

LINEAR DYNAMIC ERRORS-IN-VARIABLES MODELS Some Structure Theory

M. DEISTLER

Technical University of Vienna, A-1040 Vienna, Austria

B. D. O. ANDERSON

Australian National University, Canberra ACT 2600, Australia

This paper gives a survey of some recent results on problems of identifiability (or, more general, of the relation between the observations and certain system characteristics) for linear dynamic errors-in-variables models. For a large part of the paper the noise components are assumed to be mutually uncorrelated. After the general problem statement, a rather complete analysis of the single-input-single-output case is given. Also the case of three variables and the case where the number of inputs is equal to the number of outputs are discussed in detail. Finally, the use of higher-order cumulant spectra for identifiability is investigated.

1. Introduction

Errors-in-variables (EV) and related models, such as factor (analysis) models, have been investigated in a number of different areas as there are econometrics, psychometrics, and system engineering. In the last fifteen years there has been a resurging interest in such models [see, e.g., Aigner and Goldberger (1977), Kalman (1982, 1983), Aigner et al. (1984), T.W. Anderson (1984), Klepper and Leamer (1984), Picci and Pinzoni (1986), see also the recent books by Schneeweiß und Mittag (1986) and Fuller (1987)].

For the most part, for a long time, research has concentrated on the linear static case [see the surveys by Madansky (1959), Moran (1971), Aigner et al. (1984), T.W. Anderson (1984)]. Although it was recognized quite early [Reiersøl (1941), Geary (1943)] that a *certain kind* of dynamics may help to obtain identifiability of the system, most of the *systematic* research in the area of *dynamic* EV models is quite recent [Maravall (1979), Söderström (1980), B.D.O. Anderson and Deistler (1984), Anderson (1985a), Picci and Pinzoni (1986), Solo (1986), Deistler (1986b)].

Whereas in the 'conventional' errors-in-equations approach to linear system identification, all noise is added to the outputs (or for our purposes equivalently to the equations), in the EV approach all variables may be contaminated

by noise. Errors-in-equations modelling is appropriate for a number of purposes, e.g., for predicting the future values of the observed outputs from their own past and from the observed inputs. The main cases when the (more general) EV setting is appropriate are:

(i) If we are interested in the 'true' system generating the data (rather than in coding the data by system parameters) and if we cannot be sure *a priori* that the observed inputs are not corrupted by noise. This is the 'classical' motivation for EV modelling. Needless to say, true systems are always fictions, but they may be useful ones [compare the example in Madansky (1959) where the specific weight, measured from the variables volume and weight, corresponds to the true system].

(ii) EV modelling provides a symmetric way of noise modelling, since all variables, in principle, may be contaminated by noise. This is of particular interest in cases where there is no *a priori* classification of the variables into inputs and outputs or where even the number of equations of the system is not known *a priori*; in such cases also a symmetric system modelling (for the system relating the latent variables) is appropriate in order to avoid inadequate assumptions. This point has been emphasized by Kalman (1982, 1983).

An alternative way providing a symmetric treatment of all variables is to use autoregressive or ARMA models where the AR or ARMA process is the process of all observed variables [Sims (1980)]. However, in many cases this leads to parameter spaces with dimension being considerably higher than for the corresponding EV systems.

(iii) Under certain additional assumptions on the noise structure, EV models are equivalent to dynamic principal component models or to dynamic factor models. If we assume that the noise components are mutually uncorrelated [assumption (vi) below], then the model provides a decoupling of common and individual effects between the variables, where all common effects are attributed to the system.

As opposed to the errors-in-equations case, there is still a great number of open problems in EV model identification; for the most part these problems are connected with the more complicated structure of such models and in particular with the fact that in general the transfer functions of the systems relating the latent variables are not uniquely determined from the second moments of the observations. This is the reason why, despite of their potential great importance, the number of applications in this area is still rather limited.

2. The general problem

Let \hat{z}_t denote the n -dimensional vector of latent (i.e., in general unobserved) variables (\hat{z}_t is sometimes called the signal) which satisfy the 'exact' linear

dynamic relation

$$w(z)\hat{z}_t = 0, \quad (2.1)$$

where z is used for a complex variable as well as for the backward shift operator on \mathbb{Z} [i.e., $z(\hat{z}_t) = (\hat{z}_{t-1})$] and

$$w(z) = \sum_{l=-\infty}^{\infty} W_l z^l, \quad W_l \in \mathbb{R}^{m \times n}. \quad (2.2)$$

The W_l must be such that the infinite sum in (2.1) converges. Without restriction of generality we assume that $m \leq n$ holds and that the rows of w are linearly independent (in the sense that $w(e^{-i\lambda})$ has an m th-order determinant which is unequal to zero on a set of positive Lebesgue measure on $[-\pi, \pi]$).

In (2.1) we allow for a symmetric treatment of all components of \hat{z}_t , and we need not distinguish between inputs and outputs *a priori*. Systems of this kind are discussed in detail in Willems (1979, 1986) and Blomberg and Ylinen (1983). If the classification into inputs \hat{x}_t and outputs \hat{y}_t is known, then the conventional transfer function representation of a linear dynamic system

$$\hat{y}_t = k(z)\hat{x}_t, \quad (2.3)$$

where

$$k(z) = \sum_{j=-\infty}^{\infty} K_j z^j, \quad K_j \in \mathbb{R}^{m \times (n-m)}. \quad (2.4)$$

can be used. Clearly this can be written in the more general form (2.1) with $\hat{z}_t = (\hat{x}_t', \hat{y}_t)'$, $w(z) = (k(z), -I)$. We will use the word transfer function for both $k(z)$ and $w(z)$, the difference in meaning is indicated by the letter used.

Conversely [if there is no such *a priori* classification into inputs and outputs], for given $w(z)$ any selection of m linear independent columns of w gives rise to a selection of outputs as the corresponding selection of elements in \hat{z}_t , and to a representation (2.3). The observations z_t are of the form

$$z_t = \hat{z}_t + u_t, \quad (2.5)$$

where u_t is the noise. In the errors-in-equations case the system is of the form (2.3) and eq. (2.5) is of the form

$$z_t = (\hat{x}_t', y_t')' = (\hat{x}_t', \hat{y}_t')' + (0, w_t')'. \quad (2.6)$$

We want to emphasize both the symmetry in the system modelling (2.1) and in

the noise modelling (2.5). This may be of particular interest in econometrics where often all variables are interrelated.

Throughout we assume:

- (i) All processes considered are (wide-sense) stationary. In addition we assume that all limits of random variables are understood in the sense of mean squares convergence.
- (ii) $E\hat{z}_t = 0$ and $Eu_t = 0$.
- (iii) $E\hat{z}_t u_t' = 0$.
- (iv) The spectral density \hat{f} of the noise (u_t) exists.

Clearly (iii), (iv), and $Eu_t = 0$ are rather innocent assumptions. If \hat{z}_t has a constant, nonzero mean, we have $w(1)Ez_t = 0$, and thus we have additional information from the first moments of z_t . Also certain nonstationarities like trends in means and variances may help in identifying the system.

In addition, unless the contrary is stated explicitly, we assume:

- (v) (z_t) , (\hat{z}_t) , and (u_t) are ARMA processes (i.e., their spectral densities f , \hat{f} , and \hat{f} , respectively, are rational) and $w(z)$ is a rational matrix.

Under our assumptions we have

$$f = \hat{f} + \hat{f}, \quad (2.7)$$

where \hat{f} is singular and satisfies

$$w\hat{f}w^* = 0. \quad (2.8)$$

(λ - a.e. and thus due to (v) for all $\lambda \in [-\pi, \pi]$.) Here w^* denotes the conjugate transpose of w , and w is evaluated on the unit circle of the complex plane.

Now, for the moment, let us assume that \hat{f} (and no additional information about (2.1)) is given. Note that the rational functions form a field; matrices like \hat{f} and w are considered as matrices over this field and terms like nonsingular or linear dependent are used in this sense. Thus, in particular a rational square matrix u is nonsingular if $\det u(z) \neq 0$ for all but a finite number of $z \in \mathbb{C}$. If no *a priori* classification of the elements of \hat{z}_t into inputs and outputs is given, then every w satisfying (2.8) gives rise to a representation (2.1) and any subvector, \hat{y}_t say, of \hat{z}_t corresponding to linear dependent rows of \hat{f} may be taken as a vector of outputs. We will take the dimension of \hat{y}_t as

large as possible since we want to endogenize as much as possible. Thus we assume that, for given \hat{f} , w has the maximum number of linearly independent rows. Then, for given \hat{f} , the transfer function w is unique up to left multiplication by arbitrary, nonsingular, rational matrices u . This is easily seen from (2.8), since saying w is a transfer function corresponding to \hat{f} is the same as saying that the rows of w form a basis for the (left) kernel of \hat{f} , and thus we are free to choose any such basis. Thus, if \hat{y}_t and \hat{x}_t denote suitably selected subvectors of \hat{z}_t , then eq. (2.1) can be always transformed to (2.3); alternatively, since $w(z)$ can be written as

$$w(z) = p^{-1}(z)c(z),$$

where p is a scalar polynomial, being a least common denominator of the entries of w , and c is a polynomial matrix. After eventually suitably reordering of the components of \hat{z}_t , (2.1) can be transformed to a vector difference equation,

$$a(z)\hat{y}_t = b(z)\hat{x}_t,$$

where a and b are polynomial, $a(z)$ is nonsingular, $c = (a, -b)$ and $\hat{z}_t = (\hat{y}_t, \hat{x}_t)$ hold.

The restrictions on \hat{z}_t imposed by (2.1) may alternatively be expressed as

$$\hat{z}_t = t(z)s_t, \quad (2.9)$$

where the columns of t form a basis for the right kernel of w and where s_t is a suitably chosen 'factor process'. The formulation (2.9) is used in Picci and Pinzoni (1986).

However, in general, \hat{f} will not be given. In general terms, the problems we consider are as follows: Given the probability law of the observed process (z_t) :

- (a) Find the maximum number, m^* say, of (linearly independent) linear equations of the form (2.1) for (\hat{z}_t) compatible with the law of (z_t) .
- (b) For given m^* , give a characterization of the set of all $w(z)$ [and eventually also of all (\hat{z}_t) and (u_t)] compatible with the law of (z_t) .
- (c) What are additional conditions guaranteeing uniqueness of $w(z)$.

It should be stressed again that this setting is considerably more general than the errors-in-equations setting; in particular here the number of equations (and thus the number of outputs) does not have to be known *a priori*. The fact that we restrict our analysis to the maximum number of m^* equations means that we want to endogenize and, thus, to explain by the system as much as is possible.

As is clear already from the simplest static case, some additional structure on the noise (u_t) has to be imposed in order to give our problem a certain degree of determination. With the exception of the last section we will assume, unless the contrary is stated, that the following holds:

- (vi) \tilde{f} is diagonal, i.e., the component processes of (u_t) are mutually uncorrelated.

Partly this assumption is replaced by:

- (vii) \tilde{f} is block-diagonal, where the blocks correspond to a partitioning of z_t into inputs and outputs.

For most parts of the paper we assume that the external information available consists of the second moments f only. Then, under (vi), our problem may be formulated as follows: For given f consider all decompositions (2.7) into spectral densities \hat{f} and \tilde{f} such that \hat{f} is singular and \tilde{f} is diagonal.

- (a) What is the minimum rank, $n - m^*$ say, of \hat{f} among all these decompositions?
 (b) For given m^* , characterize the set of all \hat{f} of rank $n - m^*$ which are compatible with given f .
 (c) Under which additional assumptions is \hat{f} unique?

The following additional questions relate to causality in this context:

- (d) If we have no *a priori* causality assumptions, what kind of causality statements could be made from the knowledge of f ?
 (e) If we do know *a priori* what is an output and what an input and that the outputs are causally influenced by the inputs, what are the additional restrictions on the set of all feasible transfer functions k ?

Up to now there is no general solution to the problems listed above; in the next sections we discuss some cases which have been solved already.

It should be noted that if one drops the rationality assumption (v), then for every spectral density matrix f there exists a decomposition (2.7) [where \tilde{f}

satisfies (vi)] and, thus, a corresponding EV system (2.1), (2.5). This is easily seen from

$$\begin{aligned}
 f(\lambda) &= u(\lambda) \begin{pmatrix} \lambda_1(\lambda) & & & 0 \\ & \lambda_2(\lambda) & & \\ & & \dots & \\ 0 & & & \lambda_n(\lambda) \end{pmatrix} u(\lambda)^* \\
 &= u(\lambda) \begin{pmatrix} 0 & & & \\ & \lambda_2(\lambda) - \lambda_1(\lambda) & & 0 \\ & & \dots & \\ 0 & & & \lambda_n(\lambda) - \lambda_1(\lambda) \end{pmatrix} u(\lambda)^* \\
 &\quad + \lambda_1(\lambda) I,
 \end{aligned}$$

where $\lambda_1(\lambda) \leq \lambda_2(\lambda) \leq \dots \leq \lambda_n(\lambda)$ are the eigenvalues of $f(\lambda)$ and $u(\lambda)$ is a unitary matrix.

An alternative possibility of imposing additional structure on the noise [instead of (vi) or (vii)] is to define \hat{f} as the best rank $n - m^*$ approximation (in a certain metric) to f . This corresponds to the dynamic factor model [Brillinger (1981)]. In other cases it is possible to separate the noise u_t from the signal \hat{z}_t , by additional *a priori* information in frequency domain [Maravall (1979), Söderström (1980), Anderson and Deistler (1984), Deistler (1984)]. Closely related to this is the idea of cointegration [Engle and Granger (1987), Phillips and Ouliaris (1986)]. Recall that, if a process z_t is integrated of order 1, i.e., if

$$(1 - z)z_t = c(z)\varepsilon_t,$$

where $c(z_t)\varepsilon_t$ is stationary (written in Wold representation) whereas z_t is not stationary, then z_t is said to be cointegrated if there is a vector $\alpha \in \mathbb{R}^n$, $\alpha \neq 0$, such that αz_t is stationary. Since $c(z) = c(1) + (1 - z)c^*(z)$ for suitably chosen c^* , z_t is cointegrated if and only if $c(1)$ is singular, and defining $\hat{z}_t = (1 - z)^{-1}c(1)\varepsilon_t$ and $u_t = c^*(z)\varepsilon_t$, this gives an EV system (2.1), (2.5). All these approaches will not be discussed further here. However, in the last section the case of non-Gaussian observations, where some additional distributional assumptions on the noise and on the signal process are imposed, will be discussed.

3. Single-input–single-output systems

3.1. General case

In this section we focus our attention on systems where there are just two scalar variables. We set

$$\begin{aligned} \hat{z}_t &= (\hat{x}_t, \hat{y}_t)', & u_t &= (v_t, w_t)', \\ z_t &= \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \hat{x}_t \\ \hat{y}_t \end{pmatrix} + \begin{pmatrix} v_t \\ w_t \end{pmatrix}. \end{aligned} \quad (3.1)$$

Let f denote the spectral density of z_t . Using an obvious notation, we have [for $f_{\hat{x}}(\lambda) \neq 0$, see (ix)]

$$f = \begin{pmatrix} f_x & f_{xy} \\ f_{yx} & f_y \end{pmatrix} = \begin{pmatrix} f_{\hat{x}} + f_v & k^* f_{\hat{x}} \\ k f_{\hat{x}} & |k|^2 f_{\hat{x}} + f_w \end{pmatrix}, \quad (3.2)$$

where obviously

$$\tilde{f} = \begin{pmatrix} f_{\hat{x}} & k^* f_{\hat{x}} \\ k f_{\hat{x}} & |k|^2 f_{\hat{x}} \end{pmatrix}, \quad \tilde{f} = \begin{pmatrix} f_v & 0 \\ 0 & f_w \end{pmatrix},$$

and where k is evaluated on the unit circle and where k^* is the conjugate of k .

In order to guarantee the existence of a stationary solution, we assume:

(viii) $k(z)$ has no poles for $|z| = 1$.

In addition, for simplicity we assume:

(ix) $|k(e^{-i\lambda})| f_{\hat{x}}(\lambda) > 0$, $\lambda \in [-\pi, \pi]$.

Then we have [Anderson and Deistler (1984)]:

Theorem 3.1. *With the above assumptions, the set of all transfer functions k , satisfying*

$$\left| \frac{f_{yx}(\lambda)}{f_x(\lambda)} \right| \leq |k(e^{-i\lambda})| \leq \left| \frac{f_y(\lambda)}{f_{xy}(\lambda)} \right|, \quad (3.3a)$$

$$\arg k(e^{-i\lambda}) = \arg f_{yx}(\lambda), \quad (3.3b)$$

(where $\arg\{[ae^{i\varphi}] = \varphi\}$ is the set of all transfer functions k corresponding to given f).

Proof. Since \hat{f} in (3.2) is singular, we have

$$|f_{xy}|^2 = f_x f_y. \tag{3.4a}$$

In addition, \hat{f} satisfies

$$0 < f_x \leq f_w, \quad 0 < f_y \leq f_w, \tag{3.4b}$$

and (3.4a) and (3.4b), together with the conditions guaranteeing that \hat{f} is a spectral density matrix corresponding to a real process, are the only restrictions on \hat{f} .

Then every transfer function k compatible with given f satisfies (3.3) and conversely every transfer function k (with real coefficients K_j) satisfying (3.3a) and (3.3b) given rise to an EV system compatible with given f . Clearly, $\hat{f} = f - \hat{f}$ holds.

Note that rationality [i.e., assumption (v)] is not needed for Theorem 3.1. This theorem is a generalization of a well-known result for the static case [i.e., where k is constant and (x_t) , (v_t) , and (w_t) are white noise] – see, e.g., Frisch (1934) – which states that the set of all feasible slope parameters is given by the two ‘elementary’ regressions and that the sign of the slope parameters is unique. In (3.3a), analogously, the bounds for $|k(e^{-i\lambda})|$ correspond to the best linear (least squares) filters for (y_t) given (x_t) and for (x_t) given (y_t) , respectively, i.e., to the two extreme cases where all noise is attributed to (y_t) or to (x_t) , respectively.

Note that the solutions (3.3), in general, are far from unique. Given one $k(e^{-i\lambda})$ satisfying (3.3), we can obtain another one through multiplication by $n(\lambda)$, where n is positive and obeys the inequality constraints

$$\left| \frac{f_{xy}}{f_x k} \right| \leq n \leq \left| \frac{f_y}{f_{xy} k} \right|. \tag{3.5}$$

Clearly, once k has been uniquely determined, also f_x , f_y , and f_w are unique for given f .

In the case $n = 2$, the problem of finding the maximum number m^* of linear independent linear relations is trivial, since $m^* \geq 1$ always holds since a decomposition (2.7) always exists and since $m^* < 2$ holds when f is nondiagonal.

3.2. Causal solutions

We now turn our attention to questions relating to causality in this context. We will say that $k(z)$ is *causal* if it has a Taylor series expansion:

$$k(z) = \sum_{j=0}^{\infty} K_j z^j, \quad |z| \leq 1. \quad (3.6)$$

For the rational case (we consider here) this is equivalent to saying that $k(z)$ has no poles for $|z| \leq 1$. We say that $k(z)$ is *miniphase* if $k(z)$ has no zeros for $|z| < 1$.

Theorem 3.2. Let the earlier assumptions hold and let n_k denote the numbers of zeros of k inside the unit circle minus the number of poles in the same region. Then:

- (i) n_k is an invariant for all transfer functions in the equivalence class (3.3).
- (ii) A necessary condition for the existence of a causal transfer function in the equivalence class (3.3) is that $n_k \geq 0$ holds; a necessary condition for the existence of a miniphase transfer function is that $n_k \leq 0$ holds and $n_k = 0$ is a necessary condition for the existence of a transfer function that is both causal and miniphase.

Proof. Let k_1 and k_2 denote two transfer functions corresponding to the same f_{yx} . Then, using an obvious notation, we obtain from (3.2)

$$f_{yx} = k_1 f_{\tilde{x}_1} = k_2 f_{\tilde{x}_2}. \quad (3.7)$$

Since the $f_{\tilde{x}_i}$ are rational and positive [on $|z| = 1$], they are (when extended to \mathbb{C}) of the form

$$c \left(\prod_1^p (e^{-i\lambda} - z_j)(e^{i\lambda} - \bar{z}_j^{-1}) \right) / \left(\prod_1^q (e^{-i\lambda} - w_j)(e^{i\lambda} - \bar{w}_j^{-1}) \right), \quad (3.8)$$

for suitably chosen $|z_j| < 1$, $|w_j| < 1$; $c > 0$ and $p, q \in \mathbb{Z}_+$. Thus,

$$f_{\tilde{x}_2}^{-1} f_{\tilde{x}_1} = c p_1 \bar{p}_1 p_2^{-1} \bar{p}_2^{-1} z^{\delta p_2 - \delta p_1}, \quad c > 0, \quad (3.9)$$

where p_i , $i = 1, 2$, are polynomials satisfying

$$p_i(z) \neq 0, \quad |z| \leq 1, \quad p_i(0) = 1, \quad (3.10)$$

where δp_i denotes the degree of p_i and where, more general,

$$\tilde{p}_i(z) = p_i(z^{-1})z^{\delta_0},$$

with δ_0 being the multiplicity of the zero of p_i at $z = 0$. This proves (i).

From (3.7) and (3.9) we see that in the class of all transfer functions k corresponding to the same f_{xy} there is one which can be written as

$$k(z) = \bar{k}(z)z^{n_k}, \quad (3.11)$$

where \bar{k} is causal and miniphase and thus (ii) follows.

Remark 1. The theorem in its form above is given in Deistler (1986a, b). An analogous result, but for the *a priori* causal and not necessarily rational case, has been derived earlier by Anderson (1985a) as follows:

(i) Suppose that all transfer functions k considered satisfy

$$\sum_0^{\infty} |k_j| \rho^j < \infty \quad \text{for some } \rho > 1,$$

and that the change in argument of $f_{yx}(\lambda)$ as λ moves from 0 to 2π (which is just $-2\pi n_k$) is zero. Let \bar{k} be a causal and miniphase transfer function corresponding to f_{yx} . Then every (causal) k compatible with f is of the form:

$$k = c\bar{k}, \quad c > 0. \quad (3.12)$$

(ii) Under the same conditions as in (i), with the exception that now $n_k \geq 0$ holds, the set of all causal transfer functions compatible with given f_{yx} is given by

$$k(z) = cu(z)\bar{k}(z), \quad c > 0, \quad (3.13)$$

where

$$u(z) = \prod_1^{n_k} \frac{z - z_i}{z_i z - 1},$$

with z_i being arbitrary points in $|z| < 1$, save that complex points occur in complex conjugate pairs, and where \bar{k} is a causal and miniphase transfer function satisfying

$$\arg \bar{k}(e^{-i\lambda}) = \arg f_{yx}(\lambda) - \arg u(e^{-i\lambda}). \quad (3.14)$$

As has been pointed out in Anderson (1985a), (i), which says that a causal miniphase transfer function can be reconstructed from its phase up to a multiplicative position constant, is an old result [see Bode (1945)].

Remark 2. If only the information contained in f_{yx} (and not the additional information contained in f_x and f_y) is considered, then the conditions of Theorem 3.2 (ii) are necessary and sufficient for the existence of transfer functions with the postulated properties. This is a direct consequence of (3.11). The class of all transfer functions compatible with given f_{yx} is the set of all elements k of the form

$$k = cv\hat{k}, \quad (3.15)$$

with \hat{k} being the left-hand side in (3.11) and cv ($c < 0$) being any rational function of the form (3.8).

Thus, by Theorem 3.2, the integer n_k which can be estimated from the data contains the information about the 'causality status' of the equivalence class. Clearly, e.g., for the case $n_k = 0$, in general (without an *a priori* causality assumption), the class of all transfer functions compatible with given f_{yx} will contain also transfer functions which are neither causal nor miniphase.

When k is *a priori* known to be causal, then the class [defined by (3.15)] reduces to the set of all elements of the form

$$k = cp\bar{p}\hat{k}z^{-\delta p}, \quad (3.16)$$

where $c > 0$ and p is an arbitrary polynomial satisfying (3.10), and δp , the degree of p , satisfies $\delta p \leq n_k$ [(3.16) is the same as (3.13)]. In the case $n_k = 0$ then the (f_{yx}) equivalence class can be described by one parameter (namely c) only. For $n_k > 0$, $n_k + 1$ parameters are needed. This shows a vast reduction in complexity by an *a priori* causality assumption.

Now let us consider the additional restrictions on the equivalence classes coming from the knowledge of f_x and f_y . In the simplest case, where k is *a priori* known to be causal and $n_k = 0$, the set of feasible constants c in (3.12), is given by

$$\max_{\lambda \in [-\pi, \pi]} \left| \frac{f_{xy}(\lambda)}{f_x(\lambda)\bar{k}(e^{-i\lambda})} \right| \leq c \leq \min_{\lambda \in [-\pi, \pi]} \left| \frac{f_y(\lambda)}{f_{xy}(\lambda)\bar{k}(e^{-i\lambda})} \right|. \quad (3.17)$$

Note that, if f is constant and we assume *a priori* causality, then k must be constant and (3.12) and (3.17) give the classical result [Frisch (1934)] of the static case.

For the general case (3.15), the constant c in cv has to satisfy

$$\begin{aligned} & \max_{\lambda \in [-\pi, \pi]} \left| \frac{f_{xy}(\lambda)}{f_x(\lambda)v(e^{-i\lambda})\hat{k}(e^{-i\lambda})} \right| \\ & \leq c \leq \min_{\lambda \in [-\pi, \pi]} \left| \frac{f_y(\lambda)}{f_{xy}(\lambda)v(e^{-i\lambda})\hat{k}(e^{-i\lambda})} \right|. \end{aligned} \quad (3.18)$$

It is possible that for given v (and \hat{k}), (3.18) cannot be satisfied for any c , i.e., that certain rational functions v in (3.15) are ruled out completely. We have not been able to give an *analytical* description of the class of all k compatible with given f .

Remark 3. Clearly [under (ix)] we are free to choose either \hat{x}_t or \hat{y}_t as the input. Note that, if k is both causal and miniphase, then also \hat{y}_t can be regarded as causing \hat{x}_t through

$$\hat{x}_t = k^{-1}(z)\hat{y}_t. \quad (3.19)$$

If k is miniphase, but not causal, then \hat{y}_t causes \hat{x}_t , but not conversely.

Remark 4. There are algorithms for estimating causal miniphase transfer functions from phase data [see, e.g., Bode (1945) and Guillemin (1957)]. These algorithms assume rationality. In the nonrational case, algorithms are still available, but difficult continuity questions arise [see Anderson (1985b)].

3.3. Conditions for identifiability

Hitherto, we have presented no results which guarantee identifiability. In this subsection we aim to indicate how further assumptions can be imposed which do ensure identifiability of the transfer function k , i.e., uniqueness of k for given f . The reader is referred to Maravall (1979), Söderström (1980), Nowak (1983), Anderson and Deistler (1984), and Solo (1986).

By the rationality assumption (v) we write

$$\begin{aligned} f_{\hat{x}} &= e^{-1}d\alpha_d d^* e^{-1*}, \\ f_v &= c^{-1}h\sigma_v h^* c^{-1*}, \\ f_w &= f^{-1}g\sigma_w g^* f^*, \\ k(z) &= a^{-1}(z)b(z), \end{aligned} \quad (3.20)$$

where, e.g., $d^*(z) = d(z^{-1})$ and where the spectral densities as, e.g., f_x are defined on \mathbb{C} rather than on $[-\pi, \pi]$. [Note that a rational extension from the unit circle to the complex plane is unique.]

We assume that each of the polynomial pairs (a, b) , (d, e) , (c, h) , and (f, g) is relatively prime, $a(0) = d(0) = e(0) = c(0) = f(0) = g(0) = h(0) = 1$; $d, e, c, h, f,$ and g are all nonzero in $|z| \leq 1$. These assumptions are costless in a certain sense. In addition, let us assume that $a(z)$ is nonzero for $|z| \leq 1$ [i.e., that $k(z)$ is causal] and $\sigma_e > 0$, $\sigma_v > 0$, and $\sigma_w > 0$. Note that we do not make a miniphase assumption regarding $k(z)$. Let us make the definition

$$b(z) = b^+(z)b^-(z),$$

with $b^-(z) \neq 0$ for $|z| \geq 1$, $b^-(0) = 1$ and $b^+(z) \neq 0$ for $|z| < 1$. Then [see Anderson and Deistler (1984)]:

Theorem 3.3. With the previous assumptions, in particular if k is a priori known to be causal, the following conditions guarantee identifiability:

- (a) b^+ and b^{-*} have no common zeros.
- (b) a and e have no common zeros.
- (c) d and b^{-*} have no common zeros.
- (d) d and c have no common zeros.
- (e) $\delta d > 0$

If causality is not assumed a priori, the following conditions must be added:

- (f) With a^+, a^- defined analogously to b^+, b^- , there are no common zeros of a^+ and a^{-*} .
- (g) d and b^+ have no common zeros.
- (h) a^{-*} and e have no common zeros.

Proof. We shall establish the first part of the theorem only. For the remainder see Anderson and Deistler (1984).

We use the fact that

$$f_{yx} = kf_x = \frac{b}{a} \frac{e}{d} \frac{e^*}{d^*} \sigma_e. \quad (3.21)$$

By condition (c), d^* , and so d , is determined from the poles of f_{yx} in $|z| < 1$. Now

$$df_{yx}d^* = \frac{b^+b^-ee^*}{a} \sigma_e. \quad (3.22)$$

The root positions of a , b^+ , b^- , and e , the coprimeness of (a, b) assumed in connection with (3.20) and condition (b) now guarantee that we can find a . So we have now

$$adf_{yx}d^* = b^+b^-ee^*\sigma_e. \quad (3.23)$$

The zeros of the right side fall into two groups, those of b^+e which lie in $|z| \geq 1$ and those of b^-e^* which lie in $|z| < 1$. Condition (a) ensures that any zero of b^+e whose reciprocal is not a zero of b^-e^* is a zero of b^+ , and any zero whose reciprocal is a zero of b^-e^* is a zero of e . Hence b^- and e are uniquely determined and b^+ and σ_e are unique up to multiplication by constants. From the equation

$$df_x d^* = e\sigma_e e^* + dc^{-1}h\alpha_e h^* c^{-1} d^*,$$

then this constant can be uniquely determined due to (d) and (e) which establishes the result.

The theorem above to this point has not postulated special knowledge concerning the noise process perturbing the output y_t . By assuming it to be white, together with certain other conditions, we have identifiability. Let gcd denote the greater common divisor.

Theorem 3.4. With previous assumptions, suppose that the output noise process w_t is white, i.e., $f^{-1}g = 1$, that the input noise process is moving average, i.e., $c = 1$, and that $e(z)/\text{gcd}[a(z), e(z)]$ has positive degree. Then the system is identifiable.

The proof of this result, which is somewhat messy, can be found in Anderson and Deistler (1984).

The identifiability results of this subsection, taken with the result of the previous subsection, indicate that nonidentifiability is the rule in the absence of special assumptions, and this throws up an important, for the moment unanswered, question. Suppose that one of the sets of identifiability conditions of this section is not quite, but almost satisfied; for example, a process which should be white has an almost flat spectrum. Is it then the case that the set of transfer functions identified by the method of section 3.2 is a thin set and that, as the almost white process approaches a white process, the set of transfer functions approaches a single transfer function? Should such a result not hold, one would have to regard the unique identifiability results of this section as nonrobust and not suited to practical application.

4. The case of three variables

In this section, we consider the case $n = 3$. Let

$$z_t = \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{pmatrix}, \quad \hat{z}_t = \begin{pmatrix} \hat{z}_{1,t} \\ \hat{z}_{2,t} \\ \hat{z}_{3,t} \end{pmatrix}, \quad u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix},$$

and

$$f = (f_{ij})_{i,j=1,2,3}, \quad \hat{f} = \begin{pmatrix} \hat{f}_{11} & f_{12} & f_{13} \\ f_{21} & \hat{f}_{22} & f_{23} \\ f_{31} & f_{32} & \hat{f}_{33} \end{pmatrix}, \quad \tilde{f} = \text{diag } \tilde{f}_i.$$

It is obvious that $m^* = 3$ is equivalent to f being diagonal. So the key is to distinguish the possibilities $m^* = 1$ and $m^* = 2$. We concentrate on the generic case, where $f(\lambda)$ is positive definite and all $f_{ij}(\lambda)$ are nonzero (for all λ). Then we have [Anderson and Deistler (1987)]:

Theorem 4.1. Suppose that $f(\lambda)$ is positive definite and that $f_{ij}(\lambda) \neq 0$ (for all λ). Then the following statements are equivalent:

- (i) $m^* = 2$.
- (ii) $f_{12}f_{23}f_{31}$ is real and positive and $f_{ii} \geq |f_{ij}| |f_{ik}| |f_{jk}|^{-1}$ for all $i \neq j \neq k$.
- (iii) With $s = (s_{ij}) = f^{-1}$, $s_{12}s_{23}s_{31}$ is real and negative

In addition, in the case $m^* = 2$, \hat{f} is uniquely determined.

Remark 5. We wish to indicate the connection between this result and the result of Kalman (1982) for the static case (where f , \hat{f} , and \tilde{f} are real and constant) and for general n . As shown in Kalman (1982), for $n = 3$, $m^* = 2$ corresponds to $s_{12}s_{23}s_{31} < 0$ and $m^* = 1$ corresponds to $s_{12}s_{23}s_{31} > 0$. The point of the theorem is that, once one permits complex f , then $\text{Im}(s_{12}s_{23}s_{31}) \neq 0$ always implies $m^* = 1$.

Proof. '(i) \Rightarrow (ii)': Since \hat{f} is singular, we have

$$0 = \det \hat{f} = \hat{f}_{11}\hat{f}_{22}\hat{f}_{33} - \hat{f}_{11}|f_{23}|^2 - \hat{f}_{22}|f_{13}|^2 - \hat{f}_{33}|f_{12}|^2 + 2 \text{Re}(f_{12}f_{23}f_{31}). \quad (4.1)$$

Suppose $m^* = 2$. Then there exists a decomposition (2.7) in which \hat{f} has rank

1, so that $\hat{f}_{ii}\hat{f}_{jj} = |f_{ij}|^2$, $i \neq j$. Consequently, \hat{f} is uniquely determined by

$$\hat{f}_{ii} = |f_{ij}| |f_{ik}| / |f_{jk}|, \quad i \neq j \neq k, \quad (4.2a)$$

$$\hat{f}_{ij} = f_{ij}, \quad i \neq j. \quad (4.2b)$$

Inserting (4.2a) in (4.1) then yields

$$|f_{12}| |f_{23}| |f_{31}| = \operatorname{Re}(f_{12}f_{23}f_{31}). \quad (4.3)$$

Equivalently, $f_{12}f_{23}f_{31}$ is real and positive. The second condition in (ii) is immediate for (4.2b).

'(ii) \Rightarrow (i)': If $f_{12}f_{23}f_{31} > 0$, then we define \hat{f} by (4.2). As is easily seen, \hat{f} is nonnegative definite and satisfies (4.2). If in addition $\hat{f}_{ii} \leq f_{ii}$, then it gives a decomposition (2.7). By construction, \hat{f} has rank less than 2; thus $m^* \geq 2$, since \hat{f} is not diagonal, we have $m^* = 2$.

'(i) \Rightarrow (iii)': If $m^* = 2$, then there is an \hat{f} of rank 1. Under our assumptions, without restriction of generality, then

$$\hat{z}_{i,i} = k_i(z) \hat{z}_{1,i}, \quad i = 2, 3,$$

where $\hat{z}_{i,i}$ is the i th component of \hat{z}_i . Thus,

$$\begin{aligned} s_{12} \det f &= f_{13}f_{32} - f_{12}f_{33} \\ &= (|k_3|^2 k_2^* - k_2^* [|k_3|^2 + \varepsilon_3]) f_{11}^2 \\ &= -k_2^* \varepsilon_3 f_{11}^2. \end{aligned}$$

$$s_{31} \det f = -k_3 \varepsilon_2 f_{11}^2,$$

$$s_{32} \det f = -k_2 k_3^* \varepsilon_1 f_{11}^2,$$

where $\varepsilon_i = f_{ii} - \hat{f}_{ii} \geq 0$ and this implies (iii).

The last statement of the theorem follows immediately from (4.2).

Remark 6. Theorem 4.1 as well as its proof can be extended from the case $n = 3$ to the case of general n in a straightforward way in order to obtain a characterization for the case $m^* = n - 1$ (i.e., one input, $n - 1$ outputs); see Anderson and Deistler (1988). For the static case at least the first equivalence then is well known for a long time; see, e.g., Bekker and de Leeuw (1987).

For a discussion of questions related to causality, the reader is referred to Anderson and Deistler (1987).

5. Vector-input-vector-output systems with uncorrelated input and output noise

In the previous section, we extended the ideas of section 3 in that we expanded the two scalar variables of section 3 to three, maintaining assumption (vi), i.e., uncorrelatedness of the noise processes. In this section we make an expansion in a different direction: We postulate that \hat{z}_t is *a priori* known to be partitioned into an input variable \hat{x}_t and an output variable \hat{y}_t (but now \hat{x}_t and \hat{y}_t can be vectors). In the nonrational spectrum case, they must be of the same dimension for much of the theory to be valid. In the rational spectrum case, they can be of different dimension. Representation of the results is, however, simplified by assuming they are of the same dimension. In this section we assume:

(x) x_t and y_t have the same dimension.

Assumption (vi) is replaced by (vii). Assumption (ix) is generalized (using an obvious notation) to:

(xi) $f_{xy}(\lambda) = k^*(e^{-i\lambda})f_x(\lambda)$ is nonsingular for all $\lambda \in [-\pi, \pi]$.

Thus, we can write

$$f = \begin{pmatrix} f_x & f_{xy} \\ f_{yx} & f_y \end{pmatrix} = \begin{pmatrix} f_x + f_w & f_x k^* \\ k f_x & k f_x k^* + f_w \end{pmatrix} \quad (5.1)$$

Here the asterisk (*) denotes the conjugate transpose or, if the spectral densities are extended to \mathbb{C} , $k^*(z) = k(z^{-1})$. The task is to unravel this matrix and, in particular, the transfer function matrix k . This problem is investigated in Green and Anderson (1986) and Picci and Pinzoni (1986).

If we do not impose a causality constraint, there is not a great deal we can say. Nevertheless, formal results can be stated.

Theorem 5.1. Under the above assumptions, k can be written as

$$k = f_{yx} [f_x - A]^{-1}, \quad (5.2)$$

where $A(\lambda)$ is nonnegative Hermitian and satisfies $A(\lambda) = A(-\lambda)'$ and

$$f_x - f_{xy} f_y^{-1} f_{yx} \geq A \geq 0 \quad \text{for all } \lambda \in [-\pi, \pi]. \quad (5.3)$$

Conversely, if A is nonnegative Hermitian and satisfies $A(\lambda) = A(-\lambda)$ and (5.3) and if k is well defined by (5.2), then it gives rise to a feasible decomposition (5.1).

Proof. Suppose (5.1) holds. Identify $A = f_o$, then (5.2) is immediate from (5.1), while (5.3) is a consequence of nonnegative definite hermitian character of

$$\hat{f}_+ \begin{pmatrix} 0 & 0 \\ 0 & f_w \end{pmatrix} = \begin{pmatrix} f_x - A & f_{xy} \\ f_{yx} & f_y \end{pmatrix}.$$

The converse follows easily by taking $f_o = A$.

To make further progress, a causality assumption is needed. However, the idea of exploiting the phase information, as was done in the scalar case, is not possible here, and a different tool is needed.

This tool is a form of Wiener-Hopf factorization [see Clancey and Gohberg (1981)]. The key result is that any square matrix function $F(e^{-i\lambda})$, nonsingular for $\lambda \in [-\pi, \pi]$, can be factored as follows:

$$F(e^{-i\lambda}) = F_+(e^{-i\lambda}) \text{diag}\{(e^{-i\lambda})^{\mu_i}\} F_-(e^{-i\lambda}). \quad (5.4)$$

Here $F_+(e^{-i\lambda})$ is the restriction to the unit circle of a matrix function $F_+(z)$ which is analytic together with its inverse in $|z| \leq 1$ and $F_-(e^{-i\lambda})$ is the value on the unit circle of a matrix function $F_-(z)$ which is analytic together with its inverse in $|z| \geq 1$. The quantities μ_i are integer with $\mu_i \geq \mu_{i+1}$, and they are uniquely determined by $F(e^{-i\lambda})$. While F_+ and F_- are not unique, the family of F_+ and F_- which can appear in (5.4) is finitely parametrized. In case all μ_i are zero, F_+ is determined to within right multiplication by a nonsingular constant matrix. An algorithm for the factorization is available; see Clancey and Gohberg (1981). As in the scalar case, minimum phase problems are the easiest:

Theorem 5.2. With assumptions as above, suppose further that $k(z)$ is causal [i.e., $k(z)$ has no pole for $|z| < 1$] and minimum phase [i.e., $k^{-1}(z)$ is causal or $k(z)$ is nonsingular in $|z| \leq 1$]. Then $f_{yx}(\lambda)$ has a generalized Wiener-Hopf factorization with $\mu_i = 0$. Conversely suppose that f_{yx} has such a factorization:

$$f_{yx}(\lambda) = F_+(e^{-i\lambda}) F_-(e^{-i\lambda}), \quad (5.5)$$

and that k is causal. Then $k(e^{-i\lambda})$ is necessarily minimum phase and given for

some positive definite constant Hermitian M by

$$k(e^{-i\lambda}) = F_+(e^{-i\lambda})MF_-^{-1*}(e^{-i\lambda}), \quad (5.6)$$

where

$$M \leq F_+^{-1}(e^{-i\lambda})f_y(\lambda)F_+^{-1*}(e^{-i\lambda}), \quad (5.7a)$$

$$M^{-1} \leq F_-^{-1*}(e^{-i\lambda})f_x(\lambda)F_-^{-1}(e^{-i\lambda}), \quad (5.7b)$$

for all λ .

Before proving the theorem, we remark that M is the generalization of the positive constant c of section 3; the equality (5.6) parallels (3.8) and the inequalities (5.7a) and (5.7b) parallel (3.17).

Proof of Theorem 5.2. By assumption (xi), f_{yx} and thus f_x and k are nonsingular on $|z|=1$. Hence we can factorize $f_x(\lambda) = \theta(e^{-i\lambda})\theta^*(e^{-i\lambda})$ for some causal, minimum phase $\theta(z)$. Then,

$$f_{yx}(\lambda) = [k(e^{-i\lambda})\theta(e^{-i\lambda})]\theta^*(e^{-i\lambda}) \quad (5.8)$$

gives a decomposition of the form of (5.4), with $F_+ = k\theta$, $F_- = \theta^*$, $\mu_+ = 0$. This proves the first part of the theorem. Now suppose that (5.4) holds with $\mu_+ = 0$. Then,

$$F_+(e^{-i\lambda})F_-(e^{-i\lambda}) = [k(e^{-i\lambda})\theta(e^{-i\lambda})]\theta^*(e^{-i\lambda}). \quad (5.9)$$

It is slightly nontrivial to conclude from this equation that k is minimum phase. Take determinants on both sides and compute the change in argument as λ changes from 0 to 2π . For $\det F_+$, $\det F_-$, $\det \theta$, and $\det \theta^*$ it is necessarily zero. Hence it is zero for $\det k$. Since k is causal, this means k is also minimum phase. Then $k\theta$ is causal and minimum phase, and we have two generalized Wiener-Hopf factorizations of f_{yx} . Now, since F_+ is unique to within right multiplication by a nonsingular constant matrix, for some nonsingular H there holds

$$F_+(e^{-i\lambda})H = k(e^{-i\lambda})\theta(e^{-i\lambda}),$$

$$H^{-1}F_-(e^{-i\lambda}) = \theta^*(e^{-i\lambda}),$$

whence

$$k(e^{-i\lambda}) = F_+(e^{-i\lambda})HH^*F_-^{-1*}(e^{-i\lambda}).$$

Identify M with HH^* . Observe also that (5.7a) is equivalent to

$$0 \leq f_y - F_+ M F_+^* = f_y - f_{yx} k^* = f_w,$$

while (5.7b) is equivalent to

$$0 \leq f_x - F_-^* M^{-1} F_- = f_x - k^{-1} f_{yx} = f_v.$$

Now let us turn our attention to causal solutions which are not necessarily minimum phase. It is comparatively easy to present a necessary condition on the factorization indices of $f_{yx}(\lambda)$.

Theorem 5.3. *With assumptions as above, suppose that $k(z)$ is causal. Then $f_{yx}(\lambda)$ has a generalized Wiener-Hopf factorization in which all factorized indices are nonnegative.*

Proof. From (5.8), we have

$$\begin{aligned} f_{yx}(z) &= k(z)\theta(z)\theta'(z^{-1}) \\ &= F_+(z)\text{diag}(z^{\mu_i})F_-(z), \end{aligned}$$

where $f_{yx}(z)$ corresponds to the extension of $f_{yx}(\lambda)$ from the unit circle to the complex plane and

$$\text{diag}(z^{\mu_i})[F_-(z)\theta^{-1}(z^{-1})] = F_+^{-1}(z)k(z)\theta(z).$$

The right side is causal, i.e., has no poles in $|z| \leq 1$. Also, $F_-(z)\theta^{-1}(z^{-1})$ has no poles or zeros in $|z| \geq 1$. Suppose $\mu_i < 0$ and consider the i th row on the left side. Then examination of the left side shows that any poles must be in $|z| < 1$. Since the right side indicates all poles lie in $|z| > 1$, the row must be constant. Letting $z \rightarrow \infty$ on the left side shows it is zero. Hence the i th row of $F_+^{-1}(z)k(z)\theta(z)$ is zero, which contradicts the nonsingularity of $f_{yx}(\lambda)$.

Of course, the factorization of $f_{yx}(z)$ is the key to writing down all causal solutions $k(z)$ and, if $\mu_i < 0$ for some i , there can be no causal solution. In the nonminimum phase scalar case, there was a finite-dimensional parametrization of all solutions. This is also true in the matrix case. The result is as follows [see Green and Anderson (1986) for a proof]:

Theorem 5.4. *With assumptions as above, suppose that*

$$f_{yx}(z) = F_+(z)\text{diag}(z^{\mu_i})F_-(z) \quad (5.10)$$

is a generalized Wiener-Hopf factorization of $f_{yx}(z)$ with $\mu_i \geq 0$ for all i . Let $H(\mu)$ denote the set of matrices $H_-(z)$ for which H_- is analytic together with its inverse in $|z| \leq 1$ and $\text{diag}(z^{\mu_i})H_-(z)$ is analytic in $|z| \leq 1$. Then this set is

finitely parameterized by no more than $(n/2)(\sum \mu_i + n/2)$ parameters, where $f_{yx}(z)$ is $n/2 \times n/2$. Moreover, $k(z)$ is a causal solution of the EV problem if and only if

$$k(e^{-i\lambda}) = F_+(e^{-i\lambda}) \text{diag}\{(e^{-i\lambda})^{\mu_i}\} H_-(e^{-i\lambda}) H_-^*(e^{-i\lambda}) F_-^{-1}(e^{-i\lambda}), \quad (5.11)$$

with

$$H_-(e^{-i\lambda}) H_-^*(e^{-i\lambda}) \leq \text{diag}\{(e^{-i\lambda})^{-\mu_i}\} F_+^{-1}(e^{-i\lambda}) f_y(\lambda) F_+^{-1*}(e^{-i\lambda}) \\ \times \text{diag}\{(e^{i\lambda})^{-\mu_i}\}, \quad (5.12a)$$

$$[H_-(e^{-i\lambda}) H_-^*(e^{-i\lambda})]^{-1} \leq F_-^{-1*}(e^{-i\lambda}) f_x(\lambda) F_-^{-1}(e^{-i\lambda}). \quad (5.12b)$$

Note that $H_- H_-^*$ is a frequency dependent generalization of the constant positive definite matrix M of Theorem 5.2. Compare (5.6) and (5.11) as well as (5.7) and (5.12). Notice also for (5.11) that all the nonminimum phase zeros of $k(e^{-i\lambda})$ are contained in the zeros of $\det H_-(e^{-i\lambda})$; this means that the free parameters affect, as in the scalar case, these nonminimum phase zeros.

Finally, we remember that it is possible to specialise the theorem to the case of scalar $f_{yx}(z)$ and in this way the results to those obtained in section 3.

6. The non-Gaussian case: The use of higher-order moments

Up to now, our analysis has been based on the knowledge of the second moments of the observations only. For the static case, Geary (1942) and Reiersøl (1950) showed that non-Gaussian observations, e.g., higher-order moments of the observations, may contain sufficient information to guarantee identifiability [see also Bekker (1986) for the static multivariable case]. This analysis has first been extended to the dynamic case by Akaike (1966). [See also Deistler (1986a).]

For simplicity we assume the existence of higher-order (moments and of higher order) cumulant spectra. As has been shown in Zeisel (1986), Reiersøl's (1950) analysis, which is based on characteristic functions and which does not assume existence of moments, can be extended to the dynamic case. Since the extension of our subsequent analysis to the multivariable case is straightforward, for simplicity we consider here the case $n = 2$.

Let us repeat that, e.g., the n th-order cumulant corresponding to $\hat{y}_{t_1}, \dots, \hat{y}_{t_p}, \hat{x}_{t_1}, \dots, \hat{x}_{t_{p-1}}, \hat{x}_0$, $\text{cum}(\hat{y}_{t_1}, \dots, \hat{y}_{t_p}, \hat{x}_{t_1}, \dots, \hat{x}_{t_{p-1}}, \hat{x}_0)$ say, is the coefficient of $(i)^p s_1, \dots, s_p$ in the Taylor series expansion of $\log E \exp(i(\hat{y}_{t_1} s_1 + \dots + \hat{x}_0 s_p))$; see, e.g., Brillinger (1981). If

$$\sum_{t_1, \dots, t_{p-1} = -\infty}^{\infty} |\text{cum}(\hat{y}_{t_1}, \dots, \hat{x}_{t_{p-1}}, \hat{x}_0)| < \infty$$

holds, then the corresponding p th-order cumulant spectrum exists and is given by

$$\begin{aligned}
 & f_{\hat{y}^p}(e^{i\lambda}) (\lambda_1, \dots, \lambda_{p-1}) \\
 &= (2\pi)^{-p+1} \sum_{t_1, \dots, t_{p-1} = -\infty}^{\infty} \text{cum}(\hat{y}_{t_1}, \dots, \hat{y}_{t_p}, \hat{x}_{t_{p-1}}, \dots, \hat{x}_{t_1}, \hat{x}_0) \\
 & \quad \times \exp\left\{-i \sum_{j=1}^{p-1} \lambda_j t_j\right\}. \tag{6.1}
 \end{aligned}$$

We will impose the following additional assumptions:

- (xii) (\hat{z}_t) and (u_t) are strictly stationary and mutually independent processes. All corresponding moments and cumulant spectra up to order p , where p is sufficiently large, exist.

As is easily seen, due to linearity and continuity of cumulants with respect to one variable (when the others are kept constant) we obtain from (6.1)

$$f_{\hat{y}^p}(e^{i\lambda}) (\lambda_1, \dots, \lambda_{p-1}) = k(e^{-i\lambda_1}) f_{\hat{y}^p}(e^{i\lambda}) (\lambda_1, \dots, \lambda_{p-1}). \tag{6.2}$$

Furthermore, from the properties of cumulants, we obtain (using an obvious notation)

$$\begin{aligned}
 & f_{\hat{y}^p}(e^{i\lambda}) (\lambda_1, \dots, \lambda_{p-1}) \\
 &= f_{w^p}(e^{i\lambda}) (\lambda_1, \dots, \lambda_{p-1}) + f_{w^p}(e^{i\lambda}) (\lambda_1, \dots, \lambda_{p-1}). \tag{6.3}
 \end{aligned}$$

Thus we obtain:

Theorem 6.1. Assume that the previous assumptions except for (vi) [or (vii)] hold. Then either

- (xiii.a) (u_t) is Gaussian

and

- (xiii.b) $f_{\hat{y}^p}(\lambda_1, \dots, \lambda_{p-1}) \neq 0$, $\lambda_1 \in [-\pi, \pi]$, for suitable $\lambda_2, \dots, \lambda_{p-1}$ and $p > 2$,

or

(xiv.a) (v_r) and (w_r) are (mutually) independent

and

(xiv.b) $f_{y^{(r-1)}x^{(r-1)}}(\lambda_1, \dots, \lambda_{p-1}) \neq 0$, $\lambda_1 \in \{-\pi, \pi\}$, for suitable $\lambda_2, \dots, \lambda_{p-1}$, $p > 2$, $r-1 > 0$, $p-r > 0$,

guarantee identifiability of k .

Proof. If (u_r) is Gaussian, then $f_{w_r^{(r-1)}} = 0$ for $p > 2$ and thus, due to (xiii.b), k is uniquely determined from (6.2) with $r = 1$. If (xiv.a) holds, then $f_{w_r^{(r-1)}} = 0$ for $r > 0$, $p-r > 0$. Thus, together with (xiv.b) again gives k uniquely from (6.2).

Remark 7. Evidently conditions (xiii.b) or (xiv.b) imply non-Gaussianity. In the non-Gaussian case they are fulfilled under rather general assumptions [Deistler (1986a)]. Of course, e.g., the assumption that (\hat{z}_r) is non-Gaussian whereas (u_r) is Gaussian will only be justified in special applications where this distributional information is available (which will rarely be the case in econometrics). In principle, (6.2) could be used to estimate k from the estimated cumulant spectra. However, estimation of higher-order cumulant spectra requires a large amount of data.

References

- Aigner, D.J. and A.S. Goldberger, eds., 1977. Latent variables in socio-economic models (North-Holland, Amsterdam).
- Aigner, D.J., C. Hsiao, A. Kapteyn, and T. Wansbeck, 1984. Latent variable models in econometrics, in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics (North-Holland, Amsterdam).
- Akaike, H., 1966. On the use of non-Gaussian process in the identification of a linear dynamic system. *Annals of the Institute of Statistical Mathematics* 18, 269-276.
- Anderson, B.D.O., 1985a. Identification of scalar errors-in-variables models with dynamics. *Automatica* 21, 709-716.
- Anderson, B.D.O., 1985b. The continuity of the spectral factorization operation. *Mathematica Applicada y Computational* 4, 2.
- Anderson, B.D.O. and M. Deistler, 1984. Identifiability in dynamic errors-in-variables models. *Journal of Time Series Analysis* 5, 1-13.
- Anderson, B.D.O. and M. Deistler, 1987. Dynamic errors-in-variables systems with three variables. *Automatica* 23, 611-616.
- Anderson, B.D.O. and M. Deistler, 1988. Identification of dynamic systems from noisy data. Mimeo. (Institute for Econometrics and Operations Research, Technical University, Vienna).
- Anderson, T.W., 1984. Estimating linear statistical relationships. *Annals of Statistics* 12, 1-45.
- Bekker, P., 1986. Comment on identification in the linear errors-in-variables model. *Econometrica* 54, 215-217.
- Bekker, P. and J. de Leeuw, 1987. The rank of reduced dispersion matrices. *Psychometrika* 52, 125-135.
- Bloomberg, H. and R. Yllinen, 1983. Algebraic theory for multivariate linear systems (Academic Press, London).
- Bode, H.W., 1945. Network analysis and feedback amplifier design (Van Nostrand, New York, NY).

- Brillinger, D.R., 1981, *Time series: Data analysis and theory*, Expanded ed. (Holden-Day, San Francisco, CA).
- Clancey, K. and I.C. Gohberg, 1981, Factorization of matrix functions and singular integral operators (Birkhäuser Verlag, Basel).
- Deistler, M., 1986a, Linear dynamic errors-in-variables models, in: J. Gani and M. Priestley, eds., *Essays in time series and allied processes*, Journal of Applied Probability, Special volume 23A, 23-39.
- Deistler, M., 1986b, Linear errors-in-variables models, in: S. Bittanti, ed., *Time series and linear systems*, Lecture notes in control and information sciences (Springer-Verlag, Berlin) 37-67.
- Engle, R. and C.W.J. Granger, 1987, Co-integration and error-correction: Representation, estimation and testing, *Econometrica* 55, 251-276.
- Frisch, R., 1934, Statistical confluence analysis by means of complete regression systems. Publication no. 5 (Economic Institute, University of Oslo, Oslo).
- Fuller, W.A., 1987, *Measurement error models* (Wiley, New York, NY).
- Geary, R.C., 1942, Inherent relations between random variables, *Proceedings of the Royal Irish Academy A* 47, 63-76.
- Geary, R.C., 1943, Relations between statistics: The general and the sampling problem when the samples are large, *Proceedings of the Royal Irish Academy A* 49, 177-196.
- Green, M. and B.D.O. Anderson, 1986, Identification of multivariable errors-in-variables models with dynamics, *IEEE Transactions on Automatic Control* AC-31, 467-471.
- Guillemin, E.A., 1957, *Synthesis of passive networks* (Wiley, New York, NY).
- Kalman, R.E., 1982, System identification from noisy data, in: A. Bednarek and L. Cesari, eds., *Dynamical systems II*, a University of Florida international symposium (Academic Press, New York, NY).
- Kalman, R.E., 1983, Identifiability and modeling in econometrics, in: P.R. Krishnaiah, ed., *Developments in statistics*, Vol. 4 (Academic Press, New York, NY).
- Klepper, S. and E. Leamer, 1984, Consistent sets of estimates for regressions with errors in all variables, *Econometrica* 52, 163-183.
- Madansky, A., 1959, The fitting of straight lines when both variables are subject to error, *Journal of the American Statistical Association* 54, 173-205.
- Maravall, A., 1979, *Identification in dynamic shock-error models* (Springer-Verlag, Berlin).
- Moran, P.A.P., 1971, Estimating structural and functional relationships, *Journal of Multivariable Analysis* 1, 232-255.
- Nowak, E., 1985, Global identification of the dynamic shock-error model, *Journal of Econometrics* 27, 211-219.
- Phillips, P.C.B. and S. Ouliaris, 1986, Testing for cointegration, Cowles Foundation discussion paper no. 809 (Princeton University, Princeton, NJ).
- Picci, G. and S. Pinzoni, 1986, Dynamic factor-analysis models for stationary processes, *IMA Journal of Mathematical Control and Information* 3, 185-210.
- Reiersøl, O., 1941, Confluence analysis by means of lag moments and other methods of confluence analysis, *Econometrica* 9, 1-24.
- Reiersøl, O., 1950, Identifiability of a linear relation between variables which are subject to error, *Econometrica* 18, 375-389.
- Sims, C.A., 1980, Macroeconomics and reality, *Econometrica* 48, 1-48.
- Schneeweiß, H. and H.J. Mittag, 1986, *Lineare Modelle mit fehlerbehafteten Daten* (Physica Verlag, Heidelberg).
- Söderström, T., 1980, Spectral decomposition with application to identification, in: F. Archetti and M. Dugiani, eds., *Numerical techniques for stochastic systems* (North-Holland, Amsterdam).
- Solo, V., 1986, Identifiability of time series models with errors-in-variables, in: F. Gani and M. Priestley, eds., *Essays in time series and allied processes*, Journal of Applied Probability, Special volume 23A, 63-71.
- Willems, J.C., 1979, System theoretic models for the analysis of physical systems, *Ricerca di Automatica* 10, 71-106.
- Willems, J.C., 1986, From time series to linear systems, Part I, *Automatica* 22, 561-580.
- Zeisel, H., 1986, Identifizierbarkeit der Transferfunktion in linearen dynamischen Fehler-in-den-Variablen Modellen im nicht-gauß'schen Fall, Diplomarbeit (Institut für Ökonometrie und Operations Research, Technische Universität, Wien).