System-theoretic trends in econometrics
Schumacher, J.M.

Publication date: 1990

Citation for published version (APA):
SYSTEM-THEORETIC TRENDS
IN ECONOMETRICS

J.M. Schumacher

FEW 432
Abstract: This is a brief survey of some recent research trends in econometrics which make extensive use of techniques developed in system theory. In particular, we pay attention to the following subjects: cointegration, error correction, and the representation of systems; path controllability, system inversion, and trackability; inputs, outputs, and errors-in-variables.
Introduction

System theory interacts with the theory of economics and econometrics in rather diverse ways, and the past few decades have seen the arrival and sometimes also the departure of a rich variety of research trends in the interface. The story might begin with *The Mechanism of Economic Systems* [55], a book that was published in 1953 although it was based on notes that the author, Arnold Tustin, had written immediately after World War II. In this book, Tustin proposed to model the workings of a national economy by analog simulation using clever mechanical and electrical devices which he described in some detail. Apparently his hope, as an electrical engineer, was to use such nonlinear models to explain and remedy business cycles much in the same way as unwanted oscillatory motions in servomechanisms can be suppressed by appropriate controller design. As noted by Aoki [3], this approach doesn't seem to have had widespread influence among economists.

There have been other trends, however, which did acquire a status of permanence in the economic and econometric literature. Optimal control theory, in the style that emerged in the fifties, has found its way into the economic realm and is well and alive there. This is evidenced in recent textbooks such as [16] and [53]. Optimal stochastic control theory has found application in financial management; a recent survey is provided in [31]. There are other areas that are more or less allied to system theory and that are extensively used in economics, such as the theory of differential games, but we will leave these out of our discussion.

An example of a standard and full-fledged subject in system theory that has had an undeniable influence in econometrics is, of course, the Kalman filter. Its importance was recognized in the standard reference [22], and the Kalman filter can now be considered as one of the standard tools in the study of time series and dynamic economic models (cf. [14, 48]). Further interaction between system theory and econometrics takes place in the field of identification. The fundamental problems that are involved here were stirred up by R.E. Kalman [30]. A recent detailed elaboration of some of the points raised by Kalman can be found in [36, 37]. At a more technical level, the recent book...
by Hannan and Deistler [23] provides an excellent reference for the way that system theory and statistics interact to solve identification problems.

In this paper, we shall attempt to highlight some of the newer research trends in econometrics which make extensive use of ideas and techniques from system theory. First, we shall discuss the issue of 'cointegration' which has been heavily debated in econometric circles during the past decade. One of the central points in the discussion is a result known as the Granger representation theorem; this is basically a theorem about alternative representations for linear dynamic systems, which in system-theoretic terms would fall under the heading of realization theory (or as some would perhaps prefer to say: the theory of system representations and transformations). There is also an aspect of control in the cointegration debate; in particular, the tracking of targets is involved. The ability of a system to track a given target is a classical subject in system theory, and recently there have been some efforts to extend this older work and to apply it in specific economic contexts. We shall briefly discuss the results in this area in Section 3. Our final topic will concern the selection of 'inputs' and 'outputs' ('endogenous' and 'exogenous' variables, in econometric terminology). This subject allows a four-fold decomposition brought about by the two divisions static/dynamic and deterministic/stochastic; we shall discuss all four cases, to bring out some interesting analogies. The final Section 5 contains concluding remarks.

In this paper we will not cover all of the impulses to the application of system-theoretic ideas in economics that are due Aoki and his co-workers, such as the ideas concerning aggregation and reduction by balancing; instead we refer to Aoki’s recent book [4]. For additional material, we also refer to the special issue of the Journal of Economic Dynamics and Control on Economic Time Series with Random Walk and Other Nonstationary Components (Vol. 12-2/3 (1988), edited by M. Aoki), the special issues of Computers & Mathematics with Applications on System-Theoretic Methods in Economic Modeling (Vols. 17-8/9 (1989) and 18-6/7 (1989), edited by S. Mittnik), and the survey paper by E. J. Moore [38].
2. Cointegration, error correction, and the representation of systems

Many economic time-series show an apparent random drift, which may be explained by a lack of forces which tend to drive the variable under study to some preferred level. Since the traditional econometric methods of dealing with time-series are based on stationarity assumptions, it is standard practice (recommended for instance in [5]) to pre-filter the data by taking differences. Differencing once will reduce a 'random walk'-like behavior to stationarity. If necessary, a time-series may be differenced several times in order to achieve stationarity. A scalar time series is said to be integrated of order \(d\) if it reaches stationary after differencing \(d\) times. Since there is loss of information involved in taking differences (a differenced model can only describe relations between changes of variables, not relations between the absolute levels), over-differencing should be avoided.

In the context of vector time series, clearly there may be different orders of observation between the components of the vector; more generally, it can happen that certain linear combinations of the components have lower order of integration than the components themselves. This may be seen as strong evidence for the presence of economic forces which tend to keep a certain balance between the components, and the discovery of such relations is therefore of considerable interest. Examples are the relations between consumption and income and between short-term and long-term interest rates [9,13]. Generally speaking, cointegration is found in so-called error correction models. Suppose that we have two (vector) variables \(y_t\) and \(z_t\) which tend to satisfy a static 'target' relation

\[
Ay_t + Bz_t = 0.
\]

The presence of this target relation can be reconciled with the presence of (first-order) nonstationary dynamics by specifying a model in the form

\[
A_1(L)\Delta y_t + B_1(L)\Delta z_t + D(L) [Ay_{t-1} + Bz_{t-1}] = C(L)e_t.
\]

(The notation here is the econometric one: \(L\) is the lag operator that maps \((x_t)\) to \((x_{t-1})\); \(\Delta = I - L\) is the difference operator, which maps \((x_t)\) to \((x_t - x_{t-1})\); \(A_1(z), B_1(z), D(z),\) and \(C(z)\) are
polynomial matrices; \((\varepsilon_t)_t\) is white noise.) This way of incorporating long-term dynamics into short-term dynamic models originates in [9, 47].

A precise formulation of the connection between cointegrated models and error correction models has been proposed by C.W.J. Granger in an unpublished manuscript [20] and in the paper [13]. Specifically, Granger calls a process \((x_t)_t\) cointegrated of order \(d, b\) if all components are integrated of order \(d\), and if some nontrivial linear combination \(z_t = \alpha' x_t\) is integrated of order \(d-b\) where \(b > 0\). A process \(x_t\) in \(\mathbb{R}^n\) that is cointegrated of order \(1, 1\) is said to have cointegrating rank \(r\) if \(\alpha' x_t\) is stationary for some \(rxn\)-matrix \(\alpha'\) of full row rank, and if \(\beta' x_t\) is nonstationary for any matrix \(\beta'\) whose rank exceeds \(r\). The Granger representation theorem gives the connection between representations of 'autoregressive' and 'moving-average' type for time series that are cointegrated of order \(1, 1\). The following version uses a formulation proposed by Johansen [26].

The Granger Representation Theorem. Assume that the \(\mathbb{R}^n\)-valued process \((x_t)_t\) satisfies

\[ \Delta x_t = C(L)\varepsilon_t \tag{1} \]

where \((\varepsilon_t)_t\) is zero-mean white noise of unit variance, and \(C(z)\) is an \(nxn\) matrix-valued function that is holomorphic on the disk \(|z| < 1+\rho\) and that is nonsingular on the same disk except at 1, where \(C(1)\) has rank \(n-r\). Let \(\alpha\) and \(\beta\) be \(nxr\) matrices of full column rank such that \(\alpha' C(1) = 0\) and \(C(1)\beta = 0\). If the \(rxr\)-matrix \(\alpha'(dC/dz(1))\beta\) is nonsingular, then the process \((x_t)_t\) is cointegrated of order \(1, 1\) with cointegrating rank \(r\) and satisfies the equation

\[ \Pi_0 x_t + \Pi_1(L) \Delta x_t = \varepsilon_t \tag{2} \]

where

\[ \Pi_0 = \beta (\alpha'(dC/dz(1))\beta)^{-1} \alpha' . \]

The processes \((\Delta x_t)_t\) and \((\alpha' x_t)_t\) are stationary so that the representation (2) may be seen as an error correction representation.

Conversely, suppose the process \((x_t)_t\) satisfies an equation (2) where \((\varepsilon_t)_t\) is white noise and where the matrix function \(\Pi(z) = \Pi_0(z) + (1-z)\Pi_1(z)\) is holomorphic and nonsingular on the disk \(|z| < 1+\rho\) except at \(z=1\) where \(\Pi_0 = \Pi(1)\) has rank \(r\). Write
\[ \Pi_0 = \beta \alpha' \] where \( \alpha \) and \( \beta \) are \( nxr \)-matrices of full column rank. If the \( rxr \)-matrix \( \alpha' \Pi_1(1) \beta \) is invertible, then the process \( (x_i)_t \) is cointegrated of order 1, 1 with cointegrating rank \( r \) and satisfies an equation (1) in which \( C(z) \) is holomorphic and nonsingular in the disk \( |z| < 1 + p \) except at the point \( z=1 \), where

\[ C(1) = \beta (\alpha' \Pi_1(1) \beta)^{-1} \alpha'. \]

The proof of the Granger representation theorem in [13] is somewhat hard to follow. Engle sketches a different proof, due to B.S. Yoo, in [12]. This proof is based on what Engle calls the Smith-McMillan-Yoo form; it is actually a Smith form with respect to the ring of causal stable rational functions. In [26], Johansen uses the context of functions that are holomorphic on an open disk containing the unit circle (which is more general than the rational context used by Yoo), and he provides a third proof.

Apparently it hasn't been noticed in this literature that essentially a matrix generalization is involved here of the following simple rule from complex function theory: if \( f(z) \) is holomorphic in a neighborhood of \( z_0 \), then \( f^{-1}(z) \) has a simple pole at \( z_0 \) if and only if \( df/dz(z_0) \) is nonzero, and in that case the residue of \( f^{-1}(z) \) at \( z_0 \) (i.e. the coefficient of \( (z-z_0)^{-1} \) in the Laurent series development of \( f^{-1}(z) \) around \( z_0 \)) is given by \( (df/dz(z_0))^{-1} \). In the matrix case, one has to take directions into account, and the resulting residue formula is given below.

We shall say that a matrix function \( G(z) \) has a simple pole at a point \( z_0 \) of the complex plane if \( G(z) \) has a pole at \( z_0 \) but \( (z-z_0)G(z_0) \) doesn't have a pole there.

**Residue Formula.** Let \( F(z) \) be an \( nxn \) matrix function that is holomorphic in a neighborhood of \( z_0 \), and suppose that \( F(z) \) is nonsingular in a neighborhood of \( z_0 \) except at \( z_0 \) itself. Let the rank of \( F(z_0) \) be \( n-r \); let \( \alpha \) and \( \beta \) be \( nxr \)-matrices of full column rank such that \( \alpha' F(z_0) = 0 \) and \( F(z_0) \beta = 0 \). Under these conditions, the matrix function \( F^{-1}(z) \) has a simple pole at \( z_0 \) if and only if the constant matrix \( \alpha' (dF/dz(z_0)) \beta \) is invertible, and in that case the residue of \( F^{-1}(z) \) at \( z_0 \) is given by

\[ \text{Res} \left( F^{-1}(z); z_0 \right) = \beta (\alpha' (dF/dz(z_0)) \beta)^{-1} \alpha'. \]

The formula is given by Lancaster for the case in which \( F(z) \) is a polynomial matrix [35, pp. 60-65]; the holomorphic version is
formula (4.18) in [50]. The proof is based on a suitable ('local') version of the Smith form. To see how the residue formula applies to the Granger representation theorem, we note that \( \Pi(L) \) and \( C(L) \) should be related by

\[
\Pi(L)C(L) = \Lambda.
\]

This means that

\[
C^{-1}(z) = \Pi_0 (1-z)^{-1} + \Pi_1 (z),
\]

so that \( \Pi_0 \) is the residue of \( C^{-1}(z) \) at 1, and of course we also have

\[
\Pi^{-1}(z) = C(1)(1-z)^{-1} + (C(z)-C(1))/(1-z)
\]

so that \( C(1) \) is the residue of \( \Pi^{-1}(z) \) at 1.

Aside from the technicalities, a more fundamental point that might be brought up in connection with the Granger representation theorem is the following. The theorem purports to be a statement about different representations of the same thing, but it is actually not too clear what it is that is being represented. Statements about equivalence of representations are traditionally formulated in situations in which there is a unique stationary solution associated with each representation, and in this case there is of course no problem—what is represented is that stationary solution. If one leaves the domain of stationary series, however (as one is forced to do in order to discuss phenomena such as cointegration), then this obvious answer is no longer applicable. The difficulty is noted by Davidson, who writes: "In fact, because of missing constants of integration a process such as [given by a vector autoregressive equation \( \Pi(L)x_t = \varepsilon_t \), with \( \Pi(1) \) singular] cannot give a complete description of the generation process of the variables; it must be understood as representing a stationary process in the differences" [8, p.8/9]. A more satisfactory approach, however, should address the problem of nonunique solutions directly. The idea of considering sets of solutions rather than individual solutions is a key point in the work of J.C.Willems [57, 58], which already has given rise to an extensive theory of equivalent representations for linear deterministic systems (cf. the survey [51]). It would seem that a similar theory will have to be developed for the stochastic case in order to allow for an exact and complete
formulation of results such as the Granger representation theorem.

Now, let us consider briefly the general situation of higher-order cointegration. If \( (x_t) \) is a process that is integrated of order \( d > 0 \), then the Wold decomposition implies a representation of the form

\[
\Delta^d x_t = C(L) e_t.
\]

We shall continue to assume that the matrix function \( C(z) \) is holomorphic on an open disk containing the unit circle and that \( C(z) \) is nonsingular on the same disk except at \( z = 1 \). It is natural to define the cointegration space of order \( k \) as the set of all vectors \( \alpha \) such that \( \Delta^k \alpha' x_t \) is stationary. If we denote the dimension of this space by \( n_k \), then we may call the indices \( (n_0, n_1, \ldots, n_d) \) the cointegration indices of the process \( (x_t) \). In this terminology, an \( \mathbb{R}^n \)-valued process is integrated of order 1, 1 with cointegrating rank \( r \) if and only if its cointegration indices are \( (r, n) \). The cointegration indices can be easily expressed in terms of the coefficients of the power series development of \( C(z) \) around \( z = 1 \): writing

\[
C(z) = \sum_{j=0}^{\infty} C_j (1-z)^j,
\]

we have

\[
n_{d-1} = \dim \ker \left[ C_0 \quad C_1 \quad \cdots \quad C_{d-1} \right].
\]

The important point to note is that the cointegration indices are not in any one-one relation with the orders of the zeros at 1 of the matrix function \( C(z) \). (We recall that a nonsingular meromorphic matrix function \( F(z) \) allows, with respect to a given \( z_0 \in \mathbb{C} \), a ‘local’ Smith form

\[
F(z) = U(z) \text{diag} \left( (z-z_0)^{k_1}, \cdots, (z-z_0)^{k_n} \right) V(z) \tag{3}
\]

where \( U(z) \) and \( V(z) \) are holomorphic in a neighborhood of \( z_0 \) and invertible at \( z_0 \). The integers \( k_1, \cdots, k_n \) are called the orders of the zeros of \( F(z) \) at the point \( z_0 \).) This is seen most clearly by comparing the formula

\[
n_{d-1} = \dim \{ \alpha | (1-z)^{-1} C'(z) \alpha \in H(1) \},
\]

in which we use the notation \( H(1) \) for the space of vector
functions that are holomorphic in a neighborhood of 1, with the following formula (adapted from [41]) for the number $v_j$ of zeros at 1 of $C(z)$ of order $\geq j$:

$$v_j = \dim\{\alpha(1) \mid \alpha(z) \in H(1), (1-z)^{-j}C'(z)\alpha(z) \in H(1)\}. \quad (4)$$

Clearly we have

$$n_{d-j} \leq v_j \quad (5)$$

but equality does not hold in general, as can be seen from simple examples. The most important exception to this is, of course, the case of first-order integration.

It can easily be seen that the vector functions $\alpha(z)$ which appear in (4) may be restricted to be vector polynomials, without impairing the validity of the statement. Therefore, if we allow cointegrating vectors to be polynomial rather than constant and change the definition of 'cointegration indices' accordingly, we do obtain a one-one relation between cointegration indices and orders of zeros at 1. The importance of polynomial cointegrating vectors (PCIV's) has been emphasized by Yoo (cf. [12]). A slightly different approach is taken by Johansen [24]. He introduces what we have called the 'cointegration indices', and notes that their sum can at most be equal to the order $r$ of the zero of $\det C(z)$ at 1. The case in which equality holds is referred to by Johansen as the 'balanced' case; since it is easily verified that

$$\sum_{j=1}^{\infty} v_j = r,$$

we can see that this case is the one in which equality holds in (5) for each $j=1, \cdots, d$ and, moreover, $v_j=0$ for $j>d$. Johansen proceeds to show that, after constant row transformations which are summarized in a nonsingular matrix $T$, we can write

$$TC(z) = \begin{pmatrix}
\tilde{C}_0(z) \\
(1-z)\tilde{C}_1(z) \\
\vdots \\
(1-z)^k\tilde{C}_k(z)
\end{pmatrix}$$

where, in the balanced case, the matrix $\tilde{C}(z)=[\tilde{C}_0(z) \cdots \tilde{C}_k(z)]'$ is nonsingular at 1. We may also write this in a slightly different way:
Comparing this with (3), we see that the balanced case is characterized by the fact that the local Smith form around $z=1$ can be obtained using only a constant transforming matrix on the left side. In general, one will have to use a non-constant transformation; although the local Smith form in principle calls for holomorphic transformations, Johansen proves by a direct argument that a polynomial transformation on the left hand side will suffice. (In the rational case, one might appeal to the Smith-McMillan form to prove this; in fact, this is what Yoo does.) The polynomial transformation can then be interpreted as a transformation of the variables in which linear combinations are taken of contemporaneous and lagged components.

So, either by introduction of polynomial cointegrating vectors or by polynomial transformations of the variables, the structure of cointegrated systems can be studied through the zero structure of an associated matrix function at $z=1$. This may help to solve remaining problems, such as the formulation of analogs of the Granger representation theorem for higher-order cointegrated series (partial results on this can be found in [24] and [8]). Another important question is, to what extent polynomial cointegrating vectors (or polynomial transformations of the variables) are unique; the answer to this is of course critical to the discovery of 'target relations'.

In the above, we have emphasized what might be called the 'structural' aspect of cointegration. There is of course also a 'statistical' side to the matter, which is concerned with the testing of hypotheses about the cointegration structure and with the estimation of cointegrating vectors, and most of the journal literature in fact concentrates on this aspect (see for instance [13, 25, 42]). Virtually all of this work is concerned with first-order cointegrated systems. It seems, however, that even in this context there are some basic questions that remain to be answered, in particular in connection with hypothesis testing. Engle notes: "The null hypothesis of cointegration would be far more useful in empirical research than the natural null of non-cointegration. The selection of a 5% test of the non-cointegration null is very arbitrary and many researchers are assuming cointegration when these tests are only rejected at larger significance levels" [12, p.26/27]. One may argue about
what is natural; in a sense, the hypothesis of cointegration is the more highly structured one, and is therefore simpler and more natural. From a certain point of view, the cointegrated situation is also the more singular one, which may explain the difficulties that classical statistical methods have with adopting cointegration as the null hypothesis. Possibly the theory of zeros of matrix functions may also be of help here to unravel the singularities.

3. The tracking of targets

Although cointegration can be caused by the presence of 'common trends', another explanation that is sometimes plausible is presence of steering action. Davidson and Hendry [10] even use the word 'servo-mechanism' to describe the economic forces that keep certain variables together; Arnold Tustin would have appreciated this terminology. Error correction models are placed explicitly in a context of target following by Kloek [32]. It should then be expected that the extensive theory of tracking which has been developed in mathematical control theory should have some relevance.

There is a sizable economic literature with a clear system-theoretic motivation on the problem of exactly following a prescribed path, the so-called 'path controllability'. The problem is customarily posed in a deterministic setting and bears a mathematical-economic flavor rather than an econometric one. Path controllability can be seen as an extension of Tinbergen's concept of achievability of targets in static models [54]. When the targets are solved in terms of the instruments in a static linear model, so that we have

$$y = Gu$$

where $y$ is a vector of targets, $G$ is a constant matrix, and $u$ is a vector of instruments, then the obvious criterion for achievability of each given vector $y$ by a suitable choice of instruments $u$ is that the matrix $G$ should have full row rank. A necessary condition for this to hold is of course that the number of targets should not exceed the number of instruments; this is sometimes called the "Tinbergen policy condition". The dynamic version of target achievability was introduced in economics by Preston [44] and Aoki [2], after essentially the same idea had been
introduced into system theory (under the name of ‘functional reproducibility’) by Brockett and Mesarović [6]. In the discrete-time case, path controllability is defined to mean that, after a certain ‘adjustment time’ or ‘policy lead’, any given path of the target variables can be tracked exactly by proper choice of the instrument variables. The definition in the continuous-time case is slightly different, but the criterion (at least in the linear constant-parameter case) is the same: path controllability holds if and only if the transfer matrix $G(z)$ from instruments to targets has full row rank as a rational matrix [6, p. 559]. This is a rather attractive generalization of the static rule of Tinbergen.

Further work within the system theory community on this subject has concentrated on finding simple conditions for right invertibility in terms of the state space representation

$$x(k+1) = Ax(k) + Bu(k), \quad x(k) \in X, \quad u(k) \in U$$

$$y(k) = Cx(k) + Du(k), \quad y(k) \in Y.$$

A condition for right invertibility in terms of the parameters $A$, $B$, $C$, and $D$ was already given by Brockett and Mesarović, but this involved a rather big matrix formed from the parameter matrices. The following compact method for determining whether or not a system is right invertible is essentially due to Morse and Wonham [39]. Define recursively a sequence of subspaces of the state space $X$ by

$$T^0 = \{0\}$$

$$T^{k+1} = \{x \in X \mid x = Ax + Bu \text{ for some } x \in T^k \text{ and } u \text{ such that } Cx + Du = 0\}.$$

It is easily seen that the sequence $(T^k)_k$ is nondecreasing, and so the sequence must have a limit which is denoted by $T^*$. The system given by the parameters $(A, B, C, D)$ is right invertible (in the sense that the transfer matrix $G(z) = C(zI - A)^{-1}B + D$ is right invertible as a rational matrix) if and only if

$$CT^* + \text{im} D = Y.$$

The state-space framework suggests extensions to the non-constant-parameter case and the nonlinear case. A characterization of path controllability for linear systems with time-varying parameters has been given by Engwerda [15]; necessarily the
condition is more involved than in the constant-parameter case, but an analogy with the Morse-Wonham result can still be drawn. Necessary and sufficient conditions for (local) path controllability for discrete-time nonlinear systems have been given by Nijmeijer [40], who also establishes the close relation that exists between path controllability and decouplability (the possibility of introducing a control policy in which each target is influenced by only one instrument). Recently, state space algorithms have become available to decide on the right invertibility of systems that are given in implicit form, rather than in solved form [33]. This is a return to the original formulation by Tinbergen, who starts in [54] with implicit equations rather than with a 'final form'.

One may reasonably argue that the invertibility of dynamic systems should play an important role in dynamic economic theory, simply because invertibility is such a basic concept; so the study of system invertibility is well-motivated. However, it is also clear that exact path controllability is not a realistic goal in a world full of disturbances. Alternative formulations of tracking problems can be obtained by introducing assumptions on the variables to be tracked, and a stochastic setting can be accommodated by relaxing the condition of exact tracking. For instance, Kloek in [32] assumes that the target is integrated of order 2 and requires that the tracking error should be zero-mean and weakly stationary. Situations in which some information is available about the dynamics of the signal to be followed have been studied extensively in system and control theory; in fact, this branch of control theory has its roots in the design of certain servomechanisms that were used in World War II and that took the notion of 'target' quite literally. We refer to [34, Ch.5], [60, Ch.6-8], [49], and [18] for a sample of the modern literature on the subject. Basically, the conclusion of these studies is that, for trackability of constants and linear trends in discrete-time systems, the transfer matrix from instruments to targets should have full row rank at the point 1 of the complex plane. Moreover, if the action of the controlling mechanism is based purely on the tracking error, then it can be shown that the controller must contain what is called an 'internal model' of the signal that is to be followed.

A few remarks can be made here. Firstly, we see that again
the zero structure at 1 is of importance. Secondly, a somewhat surprising conclusion is that the trackability condition is stronger than the condition for path controllability; indeed, if the transfer matrix has full row rank at some given point of the complex plane then it will certainly have full row rank as a rational matrix. This may be explained by the fact that path controllability is achieved by open-loop control, whereas in the case of the trackability problem the solution is sought in the form of a closed-loop controller, which automatically adjusts the control action to changes in the signal to be followed. Thirdly, the presence of an 'internal model' might be an interesting hypothesis in situations in which control action is suspected, such as when time series are cointegrated. Structural constraints such as the one implied by the internal model principle may also be used in model specification. We note that the internal model principle has been mentioned recently by Salmon [46], who however seems to use the term to indicate compatibility between models; this is certainly a subject of interest, but not related directly to the tracking problem.

The role of ideas from control theory in mathematical economics can now almost be called classical; this is true for optimal control, but also for a number of other ideas in which optimization is not necessarily involved, such as path controllability. Developments that may be expected here include further elaboration of the relation between path controllability and decoupling, and study of the structure of control policies when the instruments are in the hands of various different agents. The application of control ideas in econometric modeling is more recent and, to a considerable extent, this subject still has to take shape. In many situations in which several variables of interest are studied there is a great need for structural information to be incorporated in the specification of models, and the results of control theory may help to provide such information in the form of constraints that must be satisfied for control action to be effective.
4. Inputs, outputs, and errors-in-variables

It is a generally recognized fact among econometricians that the distinction between endogenous and exogenous variables is often debatable. For this reason (and for other reasons as well) it has been argued by J.C. Willems [56] that in a general theory of systems one should start with a notion of 'observables' or 'external variables' without imposing a priori a division between inputs and outputs. This implies that one should describe the relations between the variables in a nondiscriminating way. Having done this, one may ask which choices of inputs and outputs would be reasonable; of course, exactly what is 'reasonable' in a given situation may depend on the availability of extra information which is not expressed in the system description. We shall discuss the problem of selecting inputs and outputs in four cases, corresponding to the divisions static/dynamic and deterministic/stochastic. The discussion will be limited to linear systems, however.

The deterministic static case would perhaps be considered trivial, but let us discuss it anyway for purposes of comparison. Suppose a linear relation between external variables $w_i$ is given by

$$Rw = 0$$

(6)

where we may assume that the matrix $R$ has full row rank. If we believe that it is reasonable to require that the inputs are not restricted by the equations and that the outputs are completely determined by the inputs and by the equations, then the standard procedure applies: select output variables by finding a maximal set of independent columns among the columns of $R$, name the associated components $y$, name the remaining components $u$, rewrite (6) as $R_1y + R_2u = 0$ and, noting that $R_1$ is invertible by construction, obtain

$$y = -R_1^{-1}R_2u$$

which clearly has the desired characteristics. In general, the choice of inputs is not unique; however, the number of inputs is determined by the data (6). Any selection of this number of variables will 'generically' be valid as a choice of inputs.

There is a certain asymmetry in the selection procedure based on (6) since we first select the outputs and then simply let the
inputs be what is left. However, if we would have represented the subspace \( \ker R \) which effectively appears in (6) as the image rather than as the kernel of some matrix, then we would have selected the inputs first by taking a maximal set of independent rows of the representing matrix. So the seeming priority of outputs over inputs in the selection procedure above is just a consequence of the chosen representation.

In the linear deterministic dynamic case, the problem of selecting inputs and outputs has been considered by Willems in [57]. In this case, the condition for an admissible selection of inputs and outputs might be that the transfer matrix from inputs to outputs should exist and should be proper rational. (This can be formulated more intrinsically: see [59]). The solution given in [57, 58] may be described as follows. Let a set of difference equations with constant coefficients in the variable \( w(k) \in \mathbb{R}^q \) be given by

\[
R(\sigma)w = 0,
\]

where \( \sigma \) denotes the (forward) shift and \( R(z) \) is a polynomial matrix which we may assume to have full row rank. The basic technique is to write \( R(z) \) in the form \( T(z)B(z) \) where \( T(z) \) is an invertible rational matrix and \( B(z) \) is 'right bicausal', i.e. \( B(z) \) is proper rational and has full row rank at infinity. This factorization may be achieved by the reduction of \( R(z) \) to row reduced form [27, p.386]; indeed, note that this procedure factorizes \( R(z) \) as \( U(z)\Delta(z)B(z) \) where \( U(z) \) is unimodular, \( \Delta(z) \) is diagonal with diagonal elements of the form \( z^k \), and \( B(z) \) is right bicausal. A proposed selection of inputs and outputs will induce a partitioning of \( R(z) \) as \([R_1(z) \quad R_2(z)]\) (after possible reordering of the columns), and a corresponding partitioning of \( B(z) \) as \([B_1(z) \quad B_2(z)]\). Now, \( R_1(z) \) will be invertible if and only if \( B_1(z) \) is invertible, and \( R_1^{-1}(z)R_2(z) = B_1^{-1}(z)B_2(z) \) will be proper rational if and only if \( B_1(z) \) doesn't have a zero at infinity. (The 'only if' holds because \( B_1(z) \) and \( B_2(z) \) are coprime as matrices over the ring of proper rational functions, so there can't be a pole-zero cancellation.) So the result is that the proposed selection of inputs and outputs is admissible if and only if the matrix \( B_1(\infty) \) is nonsingular. In other words, what we have to do is to select a maximal number of independent columns from the full row rank matrix \( B(\infty) \) — we might say that the
problem is reduced to the static case.

Of course, this solution is hardly surprising to the econometrician, who is used to representing transfer matrices as quotients of matrices of polynomials in $z^{-1}$ (the backward shift). In models of the form

$$B(\sigma^{-1})y = A(\sigma^{-1})u$$

where $A(z)$ and $B(z)$ are polynomial matrices, the condition that $B(0)$ should be invertible is known as the 'causality condition'; in fact, such models are often specified with the condition $B(0)=I$ (see for instance [22, p.13]).

In order to make a comparison with the stochastic situation that will be discussed below, let us see how much more difficult the problem becomes when we require that that the transfer matrix from inputs to outputs should not only be proper, but also stable. In principle, the same technique as above applies: if we can write $R(z)$ in the form $T(z)B(z)$ where $T(z)$ is an invertible rational matrix and $B(z)$ is now a proper stable rational matrix having full row rank for all $z$ with $|z|\geq 1$, then a selection of inputs and outputs will be admissible if and only if the corresponding matrix $B_1(z)$ is nonsingular for all $z$ with $|z|\geq 1$. The desired factorization of $R(z)$ can be obtained by a Wiener-Hopf factorization with respect to the unit circle [7] (cf. the interpretation of the reduction to row reduced form as a Wiener-Hopf factorization with respect to the point at infinity in [19]). So in this case, the input-output selection problem is essentially the following: given a matrix that is 'right unimodular' over the ring of proper stable functions, find a square submatrix that is unimodular. Obviously, this is not always possible. The simplest example would be that of a system with two variables in which neither the transfer matrix from the first to the second variable nor its inverse is stable.

Next, let us consider the stochastic case. If we suppose that both the observations and the additive noise are generated by mechanisms that can be modeled as zero-mean normally distributed variables, then the general linear model can be written as

$$w = Nx + \epsilon$$

where $x$ generates the observations and $\epsilon$ is noise. The observed vector $w$ will be normally distributed with zero mean and
covariance matrix $Q$, and so all observational data are summarized in $Q$. In the model (7), we could select independent rows from the matrix $N$ (which may be assumed to be of full column rank) and we might convert the model to an input-output form just as in the deterministic case. However, without further assumptions on the noise, the model (7) is hopelessly non-unique. Not even the number of inputs is well-defined; it may vary from $rkQ$ (no noise) to 0 (all noise).

One possible constraint on the noise covariance matrix $\Sigma$, which is well-motivated when the observation space $\mathbb{R}^q$ is considered as the Cartesian product of $q$ different one-dimensional spaces, is to require that $\Sigma$ should be diagonal. This, of course, leads to the factor analysis model, which has experienced renewed interest following Kalman's critique of the concept of identifiability in econometrics [28, 29]. What we called 'the number of inputs' becomes 'the number of common factors' in the context of factor analysis, and it is natural to define this number as the minimal length of the vector $x$ for which a representation of the form (7) (with $cov(\varepsilon \varepsilon')$ diagonal) is possible. In contrast to the unconstrained case, this number is now well-determined, but unfortunately its determination is an open problem.

From the point of view of selecting inputs and outputs, it may be more natural to think of $\mathbb{R}^q$ not as the product of $q$ one-dimensional spaces, but as the product of an input space and an output space (yet to be determined). A possible constraint to impose would be that the noise covariance matrix should be block diagonal corresponding to this decomposition. This leads to an alternative interpretation of the vector $x$, since it can be shown that the model

$$
\begin{bmatrix}
y \\
u
\end{bmatrix} = 
\begin{bmatrix}
H_1 \\
H_2
\end{bmatrix} x + 
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2
\end{bmatrix}
$$

(with $x$, $\varepsilon_1$, and $\varepsilon_2$ independent) holds if and only if $y$ and $u$ are conditionally independent given $x$. The conditional independence property is also used to define the notion of 'state' in stochastic systems (see for instance [52]), and so the problem of constructing a model of the form (8) for a given decomposition of $w$ into components $u$ and $y$ is sometimes called a realization problem [17, 45].

Let us say that a decomposition of $w$ into inputs $u$ and
outputs $y$ is 'admissible' if there is a model of the form (8) in which $x$ has minimal length among all models of the same type corresponding to the same decomposition, and in which the matrix $H_2$ is invertible. The invertibility of $H_2$ will allow the model to be rewritten in an input-output form:

$$\begin{align*}
\hat{y} &= H_1H_2^{-1}u \\
y &= \hat{y} + \epsilon_1 \\
u &= \hat{u} + \epsilon_2.
\end{align*}$$

This is the errors-in-variables form (see for instance [11]). The decomposition of $w$ into $y$ and $u$ leads to a partitioning of the covariance matrix $Q_{ww}$:

$$Q_{ww} = \begin{bmatrix} Q_{yy} & Q_{yu} \\ Q_{uy} & Q_{uu} \end{bmatrix}.$$ 

We claim: the decomposition of $w$ into inputs $u$ and outputs $y$ is admissible if and only if the matrix $Q_{yu}$ has full column rank. To see this, assume first that we do have an admissible decomposition and let (8) be a corresponding model. Because $x$ has minimum length, the covariance matrix $Q_{xx}$ of $x$ must be nonsingular. We have $H_1 = Q_{yx}Q_{xx}^{-1}$ since obviously $H_1x$ is the least-squares estimate of $y$ given $x$, and likewise $H_2 = Q_{ux}Q_{xx}^{-1}$. Because of the mutual independence of $x$, $\epsilon_1$, and $\epsilon_2$, one has

$$Q_{yu} = H_1E[x'x]H_2^T = Q_{yx}Q_{xx}^{-1}Q_{xx}.$$ 

Now, it is shown in [17] that the length of $x$ in a minimal representation of the form (8) is equal to the rank of $Q_{yu}$. From the formula above, we see that this implies that $Q_{yu}$ is surjective (and hence invertible) and that $Q_{yx}$ is injective. But then $Q_{yu}$ is injective too, by the same formula. Conversely, if it is given that $Q_{yu}$ has full column rank, then the construction of [17] immediately leads to a representation of the desired form.

The conclusion must be that imposing that the error covariance matrix should be block diagonal doesn’t help very much in the selection of inputs and outputs. In particular, it doesn’t rule out the possibility of attributing all observed variation to noise.

Before turning to the dynamic case, let us note that the errors-in-variables model is not uniquely determined even if we
fix the choice of inputs and outputs. It is easy to see that all possible solutions can be parametrized in terms of the 'true' input covariance matrix $Q$, and that all symmetric positive definite matrices $Q$ will qualify that satisfy the two inequalities

$$Q \leq Q_{uu}$$

and

$$Q_{yu}Q^{-1}Q_{uy} \leq Q_{yy}.$$

Using the singular value decomposition, one can easily show that the latter inequality can be rewritten as a lower bound on $Q$, of the form $Q \geq Q_{\text{min}}$. The corresponding (non-unique) 'true' linear relation between the latent variables $\hat{y}$ and $\hat{u}$ is given by $Q_{yu}Q^{-1}$.

In the static case, several proposals have been formulated to reduce the nonuniqueness of the errors-in-variables model by bringing in some extra information; see for instance [1]. Let us see what the dynamic case has to offer. We follow the development in [21] and [43].

Our goal will be to verify the admissibility of a given decomposition of $w(t)$ in inputs $u(t)$ and outputs $y(t)$. The observational data are supposed to be summarized in a spectral density matrix $Q_{ww}(z)$ for $w$, which is partitioned according to the proposed decomposition as

$$Q_{ww}(z) = \begin{bmatrix} Q_{yy}(z) & Q_{yu}(z) \\ Q_{uy}(z) & Q_{uu}(z) \end{bmatrix}.$$

We are looking for a 'true' transfer matrix $G(z)$ and a 'true' input spectral density $Q(z)$ which should satisfy

$$G(z)Q(z) = Q_{yu}(z),$$

$$Q(z) \leq Q_{uu}(z), \quad |z|=1$$

$$G(z)Q(z)G(z^{-1}) \leq Q_{yy}(z), \quad |z|=1.$$

Under suitable assumptions, the development in the static case can be followed (replace the field $\mathbb{R}$ by the field $\mathbb{R}(z)$, the partial order "\leq" by the partial order "\leq pointwise for $|z|=1"$, and the involution $M \mapsto M'$ by the involution $M(z) \mapsto M'(z^{-1})$). As in the static case, the set of all minimal solutions will be parametrized by the spectral density matrices $Q(z)$ that fall between an upper and a lower bound determined by the data, and the corresponding transfer matrices are then given by
\[ G(z) = Q_{yu}(z)Q^{-1}(z). \]

However, we want to impose both causality and stationarity and so we require \( G(z) \) to have all of its poles inside the unit disk. The problem is to find the restrictions on \( Q(z) \) that will guarantee this property for \( G(z) \).

Again, the key tool to use is the Wiener-Hopf factorization. To avoid some technical intricacies, we shall assume that both \( Q(z) \) and \( Q_{yu}(z) \) have constant rank on the unit circle. Then we can write

\[ Q_{yu}(z) = F_-(z)D(z)F_+^*(z) \]

where

\[ D(z) = \begin{bmatrix} \Delta(z) \\ 0 \end{bmatrix}, \quad \Delta(z) = \text{diag}(z^k, \ldots, z^k) \]

and where \( F_-(z) \) (\( F_+^*(z) \)) is unimodular as a matrix over the ring \( \mathbb{R}_-(z) \) (\( \mathbb{R}_+(z) \)) of rational functions having all their poles inside (outside) the unit circle. (We also used the fact that \( Q_{yu}(z) \) must have full column rank, as in the static case.) For any rational matrix \( M(z) \), write \( M^*(z) = M(z^{-1}) \); note that \( M^* \) will be \( \mathbb{R}_-(z) \)-unimodular if \( M \) is \( \mathbb{R}_+(z) \)-unimodular, and vice versa.

Now, write

\[ G = F_+ DF_+ Q^{-1} = F_+ \bar{Q} (F_+^*)^{-1} \]

where

\[ \bar{Q} = F_+ Q^{-1} F_+^* \]

Do a spectral factorization to write \( \bar{Q} = H H^*_+ \), where \( H_+ \) is \( \mathbb{R}_+(z) \)-unimodular. We then have

\[ G = [F_+]D H_+ [H_+^*(F_+^*)^{-1}] \]

Because the factors between square brackets are \( \mathbb{R}_+(z) \)-unimodular, it follows that \( G(z) \) will be causal and stable if and only if \( DH_+ \) is causal and stable, which is the same as requiring that \( \Delta H_+ \) should be causal and stable. Because \( \Delta(z) \) is diagonal, this requirement entails that all entries \( h_{ij}(z) \) of the \( i \)-th row of \( H_+(z) \), which are rational functions having all their poles outside the unit circle, should be such that the functions \( z^{k_i} h_{ij}(z) \) have all their poles inside the unit circle. Since multiplication
by a power of \( z \) can only move poles between zero and infinity, we see that the \( \kappa_i \)'s should be nonpositive and that the \( h_{ij}(z) \)'s should be polynomials of degree no higher than \(-\kappa_i\). This means that there will be no solution if one of the Wiener-Hopf indices \( \kappa_i \) is positive, and that otherwise the solution set is parametrized by at most \( \kappa \) parameters, where

\[
\kappa = - \sum_{i=1}^{i=m} \kappa_i.
\]

(Of course, we also have the requirement that \( H_4 \) should be unimodular, so the parametrization is nontrivial.)

We see that imposing the requirements of causality and stationarity may well cause a certain proposal for the selection of inputs and outputs to be rejected; if the proposed selection turns out to be admissible, then it causes the set of all possible models to be finitely parametrized. However, there is no indication how to select inputs and outputs in such a way that the associated Wiener-Hopf indices will be nonpositive; this problem was raised in [21] but apparently the question is still open. Also, as in the scalar case, the number of inputs is still undetermined and the possibility of attributing all variance to noise is not ruled out. It may be worthwhile to try out the effect of other possible constraints on the error spectral density, such as size constraints (proposed for the static case in [1]).

5. Conclusions

An econometrician once told me that he was amazed that system theory is still an active field, since he couldn't imagine that the analysis of the Kalman filter would not be completed by now. Apparently, the full variety of system-theoretic methods has as yet failed to disclose itself to the field of econometrics. System theory provides a rich set of examples which illustrate the pitfalls of modeling, and how to avoid these; Kalman has used such examples in his contributions to the ongoing debate on the fundamentals of mathematical modeling and identification. System theory also provides a large body of knowledge about state space techniques, and the applicability of such techniques to econometric problems has been shown in the work of Aoki and others. But the collection of mathematical techniques that are familiar to and developed by system theorists allows an even more
intensive contact. As shown in this paper, matrix factorizations and pole-zero considerations play an important role in econometric problems, and system theorists have applied these for a long time. There is an econometric interest in representation problems, which is something about which system theory has a lot to say. The invertibility of systems is a natural concept in dynamic economic analysis, destined to play a role similar to the invertibility of matrices in static analysis; and again, system theory provides the necessary tools. While some of the questions here are no doubt more modest than the fundamental issues with which R.E. Kalman has confronted the econometric profession, they may still be a worthwhile subject for research and lead to results that will satisfy system theorists and econometricians alike.

Acknowledgement

It is a pleasure to thank Manfred Deistler, Theo Nijman and Henk Nijmeijer for the useful information they kindly provided to me.

References


IN 1989 REEDS VERSCHENEN

368 Ed Nijssen, Will Reijnders
"Macht als strategisch en tactisch marketinginstrument binnen de distributieketen"

369 Raymond Gradus
Optimal dynamic taxation with respect to firms

370 Theo Nijman
The optimal choice of controls and pre-experimental observations

371 Robert P. Gilles, Pieter H.M. Ruys
Relational constraints in coalition formation

372 F.A. van der Duyn Schouten, S.G. Vanneste
Analysis and computation of (n,N)-strategies for maintenance of a two-component system

373 Drs. R. Hamers, Drs. P. Verstappen
Het company ranking model: a means for evaluating the competition

374 Rommert J. Casimir
Infogame Final Report

375 Christian B. Mulder
Efficient and inefficient institutional arrangements between governments and trade unions; an explanation of high unemployment, corporatism and union bashing

376 Marno Verbeek
On the estimation of a fixed effects model with selective non-response

377 J. Engwerda
Admissible target paths in economic models

378 Jack P.C. Kleijnen and Nabil Adams
Pseudorandom number generation on supercomputers

379 J.P.C. Blanc
The power-series algorithm applied to the shortest-queue model

380 Prof. Dr. Robert Bannink
Management's information needs and the definition of costs, with special regard to the cost of interest

381 Bert Bettonvil
Sequential bifurcation: the design of a factor screening method

382 Bert Bettonvil
Sequential bifurcation for observations with random errors
383 Harold Houba and Hans Kremers
Correction of the material balance equation in dynamic input-output models

384 T.M. Doup, A.H. van den Elzen, A.J.J. Talman
Homotopy interpretation of price adjustment processes

385 Drs. R.T. Frambach, Prof. Dr. W.H.J. de Freytas
Technologische ontwikkeling en marketing. Een oriënterende beschouwing

386 A.L.P.M. Hendrikx, R.M.J. Heuts, L.G. Hoving
Comparison of automatic monitoring systems in automatic forecasting

387 Drs. J.G.L.M. Willems
Enkele opmerkingen over het inversificerend gedrag van multinationale ondernemingen

388 Jack P.C. Kleijnen and Ben Annink
Pseudorandom number generators revisited

389 Dr. G.W.J. Hendrikse
Speltheorie en strategisch management

390 Dr. A.W.A. Boot en Dr. M.F.C.M. Wijn
Liquiditeit, insolventie en vermogensstructuur

391 Antoon van den Elzen, Gerard van der Laan
Price adjustment in a two-country model

392 Martin F.C.M. Wijn, Emanuel J. Bijnen
Prediction of failure in industry
An analysis of income statements

393 Dr. S.C.W. Eijffinger and Drs. A.P.D. Gruijters
On the short term objectives of daily intervention by the Deutsche Bundesbank and the Federal Reserve System in the U.S. Dollar - Deutsche Mark exchange market

394 Dr. S.C.W. Eijffinger and Drs. A.P.D. Gruijters
On the effectiveness of daily interventions by the Deutsche Bundesbank and the Federal Reserve System in the U.S. Dollar - Deutsche Mark exchange market

395 A.E.M. Meijer and J.W.A. Vingerhoets
Structural adjustment and diversification in mineral exporting developing countries

396 R. Gradus
About Tobin's marginal and average q
A Note

397 Jacob C. Engwerda
On the existence of a positive definite solution of the matrix equation $X + A'XA = I$
Paul C. van Batenburg and J. Kriens
Bayesian discovery sampling: a simple model of Bayesian inference in auditing

Hans Kremers and Dolf Talman
Solving the nonlinear complementarity problem

Raymond Gradus
Optimal dynamic taxation, savings and investment

W.H. Haemers
Regular two-graphs and extensions of partial geometries

Jack P.C. Kleijn, Ben Annink
Supercomputers, Monte Carlo simulation and regression analysis

Technologie, Strategisch management en marketing

Theo Nijman
A natural approach to optimal forecasting in case of preliminary observations

Harry Barkema
An empirical test of Holmström's principal-agent model that tax and signal hypotheses explicitly into account

Drs. W.J. van Braband
De begrotingsvoorbereiding bij het Rijk

Marco Wilke
Societal bargaining and stability

Willem van Groenendaal and Aart de Zeeuw
Control, coordination and conflict on international commodity markets

Prof. Dr. W. de Freytas, Drs. L. Arts
Tourism to Curacao: a new deal based on visitors' experiences

Drs. C.H. Veld
The use of the implied standard deviation as a predictor of future stock price variability: a review of empirical tests

Drs. J.C. Caanen en Dr. E.N. Kertzman
Inflatie neutrale belastingheffing van ondernemingen

Prof. Dr. B.B. van der Genugten
A weak law of large numbers for m-dependent random variables with unbounded m

R.M.J. Heuts, H.P. Seidel, W.J. Selen
A comparison of two lot sizing-sequencing heuristics for the process industry
414 C.B. Mulder en A.B.T.M. van Schaik
Een nieuwe kijk op structuurwerkloosheid

415 Drs. Ch. Caanen
De hefboomwerking en de vermogens- en voorraadaftrek

416 Guido W. Imbens
Duration models with time-varying coefficients

417 Guido W. Imbens
Efficient estimation of choice-based sample models with the method of moments

418 Harry H. Tigelaar
On monotone linear operators on linear spaces of square matrices
IN 1990 REEDS VERSCHENEN

419 Bertrand Melenberg, Rob Alessie
A method to construct moments in the multi-good life cycle consumption model

420 J. Kriens
On the differentiability of the set of efficient \((\mu, \sigma^2)\) combinations in the Markowitz portfolio selection method

421 Steffen Jørgensen, Peter M. Kort
Optimal dynamic investment policies under concave-convex adjustment costs

422 J.P.C. Blanc
Cyclic polling systems: limited service versus Bernoulli schedules

423 M.H.C. Paardekooper
Parallel normreducing transformations for the algebraic eigenvalue problem

424 Hans Gremmen
On the political (ir)relevance of classical customs union theory

425 Ed Nijssen
Marketingstrategie in Machtspectief

426 Jack P.C. Kleijnen
Regression Metamodels for Simulation with Common Random Numbers: Comparison of Techniques

427 Harry H. Tigelaar
The correlation structure of stationary bilinear processes

428 Drs. C.H. Veld en Drs. A.H.F. Verboven
De waardering van aandelenwarrants en langlopende call-opties

429 Theo van de Klundert en Anton B. van Schaik
Liquidity Constraints and the Keynesian Corridor

430 Gert Nieuwenhuis
Central limit theorems for sequences with \(m(n)\)-dependent main part

431 Hans J. Gremmen
Macro-Economic Implications of Profit Optimizing Investment Behaviour