MODELING, IDENTIFICATION AND CONTROL, 1985, VOL. 6, NO. 1, 3–19
doi:10.4173/mic.1985.1.1

Estimation of large scale implicit models using two-stage methods

ROLF HENRIKSEN†

Keywords: *parameter estimation, large scale systems, two-stage methods, decentralized filtering, prediction error methods, implicit models.*

The problem of estimating large scale implicit (non-recursive) models by two-stage methods is considered. The first stage of the methods is used to construct or estimate an explicit form of the total model, by constructing a minimal stochastic realization of the system. This model is then subsequently used in the second stage to generate instrumental variables for the purpose of estimating each sub-model separately. This latter stage can be carried out by utilizing a generalized least squares method, but most emphasis is put on utilizing decentralized filtering algorithms and a prediction error formulation. A note about the connection between the original TSLS-method (two-stage least squares method) and stochastic realization is also made.

1. Introduction

Identification of large scale models is often considered to be an extremely difficult problem, primarily because of the size and, possibly, the complexity of the models. Quite typically, a large scale econometric model may consist of more than a thousand equations, some of which can be non-linear, and several thousand unknown parameters. It would usually not be very clever to try and estimate all the parameters simultaneously in such a model, not only because of the very size of the model, but also because the structure of some of the sub-models and equations is only partially or even only vaguely known, and because all the restrictions on the model may be hard to impose when all the parameters are estimated simultaneously. In such a case it would be much better, in fact maybe the only meaningful way, to employ some kind of decentralized estimation scheme. Provided sufficient information about the interactions between the sub-systems, and sufficient local observations are available, a completely or partially decentralized estimated algorithm can be employed.

In addition to the size and complexity of large scale models, a large class of discrete-time models consists of so-called implicit models, models where some or all of the equations are implicit. This implies that some of the variables which appear on the right-hand side of the implicit equations may be correlated with the process noise (or equation uncertainties), causing eventual parameter estimates obtained by ordinary least squares methods to be biased. The problem of estimating implicit models has in econometrics (where implicit models are called simultaneous equation models) been solved by employing various multi-stage methods, for example, the

Received 9 January 1985.
† The Norwegian Institute of Technology, Division of Engineering Cybernetics, N-7034 Trondheim-NTH, Norway.

two-stage least squares method (TSLS) which originally is due to Theil (1953) and Basman (1957). The general idea behind these methods is to replace the variables which are correlated with the process noise with new variables that are uncorrelated with the process noise. Assuming certain conditions to be fulfilled, one can then employ a simple least squares method to estimate the unknown parameters one sub-model (or equation) after the other in a completely decentralized manner.

In more general types of models where there are observation uncertainties, one will generally have to use more sophisticated methods in order to get unbiased estimates. We shall in this paper take a closer look at the TSLS method and its relation to stochastic realization theory and instrumental variable methods. The methods derived in this paper consist of two stages, where the first stage is used to generate estimates (of states and interaction variables) which are uncorrelated with the process noise (exactly like the first stage of the TSLS method). This can be done by employing various stochastic realization algorithms or least squares methods. In the second stage the parameters in each sub-model are estimated using a least squares method that, subject to the underlying assumptions yields constant estimates, for example, the generalized least squares method, or by using decentralized filtering algorithms and a prediction error formulation. The state and interaction estimates generated in the first stage are used extensively in the second stage.

The major field of applications of the derived methods seems at first hand to be econometrics, but they can be used to any discrete-time implicit model, models that arise from systems where the sampling interval is long compared to the dynamics of the system. Also, as pointed out by Mehra (1974), the use of aggregation concepts or asymptotic expansion methods in large scale systems can lead to implicit models of the form being assumed in this paper.

The paper is organized as follows. In § 2 a description of the model and the decentralized structure is given. In § 3 we focus our attention on the TSLS method and its relation to stochastic realization theory and instrumental variables methods. § 4 is devoted to generalizing the first stage of the TSLS method, generation of intermediate state and interaction estimates (or instrumental variables) for use in the next stage. In § 5 we estimate the parameters in each sub-model by using variables generated in the previous section. Most attention is focused on the use of decentralized filtering algorithms and a prediction error interpretation. A discussion of the derived methods and conclusions which can be drawn from results on identification in closed loop are made in § 6. In § 7 a conclusion and some preliminary experiences from the derived methods are given.

2. System description

We consider a large scale discrete-time linear state space model of the form

$$x_{t+1} = Ax_t + Bx_{t+1} + Cu_t + v_t \tag{1}$$

where x_t is the state vector, u_t the control (or input) vector, $\{v_t\}$ a gaussian white zero-mean process, whereas A, B and C are matrices of appropriate dimensions. It is assumed that the matrix $(I - B)$ is non-singular, so that (1) has a unique solution.

Models of the form given by (1) occur frequently in econometrics where they are referred to as simultaneous equation models, as opposed to so-called reduced form models where $B = 0$. The real reason for this special model form is the fact that the sampling interval is so long compared to the dynamics of the system (e.g., the Nor-

wegian Central Bureau of Statistics issues only annual data whereas quarterly data are issued in some other countries). It is a well-known fact that when the time increment is long compared to some of the time constants, only implicit solution methods for differential equations can be employed in order to avoid numerical instability, so in that sense there is nothing surprising about the implicit form given by (1). In fact, as pointed out by Mehra (1974), the use of aggregation concepts and asymptotic expansion methods in large scale systems can also lead to implicit or simultaneous equation models. The implicit model concept is therefore of interest outside the field of econometrics.

The observations of the system are assumed to be of the form

$$y_t = Dx_t + w_t \tag{2}$$

where y_t is the observation vector and $\{w_t\}$ is a gaussian white zero-mean process. Errors in the input variables u_t may be accounted for by proper modeling of v_t.

The model (1) is assumed to consist of N sub-models $\mathscr{S}_i$, $i = 1, \ldots, N$, described by

$$x_{t+1}^i = A_{ii}x_t^i + C_i u_t + \Gamma_i z_{t+1}^i + v_t^i \tag{3}$$

where z_t^i is the interaction input to $\mathscr{S}_i$ from the other sub-models, viz.

$$z_{t+1}^i = \sum_{\substack{j=1 \\ j\neq i}}^{N} L_{ij}x_t^i + \sum_{\substack{j=1 \\ j\neq i}}^{N} M_{ij}x_{t+1}^i = L_i x_t + M_i x_{t+1} \tag{4}$$

where

$$\Gamma_i L_{ij} = A_{ij}, \qquad \Gamma_i M_{ij} = B_{ij} \tag{5}$$

A_{ij} and B_{ij} are, respectively, the (i, j)th block entries of the matrices A and B. Note that we have assumed $B_{ii} = 0, i = 1, \ldots, N$.

The system is said to be output decentralized (Siljak and Vukcevic, 1978) if the observation (or output) equations can be written in the form

$$y_t^i = D_i x_t^i + w_t^i, \qquad i = 1, \ldots, N \tag{6}$$

where y_t^i is the observation vector of sub-system $\mathscr{S}_i$, whereas w_t^i is the observation uncertainty.

Any linear system can be output decentralized, and it is therefore no loss of generality to assume this structure.

3. The TSLS method and some interpretations of the first stage

In this section we assume the available state information to be full (or fully reconstructable) without noise. Consider a stable linear system driven by white noise, and consider the following implicit model

$$x_{t+1} = Ax_t + Bx_{t+1} + v_t \tag{7}$$

of the system. Also, for the sake of simplicity, let us first assume the observation equations to be of the form

$$y_t = x_t \tag{8}$$

We want to estimate the parameters of the model (7), the matrices A and B, which we believe to be the 'true' model of the system. Now, let us try to estimate A and B using an ordinary least squares method with a cost functional of the form

$$J = \sum_t (x_{t+1} - Ax_t - Bx_{t+1})^T (x_{t+1} - Ax_t - Bx_{t+1}) \tag{9}$$

where summation is taken over the available sample points. Minimizing J with respect to A and B, we find

$$\frac{\partial J}{\partial A} = 0 = -2 \sum_t (x_{t+1} - \hat{A}x_t - \hat{B}x_{t+1}) x_t^T = 0$$

$$\frac{\partial J}{\partial B} = 0 = -2 \sum_t (x_{t+1} - \hat{A}x_t - \hat{B}x_{t+1}) x_{t+1}^T = 0,$$

i.e.,

$$\hat{A} \sum_t x_t x_t^T = (I - \hat{B}) \sum_t x_{t+1} x_t^T \tag{10}$$

$$\hat{B} \sum_t x_{t+1} x_{t+1}^T = \sum_t x_{t+1} x_{t+1}^T - \hat{A} \sum_t x_t x_{t+1}^T \tag{11}$$

Multiplying (7) by x_t^T and x_{t+1}^T and summing, we obtain

$$\sum_t x_{t+1} x_t^T = A \sum_t x_t x_t^T + B \sum_t x_{t+1} x_t^T + \sum_t v_t x_t^T \tag{12}$$

$$\sum_t x_{t+1} x_{t+1}^T = A \sum_t x_t x_{t+1}^T + B \sum_t x_{t+1} x_{t+1}^T + \sum_t v_t x_{t+1}^T \tag{13}$$

Now, assume the process is ergodic and stationary. Dividing (12) and (13) by the number of sample points, M say, we obtain

$$\frac{1}{M} \sum_t x_{t+1} x_t^T = A \left(\frac{1}{M} \sum_t x_t x_t \right) + B \left(\frac{1}{M} \sum_t x_{t+1} x_t^T \right) + \frac{1}{M} \sum_t v_t x_t^T \tag{14}$$

$$\frac{1}{M} \sum_t x_{t+1} x_{t+1}^T = A \left(\frac{1}{M} \sum_t x_t x_{t+1}^T \right) + B \left(\frac{1}{M} \sum_t x_{t+1} x_{t+1}^T \right) + \frac{1}{M} \sum_t v_t x_{t+1} \tag{15}$$

where

$$\frac{1}{M} \sum_t v_t x_t^T \xrightarrow[M \to \infty]{} 0 \text{ wp. 1},$$

whereas

$$\frac{1}{M} \sum_t v_t x_{t+1}^T \xrightarrow[M \to \infty]{} E(v_t x_{t+1}^T) \neq 0 \text{ wp. 1}$$

generally.

(10)–(11) and (14)–(15) now yield (wp. 1)

$$\hat{A} \xrightarrow[M \to \infty]{} (I - \hat{B})(I - B)^{-1} A \tag{15}$$

$$\hat{B} \xrightarrow[M \to \infty]{} B + (A - \hat{A}) E(x_t x_{t+1}^T)(E(x_{t+1} x_{t+1}))^{-1} + E(v_t x_{t+1}^T)(E(x_{t+1} x_{t+1}^T))^{-1} \tag{16}$$

which shows that the estimate $\hat{B}$ of B generally must be inconsistent even if $A = \hat{A}$.

In econometrics this problem is resolved by applying the two-stage least squares method (TSLS), originally due to Theil (1953) and Basman (1957), see also Theil (1971). The first stage in this consists of writing (7) in the form

$$x_{t+1} = \Phi x_t + \Gamma v_t \tag{17}$$

where $\Phi = (I - B)^{-1}A$ and $\Gamma = (I - B)^{-1}$. This is the so-called reduced form model, whereas (7) is denoted the simultaneous equations model. Estimating Φ by ordinary least squares techniques using a cost functional like

$$J = \sum_t (x_{t+1} - \Phi x_t)^T (x_{t+1} - \Phi x_t),$$

we obtain

$$\frac{\partial J}{\partial \Phi} = -2 \sum_t (x_{t+1} - \hat{\Phi} x_t) x_t^T = 0, \tag{18}$$

i.e.

$$\hat{\Phi} = \sum_t x_{t+1} x_t^T \left(\sum_t x_t x_t^T \right)^{-1} \tag{19}$$

which is a consistent estimate of Φ. Moreover, a consistent estimator of the covariance matrix of Γv_t is given by

$$\hat{\Xi} = \frac{1}{M} \sum_t (x_{t+1} - \hat{\Phi} x_t)(x_{t+1} - \hat{\Phi} x_t)^T \tag{20}$$

It could be argued that a direct approach to estimating the matrices A and B would be to estimate Φ and Γ using (17) and then to compute $B = I - \Gamma^{-1}$, $A = \Gamma^{-1}\Phi$. This would certainly be a consistent estimator. Note, however, that the structure of the system may be only vaguely known, and that the matrices A and B can be quite sparse. For example, in an econometric model, a variable would normally depend upon only a few other variables, although the lag structure etc. may be unknown. It would be very difficult, probably impossible, to impose such restrictions if A and B should be estimated from $A = \Gamma^{-1}\Phi$ and $B = I - \Gamma^{-1}$. In econometrics one usually has to estimate each equation separately. This is normally done by trying out different variables to be explanatory (i.e., on the right-hand side) in that particular equation, by trying out different lag structures, etc.

Now, let us write x_{t+1} as

$$\begin{aligned} x_{t+1} &= x_{t+1} - \hat{\Phi} x_t + \hat{\Phi} x_t \\ &= \hat{\Phi} x_t + (\Phi - \hat{\Phi}) x_t + \Gamma v_t \end{aligned} \tag{21}$$

which by substitution into (7) yields

$$x_{t+1} = A x_t + B \hat{x}_{t+1} + B(\Phi - \hat{\Phi}) x_t + \Gamma v_t \tag{22}$$

where $\hat{x}_{t+1} = \hat{\Phi} x_t$. Noting that $(\Phi - \hat{\Phi}) \xrightarrow[M \to \infty]{} 0$ and $E(\hat{x}_{t+1} v_t^T) \xrightarrow[M \to \infty]{} 0$ wp. 1, we see that consistent estimates of A and B are obtained by employing ordinary least squares techniques on (22).

Now, returning to (18), note that this equation simply expresses the fact that

$$E\{(x_{t+1} - \hat{\Phi}x_t)x_t^T\} = 0 \tag{23}$$

or

$$X_1 = \hat{\Phi}X_0 \tag{24}$$

where $X_\tau = E(x_{t+\tau}x_t^T)$. This is simply the projection theorem (Kailath, 1968).

In the slightly more general case where the observation equation takes the form

$$y_t = Dx_t, \tag{25}$$

the foregoing cost functionals and equations have to be modified accordingly. Estimates of Φ (and D) will now be obtained from the equations

$$\begin{aligned} \sum_t (y_{t+1} - \hat{D}\hat{\Phi}x_t)y_t^T &= 0 \\ \sum_t (y_{t+1} - \hat{D}\hat{\Phi}^2x_{t-1})y_{t-1}^T &= 0 \\ \sum_t (y_{t+1} - \hat{D}\hat{\Phi}^3x_{t-2})y_{t-2}^T &= 0 \\ &\vdots \end{aligned} \tag{26}$$

which simply express

$$E\{(y_{t+1} - \hat{D}\hat{\Phi}^k x_{t-k+1})y_{t-k+1}^T\} = 0. \tag{27}$$

Making use of $y_t = Dx_t$, (27) can be written as

$$Y_k = \hat{D}\hat{\Phi}^k X\hat{D}^T \tag{28}$$

where $Y_k = E(y_{t+k}y_t^T)$ and $X = E(x_t x_t^T)$. Note that (28) is also valid for $k = 0$ since, from (25), $Y_0 = \hat{D}X\hat{D}^T$.

Equation (28) expresses the fact that estimation of the reduced form model essentially means to compute the matrices D, Φ, and X from the sequence of covariance matrices $Y_0, Y_1, Y_2, \ldots$ (or estimates of these). This brings the first stage of the TSLS method in relation to the stochastic realization problem, see Faurre (1976) or Rissanen and Kailath (1972). What we have actually demonstrated is that the first stage of the TSLS method (as it is being used in econometrics) essentially amounts to constructing a simplified stochastic realization (not general) of the system. This fact was first pointed out by Kalman (1980).

The TSLS method can also be interpreted as an instrumental variable estimator (IV estimator), see Theil (1971). Using, for example, the estimated matrix $\hat{\Phi}$ from (19), we can generate instrumental variables (instruments) from the equation

$$\hat{x}_{t+1} = \hat{\Phi}x_t.$$

The particular IV estimator for this problem, which now replaces the estimator given by (10)–(11), takes the form

$$\hat{A}\sum_t x_t x_t^T = (I - \hat{B})\sum_t \hat{x}_{t+1}x_t^T$$

$$\hat{B}\sum_t \hat{x}_{t+1}\hat{x}_{t+1}^T = \sum_t \hat{x}_{t+1}\hat{x}_{t+1} - \hat{A}\sum_t x_t\hat{x}_{t+1}$$

Since $\hat{\Phi} \xrightarrow[M\to\infty]{} \Phi$ wp. 1 and $E(\hat{x}_{t+1} v_t^T) \xrightarrow[M\to\infty]{} 0$ wp. 1, we now find the asymptotic properties (provided $\Phi E(x_t x_t^T)\Phi^T$ is non-singular)

$$\hat{A} \xrightarrow[M\to\infty]{} (I - \hat{B})\Phi = (I - \hat{B})(I - B)^{-1}A$$

$$\hat{B} \xrightarrow[M\to\infty]{} B,$$

that is, consistency. More details about IV estimators and their properties can be found in Söderström and Stoica (1983).

For more about the TSLS method and its interpretations, see Chow (1964) and Dhrymes (1972).

4. Generation of interaction variable estimates

The form of the system model assumed in the TSLS method (eqns. (7) or (17) and (25)) is unfortunately not general enough to cover all kind of linear stochastic systems, not even stationary systems, and this, of course, limits the applicability of this method. Consider a vector valued weakly stationary process $\{y_t | t = \ldots, -1, 0, 1, \ldots\}$ from which estimates of the covariance matrices $Y_i = \text{cov}\,\{y_{t+i}, y_t\}$, $i = 0, \ldots,$ can be made. A system of the form

$$x_{t+1} = Fx_t + v_t \tag{29}$$

$$y_t = Dx_t + w_t \tag{30}$$

is said to be a Markovian representation of the time series Y_i, $i = 0, 1, \ldots,$ see Faurre (1976), if $\{v_t\}$ and $\{u_t\}$ are stationary white processes satisfying

$$E\{v_t v_s^T\} = Q\delta_{t,s}, \qquad E\{w_t w_s^T\} = R\delta_{t,s}, \qquad E\{v_t w_s^T\} = S\delta_{t,s}$$

Let $X = \text{cov}\,\{x_t\}$. From (29)

$$X = FXF^T + Q \tag{31}$$

and from (30)

$$Y_0 = DXD^T + R \tag{32}$$

Furthermore, from (29)–(30)

$$Y_K = DF^{k-1}K, \qquad k \geq 1 \tag{33}$$

where

$$K = FXD^T + S \tag{34}$$

The stochastic realization problem consists of finding a Markovian representation for a given covariance sequence Y_k, $k = 0, 1, \ldots,$ i.e., finding matrices F, D, K and four covariance matrices X, Q, R, S satisfying (31), (32), (33), and the relations

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \geq 0, \qquad X > 0 \tag{35}$$

A solution is said to be a minimal stochastic realization (MSR) if the dimension of x_t is the least possible.

The solution of the minimal realization problem is not unique. For one thing, the triple of matrices (F, D, K) is only unique modulo a similarity transformation T in the sense that if (F, D, K) is a solution, then $(T^{-1}FT, DT, T^{-1}K)$ is also a solution. For another, the four covariances matrices X, Q, R, S are not unique even when (F, D, K) has been selected.

An interesting observation pointed out by Faurre (1976) is the fact that there exists a minimal solution X_* which is the covariance matrix of the statistical filter $E\{x_t|\mathscr{Y}_{t-1}\}$† of the underlying process $\{x_t\}$. This filter has the form

$$\bar{x}_{t+1} = F\bar{x}_t + Hv_t \tag{36}$$

$$y_t = D\bar{x}_t + v_t \tag{37}$$

where $\{v_t\}$ is the innovation process.

Another representation is the statistical filter $E\{x_t|\mathscr{Y}_t\}$ which has the form

$$\hat{x}_{t+1} = F\hat{x}_t + H_1 v_{t+1} \tag{38}$$

$$y_t = D\hat{x}_t + Ev_t \tag{39}$$

where $H_1 = F^{-1}H$ and $E = I - DH_1$.

Since we only assume weak stationarity, i.e., covariance stationarity, we may have to assume the slightly more general form

$$\bar{x}_{t+1} = F\bar{x}_t + Gu_t + Hv_t \tag{40}$$

instead of (36). In order to find the matrix G and an initial value $\bar{x}_0$, note that (40) and (37) recursively define a set of equations of the form

$$y_k - \sum_{t=0}^{k-1} D\Gamma^{k-t-1}Hy_t = D\Gamma^{k-1}\bar{x}_0 + \sum_{t=0}^{k-1} D\Gamma^{k-t-1}Gu_t + v_k \tag{41}$$

where $\Gamma = F - HD$. Since the process $\{v_t\}$ is white and all variables in (41) except y_k are independent of v_k, we can apply simple least squares techniques to find estimates of G and $\bar{x}_0$ from this set of equations (provided F, H, D and $u_0, u_1, u_2, \ldots$ are known).

Now, after having solved the minimal realization problem, we are able to generate estimates of the interaction variables, which then can be used as instrumental variables in the equations of the original model. Before we proceed to the next section, let us briefly discuss the consequences of assuming a certain model structure, i.e., the class of models which we want our model to belong to, e.g., the class represented by (36)–(37), (38)–(39) or the general type (29)–(30) with $S = E\{v_t w_t^T\} = 0$.

First, let us start by assuming our model to belong to the latter type. The estimates $\hat{z}^i_{t+1|t}$ can then be generated from one of the preceding filter representations. Rewriting (3) as

$$x^i_{t+1} = A_{ii}x^i_t + C_i u_t + \Gamma_i \hat{z}^i_{t+1|t} + \Gamma_i(z^i_{t+1} - \hat{z}^i_{t+1|t}) + v^i_t \tag{42}$$

we see that $z^i_{t+1} - \hat{z}^i_{t+1|t}$ acts as an additional noise input. Furthermore, v^i_t is independent of $\hat{z}^i_{t+1|t}$ since v^i_t, according to the model specification, is independent of data prior to t. Also, since $\hat{z}^i_{t+1|t}$ can be written as a linear function of the innovation vectors $v_0, v_1, \ldots, v_t$, it must be independent of $(z^i_{t+1} - \hat{z}^i_{t+1|t})$ (by the projection

† $\mathscr{Y}_{t-1}$ denotes the set of observations up till time $t-1$.

theorem). However, $z^i_{t+1} - \hat{z}^i_{t+1|t}$ and v^i_t are not independent and $\{z^i_{t+1} - \hat{z}^i_{t+1|t}\}$ is not generally white, so estimation of (42) would be complicated, although a generalized least squares method can be applied.

Second, let us specify the model to belong to the class represented by (36)–(37). Noting that in this case the system is represented by the statistical filter $E\{x_t | \mathscr{Y}_{t-1}\}$ (and that the states of this representation are denoted by $\bar{x}_t$) we obtain

$$z^i_{t+1} = L_i \bar{x}_t + M_i \bar{x}_{t+1}, \tag{43}$$

and (3) is seen to take the form

$$x^i_{t+1} = A_{ii} x^i_t + C_i u_t + \Gamma_i L_i \bar{x}_t + \Gamma_i M_i \bar{x}_{t+1} + v^i_t \tag{44}$$

where, however, $\bar{x}_{t+1}$ and v^i_t are correlated (this is seen from (36)). Writing (44) as

$$x^i_{t+1} = A_{ii} x^i_t + C_i u_t + \Gamma_i L_i \bar{x}_t + \Gamma_i M_i \hat{x}_{t+1|t-1} + \Gamma_i M_i(\bar{x}_{t+1} - \hat{x}_{t+1|t-1}) + v^i_t \tag{45}$$

where

$$\hat{x}_{t+1|t-1} = F\bar{x}_t + Gu_t, \tag{46}$$

we have from (40)

$$\bar{x}_{t+1} - \hat{x}_{t+1|t-1} = Hv_t \tag{47}$$

The term Hv_t is correlated with v^i_t, but Hv_t is white. Furthermore, $\hat{x}_{t+1|t-1}$ is independent of both v_t and v^i_t. Because Hv_t is white, (45) would be easier to estimate than (42).

Third, assume the model is specified to belong to the class represented by (38)–(39). (3) now takes the form

$$x^i_{t+1} = A_{ii} x^i_t + C_i u_t + \Gamma_i L_i \hat{x}_t + \Gamma_i M_i \hat{x}_{t+1} + v^i_t \tag{48}$$

where $\hat{x}_{t+1}$ and v^i_t are correlated. Rewriting (48) as

$$x^i_{t+1} = A_{ii} x^i_t + C_i u_t + \Gamma_i L_i \hat{x}_t + \Gamma_i M_i \hat{x}_{t+1|t} + \Gamma_i M_i(\hat{x}_{t+1} - \hat{x}_{t+1|t}) + v^i_t \tag{49}$$

where

$$\hat{x}_{t+1|t} = F\hat{x}_t + Gu_t, \tag{50}$$

we have from (38)

$$\hat{x}_{t+1} - \hat{x}_{t+1|t} = H_1 v_{t+1} \tag{51}$$

$H_1 v_{t+1}$ is white, but correlated with v^i_t (according to the model specification). Both $\hat{x}_t$ and $\hat{x}_{t+1|t}$ are independent of v_{t+1} and v^i_t.

A crucial point in all this is the construction of a minimal stochastic realization of the system. Several algorithms for solving this problem can be applied, e.g., those proposed by Ho and Kalman (1966) or Rissanen and Kailath (1972), see also Faurre (1976). More recent algorithms are primarily based on the concept of singular value decomposition, see van Zee and Bosgra (1979), Damen and Hajdasinski (1982), Hajdasinski *et al.* (1984), and Aoki (1983), which more easily allows approximate lower dimension solution. It should be stressed, however, that the problem of constructing a minimal stochastic realization may be extremely complicated, at least when only estimates of the covariance matrix series Y_0, Y_1, ... exist. In such cases, it may therefore often be necessary to consider other ways of generating the instrumental variables for the interaction vectors. Mehra (1974) has considered the problem of estimating a canonical representation of the Kalman filter form given by (36)–(37). This is done by writing the canonical form as a set of difference equations in terms

of the input-output variables (y, u, v). The first of these equations is an ARMA model of the first output variable $y_1(t)$ with $u(t)$ and $v(t)$ as inputs. This equation is estimated by the TSLS method (since $y_1(t)$ is correlated with many of the input variables represented by v, such a technique is necessary). The next equation is then estimated using $u(t)$, $v(t)$, and $y_1(t)$ as inputs, and so on (the TSLS method is used all the way). Mehra (1974) also considers the problem of estimating the order of each sub-model, the order of each difference equation in the foregoing procedure.

5. Estimation of model parameters

We shall consider two methods for estimating the parameters in the model, the method of generalized least squares and a prediction error method based on decentralized filtering algorithms. Emphasis will primarily be put on the latter, whereas only brief mention of the former will be made (since this method should be well known).

5.1. Estimation of model parameters by the generalized least squares method

After the first step which was used to generate instrumental variables for the interaction inputs z_{t+1}^i, each sub-model can be rewritten in the form

$$x_{t+1}^i = A_{ii} x_t^i + C_i u_t + \Gamma_i z_{t+1|t}^i + \xi_t^i \tag{52}$$

$$y_t^i = D_i x_t^i + w_t^i \tag{53}$$

where ξ_t^i may be white or coloured, depending upon the model specification (of course, the form of w_t^i also depends upon this). Let us for simplicity assume y_t^i to be scalar. The model (52)–(53) can be rewritten in input-output form as

$$A^i(q^{-1})y_t^i = B^i(q^{-1})u_t + C^i(q^{-1})\hat{z}_{t+1|t}^i = \eta_t^i \tag{54}$$

where $\{\eta_t^i\}$ is a coloured noise term which is independent of $\hat{z}_{t+1|t}^i$. Assume $\{\eta_t^i\}$ can be represented by

$$\eta_t^i = G^i(q^{-1})e_t^i \tag{55}$$

where $\{e_t^i\}$ is white. The model (54) can then be represented by

$$A^i(q^{-1})\tilde{y}_t^i = B^i(q^{-1})\tilde{u}_t^i + C^i(q^{-1})\tilde{\hat{z}}_{t+1|t} + e_t^i \tag{56}$$

where

$$\tilde{y}_t^i = \frac{1}{G^i(q^{-1})}\, y_t^i, \qquad \tilde{u}_t^i = \frac{1}{G^i(q^{-1})}\, u_t, \qquad \tilde{\hat{z}}_{t+1|t}^i = \frac{1}{G^i(q^{-1})}\, \hat{z}_{t+1|t}^i$$

The problem of estimating the polynomials $A^i(q^{-1})$, $B^i(q^{-1})$, $C^i(q^{-1})$ and $G^i(q^{-1})$ can be solved in the way proposed by Clarke (1967).

5.2. Estimation of model parameters using decentralized filtering algorithms and a prediction error formulation

The problem of estimation and filtering in large scale systems has been treated in several papers, e.g., Sanders *et al.* (1974), Sing (1975), Siljak and Vukcevic (1978), and Tacker and Sanders (1980). The main feature about the cited papers is that some

kind of hierarchical or decentralized structure is used to simplify the filtering algorithms.

In the papers by Sanders *et al.* (1974), Tacker and Sanders (1980), and many other papers on the same subject by these authors, a decentralized structure is used in the filtering algorithms for large scale interconnected systems. Their theory is based on the assumption that interaction information is available in each sub-filter, and that only local observations (or measurements) are used in the filtering equations, whereas some centralized information structure may occur in the prediction equations. A crucial point is the assumption that the system satisfies the interaction measurement condition, which essentially means that, in addition to (6), there exists an observation μ_t^i of the form

$$\mu_t^i = H_i x_t^i + z_{t+1}^i + \xi_t^i \tag{57}$$

which is available in sub-system $\mathscr{S}_i$. We shall, for the sake of simplicity, assume $H_i = 0$. There is essentially not much loss of generality in doing so.

Based on the foregoing assumptions it can now be shown that if (A_{ii}, D_i) is an observable pair, a non-empty class $(\mathscr{F}_{SLU})$ of completely decentralized unbiased filters exists, see Sanders *et al.* (1974).

Instead of assuming interaction measurements to be available in $\mathscr{S}_i$, a partially centralized filtering structure with information exchange between the sub-filters has also been considered, see Sanders *et al.* (1978). The results there are similar to the previous ones.

Now, as a first step, let us write down the decentralized filtering equations for our implicit model.

First, assume the system has been specified to belong to the class of statistical filters $E\{x_{t+1} \mid \mathscr{Y}_y\}$. Let

$$\hat{z}_{t+1|t}^i = L_i \bar{x}_t + M_i \hat{x}_{t+1|t-1} \tag{58}$$

where $\bar{x}_t$ is obtained from the model

$$\bar{x}_{t+1} = F\bar{x}_t + Gu_t + Hv_t$$
$$y_t = D\bar{x}_t + v_t$$

whereas

$$\hat{x}_{t+1|t-1} = F\bar{x}_t + Gu_t$$

We then obtain the model form

$$x_{t+1}^i = A_{ii} x_t^i + C_i u_t + \Gamma_i \hat{z}_{t+1|t}^i + \xi_t^i \tag{59}$$

$$y_t^i = D_i x_t^i + w_t^i \tag{60}$$

where

$$\xi_t^i = v_t^i + \Gamma_i M_i H v_t \tag{61}$$

Furthermore, $v_t^i = P_v^i v_t$ and $w_t^i = P_w^i v_t$ (P_v^i and P_w^i are matrices).

We write the local filter for the ith sub-system in terms of the local estimate $\hat{x}_{t|t-1}^i$ of x_t^i. This filter has the form

$$\hat{x}_{t+1|t}^i = A_{ii} \hat{x}_{t|t-1}^i + C_i u_t + \Gamma_i \hat{z}_{t+1|t}^i + K_t^i[y_t^i - \hat{y}_{t|t-1}^i] \tag{62}$$

where $\hat{y}^i_{t|t-1} = D_i \hat{x}^i_{t|t-1}$. Specifying K^i_t to be such that $(x^i_{t+1} - \hat{x}^i_{t+1|t})$ is uncorrelated with the local innovations process $\varepsilon^i_t = y^i_t - \hat{y}^i_{t|t-1}$,

$$E\{(x^i_{t+1} - \hat{x}^i_{t+1|t})(\varepsilon^i_t)^T\} = 0 \tag{63}$$

we find

$$K^i_t = (A_{ii} X^i_{t|t-1} D_i^T + P_i)(D_i X^i_{t|t-1} D_i^T + W_i)^{-1} \tag{64}$$

where $X^i_{t|t-1} = \text{cov}\,\{x^i_t - \hat{x}^i_{t|t-1}\}$, $P_i = \text{cov}\,\{\xi^i_t, w^i_t\}$, and $W_i = \text{cov}\,\{w^i_t\}$. The covariance matrix $X^i_{t|t-1}$ satisfies the equation

$$X^i_{t+1|t} = (A_{ii} - K^i_t D_i) X_{t|t-1} (A_{ii} - K^i_t D_i)^T + \Xi_i + K^i_t W_i (K^i_t)^T - P_i (K^i_t)^T - K^i_t P_i^T \tag{65}$$

where $\Xi_i = \text{cov}\,\{\xi^i_t\}$.

Provided (A_{ii}, D_i) is an observable pair it can be shown (Sanders *et al.*, 1974) that the foregoing filter is stable, unbiased, and has a steady-state solution.

The external input to the process $x^i_t - \hat{x}^i_{t|t-1}$ is purely white noise, and it can therefore be shown that the local prediction error, or local innovations process

$$\varepsilon^i_t = y^i_t - \hat{y}^i_{t|t-1} = y^i_t - D_i \hat{x}^i_{t|t-1} \tag{66}$$

is a white process with covariance matrix

$$\text{cov}\,\{\varepsilon^i_t\} = D_i X^i_{t|t-1} D_i^T + W_i \tag{67}$$

Second, assume that system has been specified to belong to the class of statistical filters $E\{x_t | \mathscr{Y}_t\}$. Let

$$\hat{z}^i_{t+1|t} = L_i \hat{x}_t + M_i \hat{x}_{t+1|t} \tag{68}$$

where $\hat{x}_t$ is obtained from the model

$$\hat{x}_{t+1} = F\hat{x}_t + Gu_t + H_1 v_{t+1}$$

$$y_t = D\hat{x}_t + Ev_t$$

whereas

$$\hat{x}_{t+1|t} = F\hat{x}_t + Gu_t.$$

We then obtain the model form

$$x^i_{t+1} = A_{ii} x^i_t + C_i u_t + \Gamma_i \hat{z}^i_{t+1|t} + \xi^i_t \tag{69}$$

$$y^i_t = D_i x^i_t + w^i_t \tag{70}$$

where

$$\xi^i_t = v^i_t + \Gamma_i M_i H_1 v_{t+1} \tag{71}$$

Furthermore, $v^i_t = Q^i_v v_{t+1}$ and $w^i_t = Q^i_w v_t$.

We now write the local filter for the ith sub-system in terms of the local estimate $\hat{x}^i_{t|t}$ of x^i_t. The form of this filter is assumed to be

$$\hat{x}^i_{t+1|t+1} = A_{ii} \hat{x}^i_{t|t} + C_i u_t + \Gamma_i \hat{z}^i_{t+1|t} + K^i_{t+1}[y^i_{t+1} - \hat{y}^i_{t+1|t}] \tag{72}$$

where $\hat{y}^i_{t+1|t} = D_i \hat{x}^i_{t+1|t}$. Again, specifying K^i_t to be such that $(x^i_{t+1} - x^i_{t+1|t+1})$ is uncorrelated with the local innovations process $\varepsilon^i_{t+1} = y^i_{t+1} - \hat{y}^i_{t+1|t}$, we find

$$K^i_{t+1} = (X^i_{t+1|t} D_i^T + P_i)(D_i X^i_{t+1|t} D_i^T + D_i P_i + P_i^T D_i^T + W_i)^{-1} \tag{73}$$

where

$$X^i_{t+1|t} = A_{ii} X^i_{t|t} A^T_{ii} + \Xi_i \tag{74}$$

$$X^i_{t+1|t+1} = (I - K^i_{t+1} D_i) X^i_{t+1|t} - K^i_{t+1|t} P^t_i \tag{75}$$

Provided (A_{ii}, D_i) is an observable pair, it can be shown (Sanders *et al.*, 1974) that the latter filter is stable, unbiased, and has a steady-state solution.

The local prediction error

$$\varepsilon^i_t = y^i_t - \hat{y}^i_{t|t-1} = y^i_t - D(A_{ii} \hat{x}^i_{t-1|t-1} + G_i u_t + \Gamma_i \hat{z}^i_{t|t-1}) \tag{76}$$

is also in the case a white process with covariance matrix

$$\text{cov}\,\{\varepsilon^i_t\} = D_i X^i_{t|t-1} D^T_i + D_i P_i + P^T_i D^T_i + W_i \tag{77}$$

With any of the two foregoing filtering algorithms, we can formulate a local prediction error model of the form

$$\varepsilon^i_t = y^i_t - f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t) \tag{78}$$

where

$$E\{\varepsilon^i_t | \mathscr{Y}_{t-1}, \mathscr{U}_{t-1}\} = 0 \tag{79}$$

(actually ε^i_t is white, as pointed out, but (79) is sufficient for the prediction error method to work properly).

For a completely localized structure, f_i will depend only on a local parameter vector θ_i, viz.

$$f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t) = f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t; \theta_i).$$

The foregoing filters are obviously of this type.

For a particular value of θ_i, say $\theta_i = \hat{\theta}_i$, the prediction error is given by

$$\eta^i_t(\hat{\theta}_i) = y^i_t - f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t; \hat{\theta}_i) \tag{80}$$

and if θ^0_i is the true parameter value, then

$$\eta^i_t(\theta^0_i) = z^i_t - f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t; \theta^0_i) = \varepsilon^i_t \tag{81}$$

Let us introduce the two criteria

$$J^i_1(\theta_i, M) = \text{trace}\,[S_i H^i_M(\theta_i)], \qquad S_i > 0 \tag{82}$$

$$J^i_2(\theta_i, M) = \log \det\,[H^i_M(\theta_i)] \tag{83}$$

where

$$H^i_M(\theta_i) = \frac{1}{M} \sum_{t=1}^{M} \eta^i_t(\theta_i)(\eta^i_t(\theta_i))^T \tag{84}$$

Let us furthermore assume that θ^0_i is unique in the sense that there exists no other parameter $\theta^1_i \neq \theta^0_i$ such that $f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t; \theta^0_i) = f_i(\mathscr{Y}_{t-1}, \mathscr{U}_{t-1}, t; \theta^1_i)$ wp. 1. The important properties of prediction error identification methods, which have been proved in several other works, see Ljung (1976) and Caines and Ljung (1976), can now readily be transferred to out decentralized prediction error method. For an overview of the properties of prediction error identification methods, see also Goodwin and Payne (1977).

Two of the most important properties are summarized here.

Theorem

Subject to mild regularity conditions, the local prediction error estimator $\hat{\theta}_i(M)$ obtained by minimizing either $J_1^i(\theta_i, M)$ or $J_2^i(\theta_i, M)$ is strongly consistent,

$$\hat{\theta}_i(M) \xrightarrow[M\to\infty]{} \theta_i^0 \text{ wp. 1} \tag{85}$$

Furthermore, subject to mild regularity and stationarity conditions, $\hat{\theta}_i(M)$ is asymptotically normally distributed in the sense that

$$\sqrt{(M)}(\hat{\theta}_i(M) - \theta_i^0) \xrightarrow[M\to\infty]{\text{law}} \gamma_i \tag{86}$$

where $\gamma_i \sim N(0, \Sigma^i)$.

Proof

Essentially the same as Ljung (1976) and Caines and Ljung (1976), see also Goodwin and Payne (1977).

6. Identifiability of large scale systems using decentralized estimation

In the previous section we assumed the function f_i to be unique with respect to θ_i. The implications of this assumption should not be overlooked. There are certain restrictions on the input, certain types of noise-free linear feedback must be excluded. However, the outlined estimation method utilizing decentralized estimation structures is in fact an extreme case of identification in closed loop, since z_t^i is certainly generated by feedback. Identification in closed loop has been investigated in several papers, see Gustavsson *et al.* (1974, 1975, 1976), Ng *et al.* (1977), Sin and Goodwin (1980), Gevers (1976), Anderson and Gevers (1979, 1982), and Caines and Chan (1976). A review of the concept and some of the results are given in Goodwin and Payne (1977).

Now, assume we attempt to estimate the parameters in sub-system $\mathscr{S}_i$, and let $\mathscr{S}_i^e$ denote the part of the system which is external to $\mathscr{S}_i$, the remainder of the system. $\mathscr{S}_i^e$ contains the feedback to sub-system $\mathscr{S}_i$, and let $\mathscr{G}_i$ denote the part of $\mathscr{S}_i^e$ which constitutes the feedback to $\mathscr{S}_i$.

First, consider a single-input, single-output system described by

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})\varepsilon_t \tag{87}$$

where $\{\varepsilon_t\}$ is a white sequence, whereas $\{y_t\}$ and $\{u_t\}$ are the input and output sequences, respectively. The input $\{u_t\}$ is assumed to be generated by the feedback law

$$E(q^{-1})u_t = F(q^{-1})y_t \tag{88}$$

The polynomials A, B and C are assumed to be relatively prime, and the same assumption is made about the polynomials E and F. Furthermore, the polynomials $(AE - q^{-k}BF)$ and CE are also assumed to be relatively prime (the feedback does not introduce pole/zero cancellations).

The identifiability conditions, under the foregoing conditions, have been closely analysed in the paper by Gustavsson *et al.* (1975). A major result is, loosely speaking, that identification in closed loop will give consistent estimates provided the

controller, that is (88) is sufficiently complex. This can be achieved by switching between several different linear controllers, or by increasing the order of the controller, e.g., if the order of both A, B and C is n_1, if the order of both E and F is n_2, and if $k = 0$, then a sufficient condition for identifiability of A, B and C is that the order of the controller be greater than or equal to the order of the forward path,

$$n_2 \geq n_1 \tag{89}$$

If a prediction error identification method is used, it is furthermore shown that *a priori* knowledge about the controller is not needed (or even that it is known that the system is operating in closed loop), the forward path can be estimated as if the system is operating in open loop.

In the multiple-input, multiple-output case, the major result is essentially the same (Ng *et al.*, 1977), a sufficient condition for identifiability is that the minimum observability index (Wolovich, 1974) of the feedback be greater than or equal to the maximum observability index (Wolovich, 1974) of the forward path.

Now, returning to our own decentralized identification method, we note that normally $\mathscr{S}_i^e$ (or more correctly, $\mathscr{G}_i$) is of a much higher order than $\mathscr{S}_i$, e.g., in an econometric model where each equation is estimated separately, $\mathscr{S}_i^e$ will consist of all the other equations of the model. Naturally one would therefore expect the preceding conditions for closed loop identifiability to be fulfilled. We therefore conclude that the system, using the outlined prediction error method, is identifiable provided, loosely speaking, that the sub-system $\mathscr{S}_i$ is small enough compared to $\mathscr{G}_i$.

7. Conclusion

We have considered the problem of estimating large scale implicit (non-recursive) models by two-stage methods. The first stage consisted of constructing a minimal stochastic realization of the system, and we have shown the connection between this and the first stage of the original TSLS-method. The minimal stochastic realization is used to generate estimates (instrumental variables) of the interaction variables, which subsequently are used in the second stage where each sub-system is estimated separately. This latter stage may be carried out by utilizing a generalized least squares method, or, as derived in our paper, by utilizing decentralized filtering algorithms and a prediction error formulation. Some comments on the identifiability of large scale systems using decentralized estimation are also made, and it was shown that the system under fairly meaningful conditions would be identifiable even if the feedback was neglected, provided the outlined prediction error method was utilized.

We have so far not touched upon the problem of estimating the covariance matrix sequence $Y_\tau = \text{cov}\,\{y_t, y_{t+\tau}\}$, $\tau = 0, 1, \ldots$. Since our method is based on the construction of a minimal stochastic realization of this sequence, it is essential that we have good estimates of the sequence $Y_0, Y_1, \ldots$. However, it may be very hard, or even impossible, to obtain good estimates of these covariance matrices because the number of data points could be very low, e.g., a model of the post-war Norwegian national economy could only be estimated in a meaningful manner using annual data from 1950 to 1983. Another problem, noted by Caines and Chan (1976) and also observed by Wieringa (1984), is the fact that the state transition matrix Φ in a minimal stochastic realization may be very sensitive to small variations in the covariance data, often resulting in Φ being unstable. This is a serious difficulty since

our assumption about weak stationarity is based on the fact that Φ must be stable. Also, as pointed out by Caines and Chan (1976), increasing the sample sizes does not seem to increase the likelihood of obtaining a stable matrix. Caines and Chan (1976) found the algorithms of Faurre (1976) to be numerically satisfactory for matrices Φ with eigenvalues of modulus less than approximately 0·85. A similar thing was observed by Wieringa (1984), who pointed out that serious problems occur when some eigenvalues are close to 1 (in modulus), whereas good results are obtained in systems with smaller eigenvalues.

We have restricted ourselves to linear models. The methods are, however, with some modifications, also applicable to non-linear implicit models by employing various non-linear filtering techniques, see Henriksen (1979).

References

Anderson, B. D. O., and Gevers, M. R. (1979). Identifiability of closed-loop systems using the joint input/output identification method. *Proc. Vth IFAC Symposium on Identification and System Parameter Estimation*, Darmstadt, F. R. Germany, p. 645, Pergamon Press.

Anderson, B. D. O., and Gevers, M. R. (1982). Identifiability of linear stochastic systems operating under linear feedback. *Automatica*, **18**, 195.

Aoki, M. (1983). *Notes in Econometric Time Series Analysis: System Theoretic Perspectives.* Springer-Verlag.

Basman, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*, **25**, 77.

Caines, P. E., and Ljung, L. (1976). Asymptotic normality and accuracy of prediction error estimators. Research Report No. 7602, University of Toronto, Department of Electrical Engineering.

Caines, P. E., and Chan, C. W. (1976). Estimation, identification and feedback. In *System Identification: Advances and Case Studies*, (editors. R. K. Mehra and D. G. Lainiotis), p. 349, Academic Press.

Chow, C. G. (1964). A comparison of alternative estimators for simultaneous equations. *Econometrica*, **32**, 532.

Clarke, D. W. (1967). Generalized least-squares estimation of the parameters of a dynamic model. *Proc. IFAC Symposium on Identification of Automatic Control Systems*, Prague, Czechoslovakia, Paper No. 3.17.

Damen, A. A. H., and Hajdasinski, A. K. (1982). Practical tests with different approximate realizations based on the singular value decomposition of the Hankel matrix. *Proc. VIth IFAC Symposium on Identification and System Parameter Estimation*, Arlington, Virginia, USA, p. 809, Pergamon Press.

Dhrymes, P. J. (1972). Simultaneous equations inference in econometrics. *IEEE Trans. Automat. Control*, **AC-17**, 427.

Faurre, P. L. (1976). Stochastic realization algorithms. In *System Identification: Advances and Case studies*, (editors R. K. Mehra and D. G. Lainiotis), p. 1, Academic Press.

Gevers, M. R. (1976). On the identification of feedback systems. *Proc. IVth IFAC Symposium on Identification and System Parameter Estimation*, Tbilisi, USSR, Part 3, p. 625.

Goodwin, G. C., and Payne, R. L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press.

Gustavsson, I., Ljung, L., and Söderström, T. (1974). Identification of linear multivariable systems operating under linear feedback control. *IEEE Trans. Automat. Control*, **AC-19**, 836.

Gustavsson, I., Ljung, L., and Söderström, T. (1975). Identifiability condition for linear systems operating in closed loop. *Int. J. Control*, **21**, 243.

Gustavsson, I., Ljung, L., and Söderström, T. (1976). Identification of processes in closed loop—identifiability and accuracy aspects. *Proc. IVth IFAC Symposium on Identification and System Parameter Estimation*, Tbilisi, USSR, Part 1, p. 23.

HAJDASINSKI, A., VAN DEN HOF, P., and DAMEN, A. (1984). Naive approximate realizations of noisy data. *Proc. 9th IFAC World Congress*, Budapest, Hungary, Vol. X, p. 90.

HENRIKSEN, R. (1979). Nonlinear filtering in econometric models. Research Memorandum No. 240, Princeton University, Econometric Research Program, Princeton, New Jersey.

HO, B. L., and KALMAN, R. E. (1966). Effective construction of linear state-variable models from input/output functions. *Regelungstechnik*, **14,** p. 545.

KAILATH, T. (1968). An innovations approach to least-squares estimation. Part I: Linear filtering in additive white noise. *IEEE Trans. Automat. Control*, **AC-13,** 646.

KALMAN, R. E. (1980). Identifiability and problems of model selection in econometrics. *Proc. 4th World Congress of the Econometric Society*, Aix-en-Provence, France.

LJUNG, L. (1976). On the consistency of prediction error identification methods. In *System Identification: Advances and Case Studies*, (editors R. K. Mehra and D. G. Lainiotis), p. 121, Academic Press.

MEHRA, R. K. (1974). Identification in control and econometrics; similarities and differences. *Annals of Economic and Social Measurement*, **3,** p. 21.

NG, T. S., GOODWIN, G. C., and ANDERSON, B. D. O. (1977). Identifiability of MIMO linear systems operating in closed loop. *Automatica*, **13,** p. 477.

RISSANEN, J., and KAILATH, T. (1972). Partial realization of random systems. *Automatica*, **8,** p. 389.

SANDERS, C. W., TACKER, E. C., and LINTON, T. D. (1974). A new class of decentralized filters for interconnected systems. *IEEE Trans. Automat. Control*, **AC-19,** p. 259.

SANDERS, C. W., TACKER, E. C., and LINTON, T. D. (1978). Information exchange in decentralized filter via interaction estimates. *Proc. 7th IFAC Triennial World Congress*, Helsinki, Finland, p. 1367, Pergamon Press.

SILJAK, D. D., and VUKCEVIC, M. B. (1978). On decentralized estimation. *Int. J. Control*, **27,** p. 113.

SIN, K. S., and GOODWIN, G. C. (1980). Checkable conditions for identifiability of linear systems in closed loop. *IEEE Trans. Automat. Control*, **AC-25,** p. 722.

SINGH, M. G. (1975). Multi-level state estimation. *Int. J. Syst. Sci.*, **6,** p. 535.

SÖDERSTRÖM, T., and STOICA, P. G. (1983). *Instrumental Variable Methods for System Identification*. Springer-Verlag.

TACKER, E. C., and SANDERS, C. W. (1980). Decentralized structures for state estimation in large scale systems. *Large Scale Systems*, **1,** p. 39.

THEIL, H. (1953). Repeated least squares applied to complete equation systems. Central Planning Bureau, The Hague, The Netherlands.

THEIL, H. (1971). *Principles of Econometrics*. John Wiley & Sons.

VAN ZEE, G. A., and BOSGRA, O. H. (1979). The use of realization theory in the robust identification of multivariable systems. *Proc. Vth IFAC Symposium on System Identification and Parameter Estimation*, Darmstadt, F. R. Germany, p. 477, Pergamon Press.

WIERINGA, J. (1984). Minimal partial realization of covariance stationary sequences. Diploma thesis, The Norwegian Institute of Technology, Division of Engineering Cybernetics.

WOLOVICH, W. A. (1974). *Linear Multivariable Systems*. Springer-Verlag.