Proceedings of the 30th Conference
on Decision and Control
Brighton, England • December 1991

IDENTIFICATION OF DYNAMIC SYSTEMS FROM NOISY DATA: THE CASE $M^* = N - 1$[1]

B.D.O. Anderson[2], M Deistler[3]

[1] Support by the Austrian "Fonds zur Förderung der wissenshaftlichen Forshchung" "Schwerpunkt Angewandte Mathematik" (S32/02) is gratefully acknowledged.

[2] Dept of Systems Eng., RSPhysSE, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, AUSTRALIA
[3] University of Technology Vienna, Institute of Econometrics, Operations Research and Systems Theory, Argentinierstraße 8, A-1040 Vienna, AUSTRIA

ABSTRACT:
Linear dynamic errors-in-variables (or factor-) models in the framework of stationary processes are considered. The noise process is assumed to have a diagonal spectral density. We analyse the relation between the (population) second moments of the observations and the system- and noise characteristics; of particular interest are the number of equations (or the number of factors) and a description of the set of all systems compatible with the second moments of the observations. In the present paper emphasis is put on the case which can be reduced to a single factor. The problems considered arise in the context of identification and precede estimation.

Key words: identification, errors-in-variables, linear systems

1 Introduction

In identification of linear systems the "main stream" approach to noise modelling is to add all noise to the outputs (assuming orthogonality), or to the equations (which is the same for our analysis). In econometrics these models are named errors-in-equation models. Here we are concerned with the case where in principle all variables may be contaminated by noise. Such models are called *errors-in-variables* (EV) or *latent variables models*, or using a slightly different but equivalent formulation *factor models*. They have been analysed and used in econometrics, psychometrics, statistics and system engineering (see e.g. [1]-[7]).

EV modelling is appropriate for instance:

(i) If we are interested in the true system generating the data (rather than in prediction or in encoding the data by system parameters) and we cannot be sure a priori that the observed inputs are not corrupted by noise.

(ii) If we want to approximate a high dimensional data vector by a relatively small number of factors.

(iii) If we have no sufficient a priori information about the number of equations in the system or about the classification of the variables into inputs and outputs; then we have to perform a more symmetric system modelling, which in turn demands a more symmetric noise model.

The system considered is of the form

$$w(z)\hat{z}_t = 0 \tag{1.1}$$

where $\hat{z}_t$ is the n-dimensional vector of latent (i.e. in general unobserved) variables, z is used for the backward-shift on $\mathbb{Z}$ (i.e. $z(\hat{z}_t \mid t \in \mathbb{Z}) = (\hat{z}_{t-1} \mid t \in \mathbb{Z})$) as well as for a complex variable and

$$w(z) = \sum_{j=-\infty}^{\infty} W_j z^j \quad ; \quad W_j \in \mathbb{R}^{m\times n} \tag{1.2}$$

We will call $w(z)$ the *relation function* of the exact relation (1.1) (compare [8]). Clearly, systems of the form (1.1) are *symmetric* in the sense that we need no a priori classification of the variables $\hat{z}_t$ into inputs and outputs and no a priori information about causality directions; without restriction of generality we will assume that $m \leq n$ holds and that $w(z)$ contains no linearly dependent rows; and in general m is not known a priori.

The observed variables are of the form

$$z_t = \hat{z}_t + u_t \tag{1.3}$$

where u_t is the noise vector.

Throughout the paper we will assume:

(i) $(z_t), (\hat{z}_t)$ and (u_t) respectively are [wide sense] stationary processes [with real valued components and] with spectral densities $\Sigma, \hat{\Sigma}$ and D respectively. [In addition limits of random variables are understood in the sense of mean square convergence].

(ii) $E\hat{z}_t = Eu_t = 0$

(iii) $E\hat{z}_t u_s' = 0$
and finally

(iv) D is diagonal

For a discussion of assumption (iv) see e.g. [7] and [9]. In our analysis, unless the contrary is stated explicitly, the frequency λ will be kept fixed (i.e. we study (1.1) with $z = \exp i\lambda$). In this sense $\Sigma, \hat{\Sigma}$ and D are considered as (constant) Hermitian matrices rather than as spectral densities. From (1.3) we have

$$\Sigma = \hat{\Sigma} + D \tag{1.4}$$

Clearly (1.4) may also be interpreted as coming from a (static) relation between $\mathbb{C}^n$-valued random ¡variables $z, \hat{z}$ and u of the form

$$\begin{aligned} z &= \hat{z} + u \; ; \; W\hat{z} \; = \; 0, \; W \in \mathbb{C}^{m\times n} \\ \Sigma &= Ezz^* \; ; \; \hat{\Sigma} \; = \; E\hat{z}\hat{z}^* \; ; \; D \; = \; uu^* \end{aligned} \tag{1.5}$$

where $*$ denotes the conjugate transpose.

In this paper we will analyse the relation between the second moments of the observations Σ and the system and noise characteristics $w(z)$ and D. Such an analysis is a necessary first step for an analysis of the properties of estimation and inference procedures. The main questions are (compare Deistler and Anderson 1989)

(a) Find the maximum number, m^* say, of (linearly independent) rows of $w(z)$ among the set of all $w(z)$ compatible with given Σ. Sometimes we also use the symbol $mc(\Sigma)$ for m^* if we want to make the dependence of Σ explicit.

(b) Give a description of the set of all $(w(z), D)$ compatible with given Σ; in addition describe the subsets corresponding to different numbers of linear relations m.

(c) Describe the set of all Σ corresponding to a given m^*, $n > m^* \geq 1$.

Thus the problems we consider are (a) to find the (maximum) number of equations for given Σ, (b) to describe the set of all observationally equivalent (based on second moments only) signal and noise characteristics and (c) to describe the set of spectral densities corresponding to a given m^*.

There is no general solution available for these problems up to now. In this paper, the main emphasis will be on the case of general n and $m^* = n-1$, i.e. on the case which can be modelled by a single factor. (For the "opposite" case, namely the one equation case, i.e. $m^* = 1$, see [10]).

For the static case, where $\hat{z}_t$ and u_t are (real) white noise processes (and thus $\Sigma(\lambda)$ and $\hat{\Sigma}(\lambda)$ are constant with real entries) and where $w(z)$ is constant with real entries, this problem has a long history, beginning with the work of Charles Spearman, see e.g. [11], [12], and [6]. For the dynamic case see [1] and [2]. The cases $n = 2$ and $n = 3$ have been treated in detail in [4],[5].

The paper is organized as follows: In section 2 we are concerned with question (b) above, without giving a complete answer. Thereby no restriction on m^* is imposed. In sections 3 and 4 we give a rather complete analysis of the case $m^* = n-1$: We characterize the spectral densities Σ corresponding to $m^* = n-1$ and we give a complete answer to (b) for this case.

Throughout we use the following notation: If A denotes a matrix, $\mathrm{rk}(A)$ and $\ker(A)$ will denote its rank and its kernel respectively; by $\dim_{\mathbb{R}}$ and $\dim_{\mathbb{C}}$ respectively we denote real and complex dimension.

2 The Solution Set – Some General Properties

Clearly relation (1.1) implies

$$w(e^{i\lambda}).\hat{\Sigma}(\lambda) = 0 \tag{2.1}$$

If $\hat{\Sigma}$ were known and if we want to explain by the system as much as possible [in the sense that for given $\hat{\Sigma}, m$ is chosen as large as possible] and if we have no additional a priori information, then by (2.1), the rows of w are defined as a basis of the left kernel of $\hat{\Sigma}$; thus w is unique up to basis change.

Clearly in general only Σ is known and thus equation (1.4) will be the starting point of our analysis. Remember that $\Sigma, \hat{\Sigma}$ and D are nonnegative definite and that $\hat{\Sigma}$ is singular and D is diagonal. In view of this, $\hat{\Sigma}$ and D are called *feasible* for given Σ if

$$0 \leq \Sigma - D \leq \Sigma \tag{2.2}$$

holds, where $\Sigma - D = \hat{\Sigma}$ is singular and D is diagonal. As easily can be shown, for every $\Sigma \geq 0$ a feasible decomposition (1.4) exists and to every feasible decomposition a corresponding EV representation exists. To avoid having to consider a number of special cases we will assume unless the contrary is stated explicitly that

(v) $\Sigma > 0$

(vi) $\sigma_{ij} \neq 0 \quad i,j = 1 \cdots, n$
and

(vii) $\bar{s}_{ij} \neq 0 \quad i,j = 1 \cdots, n$
hold.

Here $\bar{S} = \Sigma^{-1}$ and as a general rule if e.g. Σ is a matrix, its i,j-entry is denoted by the corresponding lower case symbol σ_{ij}.

For given Σ, a vector $\bar{x} \in \mathbf{C}^n$ is called a *solution* if there exists a feasible $\hat{\Sigma}$ satisfying

$$\bar{x}\hat{\Sigma} = 0 \tag{2.3}$$

The set of all solutions corresponding to a given Σ is called the *solution set* $\bar{L}$(of Σ). Analogously we define $\mathcal{D}$ as the set of all feasible matrices D corresponding to Σ. Since $\bar{L}$ is the union of linear spaces of dimension greater than zero, we may find a normalization useful. In most parts of the paper, the first component of $\bar{x}, \bar{x}_1$ is normalized to one. In view of this, we define the *(normalized) solution set*

$$L = \{x \in \mathbf{C}^{n-1} \mid (1,x) \in \bar{L}\}$$

Let us define the matrix $S = (\bar{s}_{ij} \cdot \bar{s}_{i1}^{-1})$, $i,j = 1, \cdots, n$ and let s_j denote the j-th row of S. Now it is easily seen from

$$s_j \cdot \Sigma = (0, \cdots, 0, \bar{s}_{j1}^{-1}, 0, \cdots, 0) = s_j D$$

that s_j is the solution (with first component normalized to one) corresponding to the j-th *elementary regression*. i.e. to the case where all components of z_t, except for the j-th, are assumed to be observed free of noise; the noise covariance matrix is of the form $D_j = \mathrm{diag}\{0, \cdots, 0, \ d_{jj}^{(j)}, \ 0, \cdots, 0\}$. The vector s_j will be called the j-th *elementary solution*. Since the first elementary solution s_1 always exists, no matrix Σ is excluded by the normalization $\bar{x}_1 = 1$. However, the kernel of $\hat{\Sigma}$ may be orthogonal to $(1, 0, \cdots, 0)$ and in this sense the normalization may be a restriction of generality. This situation can not occur in the case $m^* = n-1$ since (vi) implies that every row of the corresponding $\hat{\Sigma}$ (which has rank 1) can be expressed as a linear combination of every other row. Clearly elementary solutions can also be defined for singular matrices Σ. They correspond to the projection of the j-th component of z in (1.5) on the space spanned by all other components. For the following see also [10].

Now, let us state some useful lemmas.

Lemma 2.1 *Let $\Sigma \geq 0$ (may be singular). If the n-th row of Σ, σ_n say, is linearly independent from the other rows $\sigma_1, \cdots, \sigma_{n-1}$ of Σ, then the n-th elementary regression gives a noise covariance matrix D of the form*

$$D_n = \mathrm{diag}\{0, \cdots, 0, \ d_{nn}^{(n)}\} \tag{2.4}$$

where $d_{nn}^{(n)} > 0$ and where $\mathrm{rk}(\Sigma - D_n) = \mathrm{rk}(\Sigma) - 1$ holds. If σ_n is linearly dependent on the other rows, then the noise covariance matrix of the n-th elementary regression is zero.

Proof The proof is straightforward and can be seen from projecting the n-th component of z (see 1.5) on the linear space spanned by the other components of z. ∎

Lemma 2.2 *Let $D = \mathrm{diag}\{d_{ii}\}$ be feasible and let $d_{ii}^{(i)}$ correspond to the i-th elementary regression. Then*

$$0 \leq d_{ii} \leq d_{ii}^{(i)} \tag{2.5}$$

If $d_{jj} > 0$ for some $j \neq 1$, then the second inequality in (2.5) is strict.

Proof Without restriction of generality, take $i = 1$; let $D = A + B$; where $A = \mathrm{diag}\{d_{11}, 0, \cdots, 0\}$ and $B = \mathrm{diag}\{0, d_{22}, \cdots, d_{nn}\}$. First note that for $d_{11} > d_{11}^{(1)}$, the matrix $\Sigma - A$ would not be nonnegative definite. To see this consider

$$\det(\Sigma - A) = (\sigma_{11} - d_{11}).f_1(\Sigma) + f_2(\Sigma) \tag{2.6}$$

where f_1 and f_2 depend only on Σ and where

$$f_1(\Sigma) = \det \begin{bmatrix} \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots \\ \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} > 0$$

holds. Thus $\det(\Sigma - A)$ is zero if and only if $d_{11} = d_{11}^{(1)}$; and it is negative for $d_{11} > d_{11}^{(1)}$.

Now $\Sigma - D = \Sigma - A - B = C \geq 0$ would imply $B + C = \Sigma - A \geq 0$ which is a contradiction for $d_{11} > d_{11}^{(1)}$.

To prove the last statement, perform the j-th elementary regression for $(\Sigma - D_i)$; since all elements of s_i are unequal to zero by (vii), and $s_i - (\Sigma - D_i) = 0$, the j-th row of $S - D_i$ is linearly dependent on the other rows of $(\Sigma - D_i)$. Hence by Lemma 2.1, and diagonal matrix $\triangle$ for which $0 \leq \triangle \leq \Sigma - D_i$ has $\triangle = 0$. Equivalently, the last statement of the lemma is proved. ∎

For fixed Σ, the relation between $\bar{L}$ and D is given by

$$\bar{x}\Sigma = \bar{x}D \ , \ \ \bar{x} \in \bar{L} \ , \ \ D \in \mathcal{D} \tag{2.7}$$

In order to further investigate the solution set, let us connect two points, $x, y \in \bar{L}$ by the complex line

$$\alpha x + (1-\alpha)y \ , \ \ \alpha \in \mathbf{C} \tag{2.8}$$

The question then is for which $\alpha, \alpha x+(1-\alpha)y \in \hat{L}$ holds. Note that $\alpha x + (1-\alpha)y, \alpha \in \mathbf{C}$ is a plane in $\mathbf{R}^{2n} \doteq \mathbf{C}^n$. The results obtained in this section are valid for general $mc(\Sigma)$. We start from the equation

$$(\alpha x + (1-\alpha)y)\Sigma = (\alpha x + (1-\alpha)y)D = \alpha x D_x + (1-\alpha)y D_y \quad (2.9)$$

where $x, y \in \hat{L}, x_1 = y_1 = 1$ and D_x and D_y correspond to x and y respectively; D is diagonal and the unknown variable in (2.9). Clearly $\alpha x + (1-\alpha)y \in \hat{L}$ if and only if there is a D satisfying (2.9), $D \geq 0$ and $\Sigma - D \geq 0$.

First consider the case $x = s_1, y = s_j, j > 1, D_x = D_1$ and $D_y = D_j$, i.e. we investigate the real plane given by the first and the j-th elementary solution. Then the first element in the second equation in (2.9) is of the form

$$\alpha d_{11}^{(1)} = d_{11} \quad (2.10)$$

and the j-th element gives

$$(1-\alpha)s_{jj}d_{jj}^{(j)} = (\alpha s_{1j} + (1-\alpha)s_{jj})d_{jj}$$

which gives

$$\frac{1}{1+\frac{\alpha}{1-\alpha}\frac{s_{1j}}{s_{jj}}} \cdot d_{jj}^{(j)} = d_{jj} \quad (2.11)$$

By Lemma 2, $d_{jj}^{(j)} \geq d_{jj} \geq 0$ must hold for every feasible D. Also note that $s_{1j} \cdot s_{jj}^{-1} > 0$; thus (2.10) and (2.11) imply $\alpha \in [0,1]$. Put

$$d_{ii} = 0 \text{ for } i \neq 1, \ \ i \neq j \quad (2.12)$$

then such a prescription for D satisfies (2.9) and $D \geq 0$ for every $\alpha \in [0,1]$. In order to show that a D given by (2.10) - (2.12) is feasible for every $\alpha \in [0,1]$, it remains to show that $\Sigma - D \geq 0$ holds. Note that for the j-th elementary regression $d_{jj}^{(j)}$ is the unique solution of the equation

$$\det(\Sigma - \text{diag}(0, \cdots, d_{jj}, \ \ 0, \cdots, 0)) = 0 \quad (2.13)$$

in the variable $d_{jj} \in \mathbf{R}$. This is a direct consequence of the fact that (2.13) is a linear equation with a positive coefficient for $(\sigma_{jj} - d_{jj})$ (compare 2.6). Now performing the j-th elementary regression for $\Sigma - \alpha D_1, \alpha \in (0,1)$ we see that the corresponding noise covariance matrix is $\text{diag}\{0, \cdots, d_{jj}, \ \ 0, \cdots, 0\}$ with d_{jj} given by (2.11) and thus $\Sigma - D = (\Sigma - \alpha D_1) - diag\{0, \cdots, d_{jj}, \ \ 0, \cdots, 0\} \geq 0$.

Let us consider the set

$$F = \{\sum_{j=2}^{n} \beta_j s_j \mid \sum_{2}^{n} \beta_j = 1, \ \ \beta_j \in \mathbf{C}\} \quad (2.14)$$

For every $y \in F \cap \hat{L}$ we can choose a corresponding feasible D with $d_{11} = 0$. Thus, for $x = s_1$ and $y \in F \cap \hat{L}$ in (2.8) we have (2.10) and thus $\alpha \in [0,1]$. [Clearly, here, in general not every $\alpha \in [0,1]$ gives a solution].

In an analogous way as before we proceed in the case $x = s_k$, $y = s_j, \ k, j \neq 1; \ k \neq j$. Then from (2.9) we obtain

$$\alpha s_{kk} d_{kk}^{(k)} = (\alpha s_{kk} + (1-\alpha)s_{jk})d_{kk} \quad (2.15)$$

for the k-th equation and an analogous term for the j-th equation. Again we put $d_{ii} = 0, \ i \neq k, \ i \neq j$. Equations (2.9) then are equivalent to

$$(\frac{1}{\alpha} - 1) \cdot \frac{s_{jk}}{s_{kk}} \geq 0$$

which in turn is equivalent to

$$\arg(\frac{1}{\alpha} - 1) = \arg s_{kk} - \arg s_{jk} \quad (2.16)$$

Using the same argument as above, we can show that $\Sigma - D \geq 0$ holds. Now let us discuss condition (2.16). It describes a real curve in the real plane $\alpha s_k + (1-\alpha) \cdot s_j, \ \alpha \in \mathbf{C}$. We may distinguish between three cases:

(i) If $\frac{s_{jk}}{s_{kk}} > 0$ holds, then (2.16) is equivalent to $\alpha \in [0,1]$; i.e. the solutions are the part of the line connecting s_k and s_j which is between these points

(ii) If $\frac{s_{jk}}{s_{kk}} < 0$ holds, then (2.16) is equivalent to $\alpha \in [1, \infty) \cup (0, -\infty)$; then the solutions are the part of the line connecting s_k and s_j which is outside of this point s

(iii) If $\frac{s_{jk}}{s_{kk}}$ is not real, then the solutions corresponding to the set of all feasible α are an arc of a circle

3 The Spectral Densities of the Observations for the Case $m^* = n - 1$

In this and in the next section we give a complete analysis of the case $m^* = n-1$, i.e. the case which can be described by one single factor: We give a characterization of Σ in this case, a description of the set of all Σ corresponding to $m^* = n-1$ and finally a description of the solution set (for given Σ, also in terms of different coranks m). Part of the results have been available for the static case for a long time (see e.g. [6]). For the dynamic $n = 3$ case see [5]. The next theorem gives a characterization of the case $m^* = n - 1$.

Theorem 3.1 *For $n > 3$ the following statements are equivalent:*

(a) $mc(\Sigma) = n - 1$

(b) There exists a diagonal unitary matrix U such that $U^*\Sigma U = (\tau_{ij}), \ i,j = 1, \cdots, n$ is real with all entries τ_{ij} positive and satisfying

$$\begin{aligned} \tau_{ik} \cdot \tau_{jl} - \tau_{il} \cdot \tau_{jk} &= 0; \ i,j,k,l \text{ all different} \quad (3.1) \\ \tau_{ii} \cdot \tau_{jk} - \tau_{ik} \cdot \tau_{ji} &\geq 0; \ i,j,k \text{ all different} \quad (3.2) \end{aligned}$$

(c) There exists a diagonal unitary matrix U such that $U^*\Sigma^{-1}U = (t_{ij}), \ i,j = 1, \cdots, n$ is real with all off-diagonal elements negative and satisfying

$$t_{ik} \cdot t_{jl} - t_{il} \cdot t_{jk} = 0 \ ; \ i,j,k,l \text{ all different} \quad (3.3)$$

Proof "(a)$\longrightarrow$(b)": Since there is a feasible $\hat{\Sigma}$ with rank equal to one, we can write

$$\hat{\Sigma} = \lambda\lambda^* \ ; \ \ \lambda \in \mathbf{C}^n$$

In addition, for $U = \text{diag}\{e^{i \arg \lambda_j}\}$, where $\arg\lambda_j$ denotes the phase of the j-th entry of $\lambda, \hat{\lambda} = U^*\lambda$ has all elements positive (Note: since Σ has all entries nonzero, so has λ). Thus all entries in

$$U^*\Sigma U = \hat{\lambda}\hat{\lambda}' + D \quad (3.4)$$

are positive. (3.1) and (3.2) are straightforward.

"(b)$\longrightarrow$(a)": Consider the matrix $T = \{\tau_{ij}\} = U^*\Sigma U$; then we can find a decomposition $T = \hat{T} + \tilde{T}$ defined by $\hat{\tau}_{ij} = \tau_{ij}$, for $i \neq j$ and by

$$\hat{\tau}_{ii} = \tau_{ik} \cdot \tau_{ji} \cdot \tau_{jk}^{-1} \quad ; \quad i,j,k, \text{ all different}$$

Then as easily can be checked from (3.1), $\hat{\tau}_{ii}$ is independent of j,k and all minors of $\hat{T}$ of order 2 are equal to zero and thus $\hat{T}$ has rank 1. Clearly $\hat{\tau}_{ii} > 0$ since $\tau_{ij} > 0, i,j = 1, \cdots, n$, and thus $T \geq 0$. Furthermore $\tilde{T} \geq 0$ holds by (3.2) and thus $\hat{T}$ and $\tilde{T}$ are feasible for T, which immediately implies that $U\hat{T}U^*$ and $U\tilde{T}U^*$ are feasible for Σ.

"(b)$\longrightarrow$(c)": Suppose temporarily that $D > 0$ holds. It is easily to check that (3.4) implies

$$(U^*\Sigma U)^{-1} = D^{-1} - \frac{1}{1+\hat{\lambda}'D^{-1}\hat{\lambda}} \cdot D^{-1}\hat{\lambda}\hat{\lambda}'D^{-1} = D^{-1} - \hat{\mu}\hat{\mu}' \quad (3.5)$$

where $\hat{\mu}$ is a vector with all positive entries. It is immediate that all off-diagonal entries of $(U^*\Sigma U)^{-1}$ are negative.

Now suppose $D \geq 0$. Since Σ is nonsingular, the entries of $(U^*(\Sigma + \epsilon I)U)^{-1}$ $\epsilon \geq 0$ depend continuously on ϵ. For all $\epsilon > 0$ the off diagonal elements are negative; hence in the limit for $\epsilon \to 0$ they are nonpositive and thus by assumption (vii) negative also. (3.3) directly follows from (3.4).

"(c)$\longrightarrow$(b)": Let $T = (t_{ij}) = U^*\Sigma^{-1}U$ and $T = \hat{T} + \tilde{T}$ where $\hat{T}$ is defined by $\hat{t}_{ij} = t_{ij};\ i \neq j$ and

$$\hat{t}_{ii} = \frac{\hat{t}_{ik} \cdot \hat{t}_{ji}}{\hat{t}_{jk}}\ ;\ i, j, k, \text{ all different}$$

Again it is easy to see that $\hat{t}_{ii}$ is independent of j, k and that $\hat{T}$ has rank 1. Furthermore since all $t_{ij},\ i \neq j$, are negative, $\hat{T}$ consists of negative elements only and thus can be written as

$$\hat{T} = -\hat{\mu}\hat{\mu}'\ ;\ \hat{\mu} \in \mathbf{R}^n_+ \quad (\text{i.e. } \hat{\mu}_i > 0)$$

$\tilde{T}$ then is diagonal with strictly positive diagonal elements. From (3.5) we then obtain

$$T^{-1} = U^*\Sigma U = \tilde{T}^{-1} + \frac{1}{1 + \hat{\mu}'\tilde{T}^{-1}\hat{\mu}} \cdot \tilde{T}^{-1}\hat{\mu}\hat{\mu}'\tilde{T}^{-1}$$

which implies (b). ∎

Remark 1 Note that for $m^* = n - 1$, by

$$x(\Sigma - D) = 0 \Longleftrightarrow xUU^*(\Sigma - D)U = 0$$

(with U defined as above), for every complex problem $x\ U$ gives a one-to-one relation to a solution with a real Σ. Note that in general the entries of $x\ U$ will be complex.

Remark 2 If we drop assumptions (vi) and (vii), then Theorem 3.1 remains true if we replace (b) by

(b^*): There exists a diagonal unitary matrix U such that $U^*\Sigma U = (\tau_{ij}),\quad i, j = 1, \cdots, n$ is real with all entries τ_{ij} nonnegative and satisfying (3.1),(3.2) and

$$\tau_{ij} > 0,\ i \neq j, \text{ for at least one pair } (i, j)$$

and (c) by:

(c^*): There exists a diagonal unitary U such that either (c) or

(c_1): All off-diagonal elements in one row and column of (τ_{ij}) are negative and zero in all other rows and columns hold.

Remark 3 Next let us investigate the set Σ_{n-1} of all Σ (satisfying our assumptions) with $mc(\Sigma) = n - 1$. Consider the mapping i defined by $i(\lambda, D) = \lambda\lambda^* + D$, where $\lambda \in \mathbf{C}^n$ and $D \geq 0$, diagonal. If we impose the additional normalization assumption $\lambda_1 > 0$, it is easily seen from the condition that all 2×2 minors of $\hat{\Sigma}$ are zero that, for $n \geq 3$ (and imposing (vi)), λ and D are uniquely determined from Σ and that i and i^{-1} are continuous. Thus Σ_{n-1} is a subset of real dimension $3n - 1$ of the set Σ of all $\Sigma > 0$, which is of real dimension n^2 and i^{-1} provides a useful parametrization for Σ_{n-1}. In particular we see that for $n \geq 3$ the single factor case is highly nongeneric for the spectral densities of the observations.

4 The Solution Set for $m^* = n - 1$

In this section we will give a complete description of the set of all solutions corresponding to a given Σ with $mc(\Sigma) = n - 1$. We also provide a characterization of solutions corresponding to different m. By L_m we denote the set of all solutions $x \in L$ where there is a feasible $\hat{\Sigma}$ such that $x\hat{\Sigma} = 0$ and $\dim_{\mathbf{C}} \ker\hat{\Sigma} = m$ hold. For $x = (x_1, \cdots, x_n) \in \mathbf{C}^n$ we define the projection π by $\pi(x) = (x_2, \cdots, x_n) \in \mathbf{C}^{n-1}$ and analogously we define $\pi(M) = \{\pi(x) \mid x \in M\}$ for a set $M \subset \mathbf{C}^n$.

Theorem 4.1 *Let $n \geq 3$ and $mc(\Sigma) = n - 1$. Then*

(a) There is a unique feasible $D, D^* = \operatorname{diag}\{d^*_{ii}\}$ say, given by

$$d^*_{ii} = \sigma_{ii} - \frac{\sigma_{ik} \cdot \sigma_{ji}}{\sigma_{jk}}\ ;\ i, j, k \text{ all different} \tag{4.1}$$

such that $\Sigma - D^*$ has rank equal to one; in this case

$$L_{n-1} = \pi(\ker(\Sigma - D^*) \cap \{x \in \mathbf{C}^n \mid x_1 = 1\}) \tag{4.2}$$

and

$$\dim_{\mathbf{C}}(L_{n-1}) = n - 2 \tag{4.3}$$

hold.

(b) Let

$$x_j = \alpha^* s_1 + (1 - \alpha^*)s_j\ ;\ j = 2, \cdots, n \tag{4.4}$$

where

$$\alpha^* = \left[1 - \frac{s_{13}}{s_{23}}\right]^{-1} \tag{4.5}$$

Then all entries of x_j not in positions 1 and j are equal to zero and

$$L_{n-1} = \pi\left[\left\{\sum_{i=2}^{n} \beta_j x_j \mid \sum_{2}^{n} \beta_j = 1\right\}\right] \tag{4.6}$$

(c) For every m such that $1 < m < n - 1, L_m$ is a nonvoid subset of L_{n-1}, additionally satisfying $L_m \subset L_{m+1}$.

(d) L_1 can be written as $L_1 = L_{1,U} \cup L_{1,L}$ where

$$L_{1,U} = \pi(\{\alpha s_1 + (1 - \alpha)\sum_{j=2}^{n} \beta_j x_j \mid \alpha, \beta \in [0, 1], \sum_{2}^{n} \beta_j = 1\}) \tag{4.7}$$

$$\begin{aligned} L_{1,L} &= \pi(\{\alpha s_1 + (1 - \alpha)\sum_{j=2}^{n} \beta_j s_j \mid \alpha \in [0, \alpha^*), \\ \sum_{2}^{n} \beta_j &= 1,\ \beta_k \geq 1 \text{ for some } k \geq 2 \\ &\quad \text{and } \beta_j \leq 0 \text{ for } j \neq 1, j \neq k\}) \end{aligned} \tag{4.8}$$

Proof The first part of (a) directly follows from the fact that all 2×2 minors of $(\Sigma - D^*)$ have to be zero and from assumption (vi), which together imply that D^* is uniquely determined. In order to show (4.3) note that the intersection between $\ker(\Sigma - D^*)$ and $\{x \in \mathbf{C}^{n-1} \mid x_1\}$ is always nonvoid: Since $\Sigma - D^*$ has rank equal to one and due to (vi) it is always possible to express the first row of $\Sigma - D^*$ as a linear combination of the others.

(b) Consider the equation system

$$\alpha s_1 D_1 + (1 - \alpha)s_j D_j = (\alpha s_1 + (1 - \alpha)s_j)D \tag{4.9}$$

From section 2 we know that for every $\alpha \in [0, 1], x = \alpha s_1 + (1 - \alpha)s_j$ is a solution. From (3.5) we see that if we put $\alpha^* = (1 - s_{1i} \cdot s_{ji}^{-1})^{-1}, i \neq 1, i \neq j$; then $s_{1i} \cdot s_{ji}^{-1} < 0$ implies $\alpha^* \in (0, 1)$ and all entries of x not in positions 1 and j are equal to zero. Furthermore α^* does not depend on i. Now define $D = D^*$ in (4.9) by

$$d^*_{11} = \alpha^* d^{(1)}_{11}\ ;\ d^*_{jj} = (1 + \alpha(1 - \alpha)^{-1} s_{1j} \cdot s_{jj}^{-1})^{-1} d^{(j)}_{jj} \tag{4.10}$$

(see (2.11) and (2.14)) and

$$d^*_{ii} = \sigma_{ii} - \frac{\sigma_{ik}\sigma_{1i}}{\sigma_{1k}}\ ;\ k \neq i, 1 \neq i, 1 \neq k \tag{4.11}$$

Then it is straightforward to show that the d^*_{ii} do not depend on $1, k$ and that $\Sigma - D^* \geq 0$. Furthermore for given d^*_{11}, d^*_{jj}, any $d_{ii} > d^*_{ii}\ ;\ i \neq 1,\ i \neq j$, would violate the nonnegative definiteness condition on $\hat{\Sigma}$. Now it is easy to check that $\Sigma - D^*$ has rank 1; thus $x_j \in \ker(\Sigma - D^*)\ ;\ j = 2, \cdots, n$ and these points generate L_{n-1} by (4.6).

(c) First we show that if D is a feasible matrix with $\dim_{\mathbf{C}} \ker(\Sigma - D) = m$, $1 < m < n-1$, then D must be of the form

$$D = D^* - \tilde{D} \tag{4.12}$$

where $\tilde{D}$ is a nonnegative diagonal matrix. This follows from the fact that for every feasible D, all 2×2 principal minors of $\Sigma - D$ are nonnegative, and then D^* can have at most one element smaller than the corresponding element in D. If there is such a strictly smaller element in D^* then all other elements in D^* must be strictly larger than the corresponding elements in D, which implies $\dim_{\mathbf{C}} \ker(\Sigma - D) = 1$ and thus D with $\dim_{\mathbf{C}} \ker(\Sigma - D) = m$, $1 < m < n-1$ must be of the form (4.12) with $\tilde{D} \geq 0$. Note that for two matrices $A \geq 0, B \geq 0$ we have

$$\ker(A+B) = \ker A \cap \ker B$$

Thus $\ker(\Sigma - D) = \ker(\Sigma - D^*) \cap \ker\tilde{D}$; as easily can be seen we also have $\dim \ker_{\mathbf{C}}(\Sigma - D) = (n-1) - \dim \ker_{\mathbf{C}}(\tilde{D})$. In particular, for given $\tilde{D}$ with the non-zero entries in positions $1 \leq j_1 \cdots < j_{n-m+1} \leq n$, then all x_j such that $j \neq j_1, \cdots, j \neq j_{n-m+1}$ form a basis for $\ker(\Sigma - (D^* + \tilde{D}))$.

(d) Consider the equation system

$$\begin{aligned}(\alpha s_1 + (1-\alpha) \cdot \sum_{j=1}^{n} \beta_j x_j)\Sigma &= \alpha s_1 D_1 + (1-\alpha)[\sum_{2}^{n} \beta_j x_j] D^* \\ &= (\alpha s_1 + (1-\alpha) \sum_{2}^{n} \beta_j x_j) D \ ; \\ & \sum_{2}^{n} \beta_j = 1 \end{aligned} \tag{4.13}$$

The first equation in (4.13) gives

$$\alpha(d_{11}^{(1)} - d_{11}^*) = d_{11} \tag{4.14}$$

and the remaining equations give

$$d_{jj} = d_{jj}^* \left(1 + \frac{\alpha s_{1j}}{(1-\alpha)\beta_j x_{jj}}\right)^{-1} \ ; \ j \geq 2 \tag{4.15}$$

where x_{jj} denotes the j-th component of x_j. From section 2 we know that $\alpha \leq 1$ and $(d_{11}^{(1)} - d_{11}^*) > 0$ hold. Thus, if we first consider the case $\alpha \geq 0$, from (4.14) we have $d_{11} \geq d_{11}^*$ where the strict inequality holds for $\alpha > 0$. Thus by the same argument as in the proof of (c) we see that if $\alpha > 0$ gives a solution, then it corresponds to $m = 1$. Furthermore $d_{11} \geq d_{11}^*(d_{11} > d_{11}^*)$ implies $d_{jj} \leq d_{jj}^*(d_{jj} < d_{jj}^*), j > 1$ and as easily can be checked $s_{1j} \cdot x_{jj}^{-1} > 0$ holds. Thus (4.15) implies $\beta_j \geq 0 (\beta_j > 0)$ for $\alpha \in (0,1)$.

Next we have to show that for every $\alpha \in [0,1], \beta_j \in [0,1]$, $\sum_2^n \beta_j = 1$, the element $\alpha s_1 + (1-\alpha)[\sum_2^n \beta_j x_j]$ is a solution: This is straightforward from the fact that D in (4.13) can be obtained from the first elementary regression for $\Sigma - \mathrm{diag}\{0, d_{22}^*, \cdots, d_{nn}\}$ [note that $\pi(\Sigma \beta_j x_j) \in L_{n-1}$ and see section 2]. This shows the part for $L_{1,U}$. The proof for $L_{1,L}$ is similar: There we consider the case $\alpha < 0$. Then (4.14) implies $d_{11} < d_{11}^*$ and (since $d_{11} \geq 0$ holds) $\alpha \geq \alpha_{\min} = -d_{11}^* \cdot (d_{11}^1 - d_{11}^*)^{-1}$. Since $d_{jj} > d_{jj}^*$ must hold with possible exception of one $k \geq 2$, this implies $\beta_j \leq 0$. But then from $\sum_2^n \beta_j = 1$ we must have $\beta_k \geq 1$ for one $k \geq 2$; in this case $d_{kk} > d_{kk}^*$, thus these solutions correspond to $m = 1$. Further note that $\alpha_{\min} s_1 + (1-\alpha_{\min}) \cdot x_j = s_j$ holds. ∎

In the next figure we show a solution set L for $n = 4, m^* = 3$, in the static (real) case. Note that for the real static case $\dim_{\mathbf{R}} L_m$ is equal to $\dim_{\mathbf{C}} L_m$ for the complex case.

Remark 1 For $n = 2$, in general D^* would not be unique (see [4]). Also for general n, if (vi) is violated, D^* may not be unique, as can be easily seen from the case $\sigma_{ij} = 0$, for all $i \neq j$ except for $(i,j) = (1,2)$. On the other hand for n sufficiently large in relation to m^*, D^* may be unique also for $m^* < n-1$ (under additional conditions which are generically satisfied). For example for $m^* = n-2$ and $n \geq 5$ we can calculate the corresponding D^* from the conditions that all 3×3 minors of $\hat{\Sigma}$ are zero.

Remark 2 For most cases, the main interest is in the solutions corresponding to m^*; this corresponds to the maximum degree of explanation by the system versus the "outside world". Note that the x_j in (4.4) belong to all L_m, $1 \leq m \leq n-1$. Using y_t as a symbol for outputs and x_t as a symbol for inputs we may rewrite equation (1.3) in an evident notation as

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{n,t} \end{bmatrix} \tag{4.16}$$

The systems corresponding to $m = n-1$ have one input and $n-1$ outputs. Moreover, by (vi) every entry of $\hat{z}_t$ may be chosen as an input (which of course in general will not be causal for the outputs). Once the decision which entry of $\hat{z}_t$ is chosen as the input has been made, L_{n-1} corresponds to a single system.

Remark 3 As is easy to see, for $m^* = n-1$, the solution sets L_m continuously depend on Σ.

Let for instance the first component of $\hat{z}_t, \hat{z}_{1,t} = \hat{x}_t$ say denote the "true" input; then using the x_j in (4.4), $j = 2, \cdots, n$, we have the true outputs:

$$\hat{z}_{j,t} = \hat{y}_{t-1,t} = k_{j-1}(z)\hat{x}_t \ ; \ j = 2, \cdots, n \tag{4.17}$$

where $k_{j-1}(e^{i\lambda}) = (-x_{jj})^{-1}$. Clearly (4.16) and (4.17) has a direct interpretation as a dynamic single factor model.

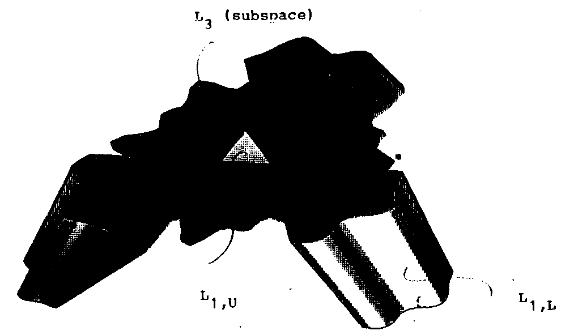

Fig.1: A "typical" solution-set for $m^* = n-1 : n = 4, m^* = 3$, static (real) case

5 References

[1] J.F. Geweke, "The dynamic factor analysis of economic time series", in D. Aigner and A. Goldberger, eds., *Latent Variables in Socioeconomic Models*, North Holland, Amsterdam, 1977, pp. 365-383.

[2] R. Engle and M. Watson, "A one-factor multivariate time series model of metropolitan wage rates", *Journal of the American Statistical Association*, 1981, Vol. 76, pp. 774-781.

[3] R.E. Kalman, "System identification from noisy data", in A. Bednarek and L. Cesari, eds., *Dynamic Systems II, a University of Florida International Symposium*, New York: Academic Press, 1982.

[4] B.D.O. Anderson and M. Deistler, "Identifiability in dynamic errors-in-variables models", *Journal of Time Series Analysis*, 1984, Vol. 5, pp. 1-13.

[5] B.D.O. Anderson and M. Deistler, "Dynamic errors-in-variables systems with three variables", *Automatica*, 1987, Vol. 23, pp. 611-616.

[6] P. Bekker and J. de Leeuw, "The rank of reduced dispersion matrices", *Psychometrica*, 1987, Vol. 52, pp. 125-135.

[7] M. Deistler and B.D.O. Anderson, "Linear dynamic errors-in-variables models, some structure theory", *Journal of Econometrics*, 1989, Vol. 41, pp. 39-63.

[8] J.C. Willems, "From time series to linear systems, Part I", *Automatica*, 1986, Vol. 22, pp. 561-580.

[9] M. Deistler, "Symmetric modelling in system identification" in H. Nijmeijer and J. M. Schumacher, eds., *Three Decades of Mathematical System Theory*, Springer Lecture Notes in Control and Information Sciences, no. 135, 1989, Springer-Verlag, Berlin.

[10] M. Deistler and B.D.O. Anderson, "Identification of dynamic systems from noisy data: The case $m^* = 1$", submitted for publication.

[11] C. Spearman, "General intelligence, objectively determined and measured", *American Journal of Psychology*, 1990, Vol. 15, pp. 201-293.

[12] T.W. Anderson and H. Rubin, "Statistical inference in factor analysis", in J. Neyman, ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press: Berkeley, 1956, pp. 11-150.