IDENTIFIABILITY IN DYNAMIC ERRORS-IN-VARIABLES MODELS

B.D.O. Anderson[1] and M. Deistler[2]

1) Department of Systems Engineering, Research School of Physical Sciences, Australian National University, GPO Box 4, Canberra ACT 2601, Australia.

2) Department of Econometrics and Operations Research, University of Technology Vienna, Argentinierstr 8, A-1040 Vienna, Austria.

Abstract

This paper is concerned with the identifiability of scalar linear dynamic errors-in-variables systems. The analysis is based on second moments only. The set of feasible systems corresponding to given second moments of the observations is described and conditions for identifiability are derived for the case of rational transfer functions.

1. Introduction

In most cases the statistical theory of linear systems is concerned with stochastic models where the inputs are observed without errors. Models where both inputs and outputs are contaminated by errors are usually called errors-in-variables (EV) models. In many situations - especially if we are concerned with the analysis of the "true" system rather than with prediction of one observed process from another one - EV models provide a more adequate way of stochastic modelling, allowing for a symmetric treatment of all variables. This, together with the intrinsic difficulties of the statistical analysis of EV models, has recently been pointed out by Kalman (1981). Note that errors in measurements is only one special interpretation for the errors contaminating inputs and outputs respecitively.

We will only deal with a special case here, namely (mainly) with scalar linear dynamic errors in variables models and our analysis is based on second moments only.

It is well known, that special dynamic structures (eg in the input process) can help to guarantee identifiability in EV models (Reiersøl (1941)). Since then identifiability in dynamic EV models has been investigated by a number of authors (see eg Maravall (1979), Aigner, Hsiao, Kapteyn and Wansbeek (1982), Hsiao (1982), Wegge (1982)). In these investigations local identifiability is shown under certain assumptions using an inverse function theorem. According to our knowledge the only paper dealing with global identifiability in this context is Söderström (1980)

The systems considered are of the form:

$$y_t^* = w_1(z)x_t^* \qquad (t \in Z) \tag{1}$$

where z denotes the backward-shift operator (as well as a complex variable),

$$w_1(z) = \sum_{i=-\infty}^{\infty} W_1(i)z^i$$

and where (y_t^*) and (x_t^*) are the unobserved stationary output and input process respectively. The unobserved inputs are assumed to be a linearly regular stationary process with Wold-decomposition

$$x_t^* = w_3(z).\varepsilon_t \quad ; \quad w_3(z) = \sum_{i=0}^{\infty} W_3(i)z^i \tag{2}$$

and the observed inputs and outputs respectively, (x_t) and (y_t) say, are given by the "measurement error equations"

$$x_t = x_t^* + w_2(z)u \tag{3}$$

and

$$y_t = y_t^* + w_4(z)v_t \tag{4}$$

where

$$w_2(z) = \sum_{i=0}^{\infty} W_2(i)z^i \quad ; \quad w_4(z) = \sum_{i=0}^{\infty} W_4(i)z^i$$

and $w_2(z)u_t$ and $w_4(z)v_t$ are the Wold-decompositions of the "measurement-errors". The $W_j(i)$ are square summable (over i) for $j = 1...4$ and the $W_1(i)$ are assumed to be such that y_t^* exists in the mean square sense. The assumptions that (x_t^*), $w_2(z)u_t$ and $w_4(z)v_t$ are regular stationary processes are made for notational convenience and may easily be relaxed. The same holds for the assumption that $w_1(z)$ can be represented as $\Sigma W_1(i)z^i$. Note also, that the summation limits in defining $w_1(z)$ are set so as not to force $w_1(z)$ a priori to be causal at this point; causality would correspond to a summation running over the non-negative integers only.

Every system satisfying (1)-(4) is called an error-in-variables (EV) system. We will consider the scalar case only if the contrary is not stated explicitly. Throughout the paper we will assume:

(I) The joint process $(\varepsilon_t,u_t,v_t)'$ is white noise with a diagonal variance-covariance matrix i.e.

$$E(\varepsilon_t,u_t,v_t) = 0$$

$$E(\varepsilon_s,u_t,v_t)'(\varepsilon_t,u_t,v_t) = \delta_{st}\cdot\begin{pmatrix}\sigma_\varepsilon & 0 & 0\\ 0 & \sigma_u & 0\\ 0 & 0 & \sigma_v\end{pmatrix}.$$

where $\sigma_u,\sigma_v,\sigma_\varepsilon \geq 0$

0191-2216/83/0000-0317 $1.00 © 1983 IEEE

(II) $W_j(0) = 1 \quad ; \quad j = 2,3,4$

Note that we have no analogous assumption for $W_1(0)$. Also no assumption about the roots of $w_1(z)$ is made a priori.

(1)-(4) can be written as a linear system with two outputs and three inputs:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} w_1 w_3 & 0 & w_4 \\ w_3 & w_2 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ u_t \\ v_t \end{pmatrix} \tag{5}$$

Let f_x and f_y denote the spectral densities of (x_t) and of (y_t) respectively and let f_{xy} denote the corresponding crosspectrum. Then the relation between the second moments of the observed processes and the transfer functions w_i, $i = 1...4$ and the parameters $\sigma_\varepsilon, \sigma_u, \sigma_v$ is given (omitting a factor of $(2\pi)^{-1}$) by:

$$f_x = w_3 \cdot \sigma_\varepsilon \cdot w_3^* + w_2 \cdot \sigma_u \cdot w_2^* \tag{6}$$

$$f_{xy} = w_3 \cdot \sigma_\varepsilon \cdot w_3^* \cdot w_1^* \tag{7}$$

$$f_y = w_1 w_3 \sigma_\varepsilon w_3^* w_1^* + w_4 \sigma_v w_4^* \tag{8}$$

where the spectral densities are considered as being defined on C (except for the poles) rather than on $[-\pi,\pi]$ and where

$$(\Sigma P(i) z^i)^* = \Sigma P(i) z^{-i}$$

If f_{x^*}, f_{y^*}, g_x and g_y denote the spectral densities of $(x_t^*), (y_t^*), w_2(z)u_t$ and $w_4(z)v_t$ respectively then (6)-(8) can be written as

$$f_x = f_{x^*} + g_x \tag{9}$$

$$f_{xy} = f_{x^*} \cdot w_1^* \tag{10}$$

$$f_y = f_{y^*} + g_y \tag{11}$$

2. Nonparametric Analysis

Here we discuss the case where there are no additional rationality assumptions on the transfer functions.

First we consider the question of characterizing the triplets f_x, f_y, f_{xy} such that there exist w_i, $i = 1...4$, $\sigma_\varepsilon, \sigma_u, \sigma_v$ for which (6)-(8) holds.

Theorem 1: For every spectral matrix $\begin{pmatrix} f_x & f_{xy} \\ f_{yx} & f_y \end{pmatrix}$ (satisfying our general assumptions) there exist w_i, $i = 1...4$, $\sigma_\varepsilon, \sigma_u, \sigma_v$ satisfying (6)-(8).

Proof: A spectral matrix is non-negative definite, i.e.

$$f_x(e^{-i\lambda}) \geq 0, f_y(e^{-i\lambda}) \geq 0, f_x(e^{-i\lambda}) f_y(e^{-i\lambda}) - |f_{xy}(e^{-i\lambda})|^2 \geq 0 \tag{12}$$

(and integrable). Now let f_{x^*}, f_{y^*} be any two functions (which as functions on $[-\pi,\pi]$ have the property $f(\lambda) = f(-\lambda)$ such that

$$0 \leq f_{x^*}(e^{-i\lambda}) \leq f_x(e^{-i\lambda}) \tag{13}$$

$$0 \leq f_{y^*}(e^{-i\lambda}) \leq f_y(e^{-i\lambda}) \tag{14}$$

$$f_{x^*}(e^{-i\lambda}) \cdot f_{y^*}(e^{-i\lambda}) - |f_{xy}(e^{-i\lambda})|^2 = 0 \tag{15}$$

(and such that in addition f_{x^*}, $f_x - f_{x^*}$ and $f_y - f_{y^*}$ correspond to regular processes i.e. these densities satisfy (see e.g. Hannan 1970) (as functions on $[-\pi,\pi]$):

$$f \neq 0 \quad \lambda\text{-a.e.} \quad \text{and} \quad \int_{-\pi}^{\pi} \log f(\lambda) d\lambda > -\infty)$$

or (16)

$$f = 0 \quad \lambda\text{-a.e.}$$

We now define

$$w_1 = \begin{cases} \dfrac{f_{xy}^*}{f_{x^*}} & \text{for } f_{x^*} \neq 0 \quad \lambda\text{-a.e.} \\ 0 & \text{for } f_{x^*} = 0 \quad \lambda\text{-a.e.} \end{cases} \tag{17}$$

and $g_x = f_x - f_{x^*} = w_2 \sigma_u w_2^*$; $g_y = f_y - f_{y^*} = w_4 \cdot \sigma_v w_4^*$; $f_{x^*} = w_3 \sigma_\varepsilon w_3^*$ denote the usual spectral factorizations.

Remark 1: Note that the spectral factorizations corresponding to the Wold decomposition are unique in the scalar case if $\sigma > 0$ and thus if $\sigma_\varepsilon, \sigma_u, \sigma_v > 0$ there is a one-to-one relation between $(f_{x^*}, f_{y^*}, g_x, g_y)$ and $(w_1, w_2, w_3, w_4, \sigma_\varepsilon, \sigma_u, \sigma_v)$. If $\sigma_\varepsilon = 0$ then w_1 and w_3 may be chosen arbitrary.

The following result characterizing all possible EV-systems for given f_x, f_y, f_{xy} is an immediate consequence of the proof above:

Corollary: For given f_x, f_y, f_{xy} the set of all corresponding $f_{x^*}, f_{y^*}, g_x, g_y$ is given by (13)-(16).

Remark 2: Neglecting condition (16) (which is rather technical and of minor importance) the allowable ranges are shown in figures 1 and 2:

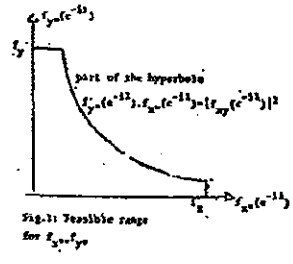

Fig.1: Possible range for f_{x^*}, f_{y^*}

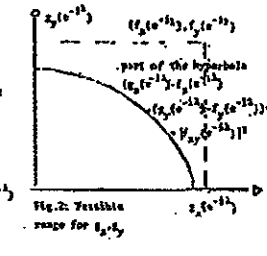

Fig.2: Possible range for f_{x^*}, f_{y^*}

From (13)-(15) we obtain

$$f_y(e^{-i\lambda}) \cdot f_{xy}^{-1}(e^{-i\lambda}) \geq w_1(e^{-i\lambda}) \geq f_{y^*}(e^{-i\lambda}) \cdot f_x^{-1}(e^{-i\lambda}) \tag{18}$$

where $a \geq b$, $a,b \in C$ means that there exists a $c \in [0,1)$ such that $ca = b$. (18) shows the analogy to the static case (i.e. $W_j(i) = 0$, $i \neq 0$, $j = 1...4$). As is well known, in this case

$$\hat{a} \geq W_1(0) \geq \hat{b}$$

where a and b are the OLS estimates of the regressions of y_t on x_t and of x_t on y_t respectively.

This shows an "essential nonuniqueness" (compare Kalman (1981)) for our problem in general which cannot be resolved without imposing further assumptions. Note that the extreme points of the parts of the hyperbolas shown in figures 1 and 2 correspond to the cases where either the inputs or the outputs are not subject to errors.

If $\begin{pmatrix} f_x & f_{xy} \\ f_{xy} & f_y \end{pmatrix}$ is singular then we must have $g_x = g_y \equiv 0$ (λ-a.e.) and thus $f_{x^*} = f_x$ and $f_{y^*} = f_y$ are unique.

Remark 3: If f_x, f_y, f_{xy} are perturbed, the set of f_{x^*}, f_{y^*}, g_x, g_y which are allowed is also perturbed, with the perturbation being continuous.

Remark 4: Theorem 1 can be extended to the multivariable case: If (6)-(8) are now interpreted as matrix equations (where * denotes the conjugate-transpose) then it is straight forward to show that there exist w_i, $i = 1 \ldots 4$ and $\sigma_\varepsilon, \sigma_u, \sigma_v$ satisfying these equations for every spectral matrix $\begin{pmatrix} f_x & f_{xy} \\ f_{yx} & f_y \end{pmatrix}$ (satisfying our general assumptions).

Remark 5: It should be stressed again, that so far we did not impose a causality requirement on $w_1(z)$ and thus we provided a symmetric treatment of (x_t) and (y_t). If we do a-priori know that $w_1(z)$ is causal, i.e. that $W_1(i) = 0$, $i < 0$, then we have additional a-priori restrictions, as (17) does not define a causal w_1 in general; and Theorem 1 and its corollary no longer hold true in general with the additional causality assumption imposed. We consider the following example: Let

$$f_x = \sigma_\varepsilon + \sigma_{\hat{u}} \quad , \quad \sigma_\varepsilon > \sigma >$$
$$f_{xy} = \sigma_\varepsilon(1+bz)^* \quad , \quad |b| < 1$$
$$f_y = |1+bz|^2\sigma_\varepsilon + \sigma_v$$

then, due to the corollary above, the set of all feasible f_{x^*} is given by (see Fig.3)

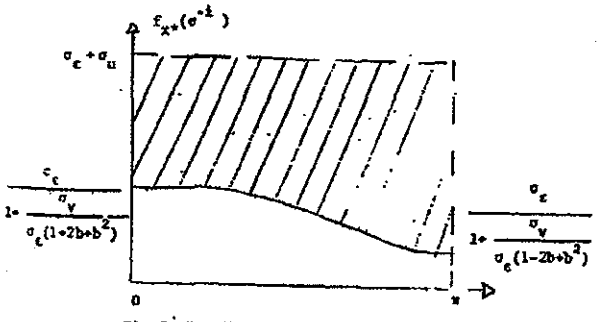

Fig.3: Feasible range for f_{x^*}

The feasible range of f_{x^*} contains e.g. as easily shown all rational spectral densities of the form

$$f_{x^*}(e^{-i\lambda}) = \frac{c_1 \prod_1^p (e^{-i\lambda} - z_j)\cdot \prod_1^p \overline{(e^{-i\lambda} - z_j)}}{c_2 \prod_1^q (e^{-i\lambda} - w_j)\cdot \prod_1^q \overline{(e^{-i\lambda} - w_j)}}$$

for arbitrary $|z_j| > 1, |w_j| > 1, p, q$ and suitable chosen c_1, c_2. From (17) we have

$$w_1^* = \frac{\sigma_\varepsilon(1+bz)^*}{ww^*} ;$$

where $f_{x^*} = ww^*$ corresponds to the Wold decomposition of f_{x^*}.

If we restrict ourselves to the case of rational f_{x^*} then it is clear that w_1 is causal if and only if either (x_t^*) is purely autoregressive or if (x_t^*) is an ARMA process with MA order equal to one and MA root being equal to $z_1 = -b$. In the non-rational case an analogous result holds.

This example also shows that, even if the true system is causal, then in general there may also be non-causal systems generating the same observed processes (as these noncausal systems are contained in the feasible set described in the corollary above). (This has been pointed out by Newbold (1978)). E.g. it may happen that the true system is causal whereas the best linear filter between the observed process is not. This of course very much complicates problems of testing for causality in EV models, we also see that even if the true transfer functions are rational, then in general the assumption of rationality of all transfer functions rules out feasible non-rational transfer functions and thus provides a restriction of generality.

3. Rational Transfer Functions: Conditions for Identifiability

We now consider the case where the transfer functions are assumed to be rational, i.e.: (we will often omit the argument z in complex functions)

$$w_1 = a^{-1}.b; \; w_3 = d^{-1}.e; \; w_2 = c^{-1}.h; \; w_4 = f^{-1}.g$$

where

$$a(z) = \sum_{i=0}^{na} A(i)z^i; \; b(z) = \sum_{i=0}^{nb} B(i)z^i; \; d(z) = \sum_{i=0}^{nd} D(i)z^i$$

$$e(z) = \sum_{i=0}^{ne} E(i)z^i; \; c(z) = \sum_{i=0}^{nc} C(i)z^i; \; h(z) = \sum_{i=0}^{nh} H(i)z^i$$

$$f(z) = \sum_{i=0}^{nf} F(i)z^i; \; h(z) = \sum_{i=0}^{nh} H(i)z^i$$

The assumption of rationality is imposed in a great number of applications, even if we do not a-priori know that the true transfer functions are rational. An advantage of rational transfer functions is, that they can approximate every transfer function with arbitrary accuracy and only a finite number of parameters is needed for description.

The proof of our theorems concerning identifiability often will not depend on the a-priori knowledge of the degrees at least of some of the polynomials considered. For estimation methods like the maximum likelihood

method however, the maximum degrees of a,...g, na...ng say, have to be prescribed and this is the reason that we fomulate our problem for given na,...,ng.

Note that the prescription of na,...,ng does not necessarily mean that the correct degrees are a-priori known. It is even possible that we may not insist that the observed processes are described in our model-class.

We impose the additional assumptions:

(III). a,b are relatively prime, and so are d,e and c,h and f,g.

(IV) $a(z)\neq 0\ |z|=1$; $d(z)\neq 0\ |z|\leq 1$; $e(z)\neq 0\ |z|<1$;
$c(z)\neq 0\ |z|\leq 1$; $h(z)\neq 0\ |z|<1$; $f(z)\neq 0\ |z|\leq 1$;
$g(z)\neq 0\ |z|<1$;

(V) $a(o)=d(o)=e(o)=c(o)=f(o)=g(o)=h(o)=1$; $\sigma_\varepsilon > 0$.

Again, assumptions corresponding to (IV) and (V) would not be justified for b(z), because we are directly interested in the transfer-function $w_1(z)$ of the system rather than in its output spectrum. This causes a part of the complications of our analysis.

If not explicitly mentioned here we will in addition impose the causality requirement

(VI) $a(z)\neq 0 |z|\leq 1$

In this case both $a(z)\neq 0|z|=1$ and $a(o)=1$ provide no restriction of generality; otherwise these assumptions are maintained for the sake of notational convenience.

The parameters of interest are θ = (A(1)...A(na), B(o),...B(nb), D(1),..,D(nd), E(1),..E(ne), C(1),..C(nc), H(1)...H(nh), F(1),..F(nf), G(1)...G(ng), $\sigma_\varepsilon,\sigma_u,\sigma_v$) and thus our parameter-space, say, is a subset of a Euclidian $\mathbb{R}^n$, where n = na+....+ng+4.

Two EV systems (two parameters $\theta_1,\theta_2\in\Theta$) are called observationally equivalent (with respect to their second moments[1]) if they generate the same spectral densities f_x, f_{xy} and f_y in (6)-(8). A class of EV systems (satisfying (I)-(VI) (or (I)-(V)) (a parameter-space Θ) is called identifiable if it contains no different observationally equivalent parameters.

Now let us investigate additional assumptions, under which identifiability can be obtained:

First let us note that in the case c = 1 (i.e. when the input errors are MA processes) d is uniquely determined from f_x: In this case the poles of f_x which are located outside the unit circle are the zeros of d and as d(o) = 1, d is uniquely determined from these poles.

If p is any polynomial we use δp for its actual degree (as opposed to the prescribed one).

For the sake of completeness let us introduce two cases which have already been investigated by Söderström (1980)

(i) If $c^{-1}.h=1$ (i.e. if the input errors are white noise) and if $\delta d > \delta e$ then we have identifiability: From (6) we obtain

$$df_x d^* = e\sigma_\varepsilon e^* + d\sigma_u d^* \qquad (19)$$

Since d is uniquely determined from f_x, a comparison of coefficients corresponding to power δd in (19) gives σ_u. Then e and σ_ε are determined from $(f_x - \sigma_u)$ by the usual factorization, a and b are uniquely determined from (7) and finally f,g and σ_v are uniquely determined from (8). The same analysis holds in the multivariable case.

(ii) If e = 1 and c = 1 (i.e. if x_t^* is purely autoregressive and if the input errors are of MA type) and if $\delta d > 0$ then we have identifiability: Here (6) is of the form

$$f_x = d^{-1}.\sigma_\varepsilon d^{-1*} + h.\sigma_u h^*$$

and thus

$$df_x d^* = \sigma_\varepsilon + dh\sigma_u h^* d^*$$

Let z_1 be a zero of d then $\sigma_\varepsilon = df_x d^*(z_1)$.h and σ_u are obtained from $f_x - d^{-1}\sigma_\varepsilon d^{-1*}$ and the rest follows as before.

The same argument also holds for d = 1, h = 1, $\delta c > 0$.

The common feature of these two special cases is the fact that f_x^* is computable from f_x and the extra structure imposed. (Notice that once f_x^* is known, the identifiability problem becomes trivial.) The fact that f_x^* is computable from f_x in the first case is intuitively obvious in continuous time: f_x is the sum of a white noise and a bandlimited spectrum, and the decomposition is immediate. In the second case, f_x^* is identifiable because all poles of f_x must be poles of f_x^* and vice versa (the measurement noise is MA), while f_x^* has no zeros (x_t^* is an AR process).

Notice also that if f_y^* is in some way computable from f_y, the problem is easy to solve.

For $a^{-1}.b = f_y^*.f_{xy}^{-1}$ and the remaining quantities are also straightforward to obtain. These results remain valid, if the causality assumption (VI) is not imposed. Both cases (i) and (ii) show that an a-priori causality requirement on $w_1(z)$ in general restricts the set of possible spectral density matrices

$$\begin{pmatrix} f_x & f_{xy} \\ f_{yx} & f_y \end{pmatrix}$$

as in general $w_1 = f_{yx}.f_{x^*}^{-1}$ (f_{x^*} is unique) will not be causal.

Note that if (except rationality) no extra structure has been imposed, then equation (6) alone does not contain information to disentangle f_{x^*} and g_x exceeding (13), because $f_x - f_{x^*} = g_x$ is rational again.

Now we shall turn to a more subtle condition, in which the key is the unravelling of f_{xy}, this being possible because of an assumed genericity.

Let

$b = b^+.b^-$

where b^- has no zeros outside or on and where b^+ has

[1] As well known, in our case higher order moments (in the non-Gaussian case) may contain additional information concerning the parameters (see e.g. Reiersøl (1950)). However we will not investigate this here.

no zeros inside the unit circle and where $b^-(0) = 1$.

Theorem 2: A class of EV systems satisfying

(i) b^+ and b^{-*} have no common zeros

(ii) a and e have no common zeros

(iii) d and b^{-*} have no common zeros

(iv) d and c have no common zeros

(v) $e(z) \neq 0 \quad |z| \leq 1$

(vi) $\delta d > 0$

is identifiable. (See Appendix for Proof)

Remark 6: Conditions (i) - (iv) and the additional requirement $e(z) \neq 0 \quad |z| = 1$ are essentially requirements that the scheme be in the same way generic[2]. Condition (vi) is a requirement of minimal dynamics, which in some form is necessary, as the completely static scheme is known to be nonidentifiable (from second moments).

Remark 7: If we do not impose the causality requirement (VI), then the following further conditions have to be added in the last theorem to guarantee identifiability:

(vii) a^+ and a^{-*} have no common zeros; thereby a^+ and a^- are defined analogously to b^+ and b^- (and $a=a^+.a^-$).

(viii) d and b^+ have no common zeros.

(ix) a^{-*} and c have no common zeros.

In this case d can be determined from (7) from those poles z_i of f_{xy} with the property that their reflections $\bar{z}_i^{-1}$ are again poles of f_{xy}; the remaining quantities are obtained as in the proof above.

Remark 8: Consider e.g. the case where b^- and b^{-*} may have common roots, but where the other conditions of theorem 2 hold. Assume that there is only one zero $z_1, |z_1| > 1$ of

$$df_{xy}d^*a^* = e.b^{-*}.\sigma_\varepsilon.e^*b^{+*} \tag{20}$$

such that $\bar{z}_1^{-1}$ is a zero of (20) too. If $e_1 = (1 - \frac{z}{z_1})$, b_1^-, b_1^+, σ_ε satisfy (20), then also $e_2 = 1$, $b_2^- = b_1^-.(1 - z.\bar{z}_1)$, $b_2^+ = b_1^+(1 - \frac{z}{z_1})$, σ_ε.

Let

$$f_{x^*,1} = d^{-1}e_1\sigma_\varepsilon e_1^*d^{-1*}$$

$$f_{x^*,2} = d^{-1}e_2\sigma_\varepsilon e_2^*d^{-1*}$$

If the spectral density f_x of (x_t) satisfies $f_x(e^{-i\lambda}) > \max(f_1(e^{-i\lambda}), f_2(e^{-i\lambda}))$ then (if the degrees of the polynomials are not a priori known), we can obtain no further information from (6). An analogous statement holds for (8).

Of course if e.g. $f_{x^*,1}(e^{-i\lambda}) > f_x(e^{-i\lambda})$ for some λ then f_1 cannot correspond to the "true" parameters.

If we have a priori knowledge about the structure of the measurement errors, we do not need conditions (i)-(iv) and (vi) in theorem 2 as the next theorem shows.

Let $\hat{a}$ denote the greatest common divisor of a and e (with the normalization $\hat{a}(0) = 1$) and let $(a,e) = \hat{a}^{-1}.(a,e)$.

Theorem 3: A class of EV systems satisfying:

(i) $c = 1$

(ii) $f^{-1}.g = 1$

(iii) $\delta\tilde{e} > 0$

is identifiable. (See Appendix for Proof)

Last, we shall consider the case again, discussed at the beginning of this section, where $c^{-1}.h = 1$, but no longer do we assume that $\delta d > \delta e$. As earlier, we can at once identify d. As it turns out, we can also identify other quantities in the generic case, where the appropriate notion of genericity is different to that used earlier and will become clear in the description of the identification process. For the sake of simplicity let us assume $e(z) \neq 0 \quad |z| \leq 1$.

From

$$df_{xy}d^* = e.b^{-*}.\sigma_\varepsilon.e^*.b^{+*}.a^{-1*} \tag{21}$$

it is trivial to identify all those z_i which have the potential to be a zero of e - any zero of d $f_{xy}d^* \quad |z| < 1$ can have this property. (This is so irrespective of whether z^{-1} is also a zero; also a zero of $df_{xy}d^*$ in $|z| > 1$ may be a zero of b^*.) There are evidently a finite number of candidates for e(z).

Now we can appeal to genericity in one of two ways: If b is not such that it has two zeros with product 1 and if e and a have no common zero, then the only candidate possible for e is $\Pi(z - z_i)$ where z_i and z_i^{-1} are both zeros of $df_{xy}d^*$. Alternatively, considering (19) we note that generically only one candidate e(z) will be such that constants σ_ε and σ_u exist to satisfy this equation.

Of course, once d and e are known, the rest is easy.

References

Aigner, D.J., Hsiao, C., Kapteyn, A. and Wansbeek, T. (1983): Latent Variable Models in Econometrics. In: Griliches, Z. and Intriligator, M.D. (Eds.) Handbook of Econometrics. North Holland Publishing Co. Amsterdam.

Hannan, E.J. (1970): Multiple Time Series. Wiley, New York.

Hsiao, C. (1983): Identification. In: Griliches, Z. and Intriligator, M.D. (Eds.) Handbook of Econometrics. North Holland Publishing Co. Amsterdam.

Kalman, R.E. (1981): System Identification from Noisy Data. Proc. International Symposium on Dynamical Systems (Gainesville, Fl.)

[2] A property is called generic (in the parameter space $\mathbb{R}^n$) if it holds for all points of the parameter space with the possible exception of a proper variety (i.e. the locus of common zeros of a finite number of polynomials over $\mathbb{R}^n$ which is not the $\mathbb{R}^n$ itself). Especially if $nd > 0$ then the class of parameters satisfying (i)-(v) are open and dense relative to the parameter space and its complement is of Lebesgue measure equal to zero.

Maravall, A. (1979): Identification in Dynamic Shock-Error Models, Springer Verlag, Berlin.

Newbold, P. (1978): Feedback Induced by Measurement Errors. Int. Economic Rev. 19, 787-791.

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.

Söderström, T. (1980): Spectral Decomposition with Application to Identification. In: Archetti, F. and Cugiani, M. (Eds.) Numerical Techniques for Stochastic Systems. North Holland Publishing Co. Amsterdam.

Wegge, L. (1981): ARMAX-Model Parameter Identification without and with Latent Variables. Working Paper. Department of Economics, University of California, Davis.

Appendix

Proof of Theorem 2:

Due to (iii), d can be determined from those poles of f_{xy} which are located outside of the unit circle. Then from (7) we have

(A1) $df_{xy}d^* = e.b^{-*}.\sigma_\varepsilon.e^*.b^{+*}.a^{-1*}$

By (ii) we can determine a from the poles of $df_{xy}d^*$. Then due to (i) and (v) e can be determined from

(A2) $df_{xy}d^*a^* = e.b^{-*}.\sigma_\varepsilon.e^*b^{+*}$

as e corresponds to all zeros of the left hand side of (A2), $z_1,\ldots,z_{\delta e}$ say, which are outside the unit circle and for which $\bar{z}_1^{-1},\ldots,\bar{z}_{\delta e}^{-1}$ are zeros of (A2) too:

$$e(z) = \prod_1^{\delta e}(1-\frac{z}{z_i})$$

But then b^- can be determined from the remaining zeros of (A2) which are located outside the unit circle and analogously b^+ can be determined up to a multiplicative constant. Multiplying (6) by dd^* and evaluating this expression at a zero of d, (by (vi), there is one at least) gives σ_ε and then b^+ is determined from (A2). c,h,σ_u,f,g and σ_v then can be determined from usual factorizations.

Proof of Theorem 3:

Due to (i) d is obtained from (6). Furthermore, we obtain $\tilde{a}$ from the poles of the left hand side in (20) and we have:

(A3) $df_{xy}d^*\tilde{a}^* = \hat{a}.e.b^{-*}.\sigma_\varepsilon\tilde{e}^*b^{+*}$

Now let $z_1,\ldots,z_{2r}$ denote all those roots of b, with the property that their reflections z_i^{-1} are again roots of b, arranged such that $|z_i| \geq z_{i+1}|$, $\bar{z}_i^{-1} = z_{2r-i}$; $i = 1\ldots r$; Furthermore let

$$b_r^+ = \prod_{i=1}^{r}(1-\frac{z}{z_i})$$

We can obtain $\tilde{e}b_r^+$ from all those roots in (A3) which are outside or on the unit circle and where their reflections are roots in (A3) too; also $b_r^+ b_r^{+-1}$ then can be determined from (A3) up to a multiplicative constant.

From (8) we have

$$adf_y d^*\tilde{a}^* = b^+.b^{-*}.\tilde{e}.\sigma_\varepsilon.\tilde{e}^*b^-b^{+*} + \tilde{a}d\sigma_v d^*\tilde{a}^*$$

By (iii) there is a z_i such that $\tilde{e}(z_i) = 0$; then $d(z_i) \neq 0$, and $\tilde{a}(z_i) \neq 0$ and thus there is at least one zero of $\tilde{e}b_r^+$ for which (A4) is unequal to zero and for such a zero we obtain

(A5) $f_y(z_i) = \sigma_v$

Multiplying the expression

(A6) $\tilde{a}d(f_y - \sigma_v)d^*\tilde{a}^* = b^+b^{-*}\tilde{e}\sigma_\varepsilon\tilde{e}^*b^-b^{+*}$

with $(b^+.b_r^{+-1}.\tilde{e}.b_r^+)^{-1}(b^{+*}.\tilde{e}^*)^{-1}$ we obtain b^-; but then e can be determined from

$$df_{xy}d^*\tilde{a}^*.b^{-*-1}(b^{+*}\tilde{e}^*)^{-1}.B^{-1}(o) = \hat{a}\tilde{e}\sigma_\varepsilon.B^{-1}(o)$$

But then also $\tilde{e}$ (from $\tilde{e}$ and eb_r^+) and a are uniquely determined and also b^+ up to a multiplicative constant. This constant as well as σ_ε, is determined by (A3) and (A6), h and σ_u are given by (6).