from a statistical point of view the identifiability with respect to the class $\mathcal{P}_n$ is more interesting. Problems of that kind are hardly treated in literature. The only paper known to me is Tigelaar [14], which contains some results for the case of finite sums. It would be valuable to extend these results and to generalize them to multivariate difference equations and, even more general, to simultaneous equation systems with lagged variables.

References

[1]   Anderson, T.W. and Rubin, H. (1956),
      Statistical inference in factor analysis, Proc. 3rd Berkeley Symp.
      Math. Stat. Prob. 5, 111-150.

[2]   Bose, R.C. (1944),
      The fundamental theorem of linear estimation, Proc. 31th Indian
      Sci. Congress (abstract), 2-3.

[3]   Ferguson, T.S. (1967),
      Mathematical statistics: a decision theoretic approach,
      Academic Press (New York).

[4]   Fisher, F.M. (1966),
      The identification problem in econometrics, Mc Graw Hill (New York).

[5]   Hannan, E.J. (1969),
      The identification of vector mixed autoregressive-moving average systems,
      Biometrika 56, 223-225.

[6]   Hannan, E,J, (1971),
      The identification problem for multiple equation systems with moving
      average errors, Econometrics 39, 751-756.

[7]   Koopmans, T.C., Rubin, H., Leipnik, R.B. (1950),
      Measuring the equation systems of dynamic economics, in: Koopmans, T.C.
      e.d, Statistical inference in dynamic economic models, Cowles
      Commission Monograph 10, Wiley (New York).

[8]   Lawley, D.N. and Maxwell, A.E. (1971),
      Factor analysis as a statistical method, 2nd edition
      Butterworths (London).

[9] Malinvaud, E. (1970),
     Statistical methods of econometrics,
     North-Holland Publ. Comp. (Amsterdam).

[10] Scheffé, H. (1959),
     The analysis of variance,
     Wiley (New York)

[11] Schönfeld, P. (1971),
     Methoden der Ökonometrie, Band 2,
     Verlag Franz Vahlen (München).

[12] Schönfeld, P. (1975),
     A survey of recent concepts of identification,
     Core discussion paper No. 7515, Catholic University (Leuven).

[13] Theil, H. (1971),
     Principles of econometrics,
     Wiley (New York).

[14] Tigelaar, H.H. (1976),
     Identifiability in models with lagged variables,
     Series "Ter Discussie" No. 76.020, Catholic University (Tilburg).

[15] Zellner, A. (1962),
     An efficient method of estimating seemingly unrelated regressions
     and tests for aggregation bias",
     Amer. Stat. Assoc. J. 57, 348-368.

PREVIOUS NUMBERS:

EIT 46 TH. van de Klundert On shift and share of durable
      A van Schaik capital.

EIT 47 G.R. Mustert The development of income
      distribution in the Netherlands
      after the second world war.

EIT 48 H. Peer The growth of the labor-management
      in a private economy.

EIT 49 J.J.M. Evers On the initial state vector in
      linear infinite horizon
      programming.

EIT 50 J.J.M. Evers Optimization in normed vector
      spaces with applications to
      optimal economic growth theory.

EIT 51 J.J.M. Evers On the existence of balanced
      solutions in optimal economic
      growth and investment problems.

EIT 52 B.B. van der Genugten An (s,S)-inventory system with
      exponentially distributed lead
      times.

EIT 53 H.N. Weddepohl Partial equilibrium in a market
      in the case of increasing
      returns and selling costs.

EIT 54 J.J.M. Evers A duality theory for convex
      ∞-horizon programming.

EIT 55 J. Dohmen Approximated fixed points.
      J. Schoeber

EIT 56 J.J.M. Evers Invariant competitive equili-
      brium in a dynamic economy
      with negotiable shares.

EIT 57 W.M. van den Goorbergh Some calculations in a three-
      sector model.

EIT 58 W.G. van Hulst Investment/financial planning
      J.Th. van Lieshout with endogeneous lifetimes:
      a heuristic approach to mixed-
      integer programming.

FEW 59 J.J.M. Evers A dynamic economy with shares,
      M. Shubik fiat, bank and accounting money.

FEW 60 J.M.G. Frijns A dynamic model of factor
      demand equations.