ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 6/7 — 1987

Studies of Economic Interdependence by State Space Modeling of Time Series: US-Japan Example

Masanao AOKI *

ABSTRACT. – This paper proposes a new way of studying interdependence of countries by building models for selected macroeconomic time series of the countries involved, by a state space model which incorporates moving average components in addition to autoregressive ones. The singular value decomposition is performed of the covariance matrix between the data and future realization of the data vector. A crucial step is the choice of a particular coordinate system which resolved indentifiability question and leads to estimates of the model system matrices which enjoy certain nesting or orthogonality properties. The estimates are consistent even when the state vector dimension is unspecified. The paper then discusses the possible presence of unit roots in the time series and develops a canonical representation of time series containing unit root components. Unit roots are removed by prior detrending in this paper. (The author has since developed a two-step procedure that eliminates the need for this prior detrending in AOKI (1987a)). The method is then applied to study interaction of the industrial production indices and *ex ante* real interest rates in the USA and Japan.

Études de l'interdépendance économique par la modélisation espace-état : l'exemple des États-Unis et du Japon

RÉSUMÉ. – Cet article propose une nouvelle façon d'étudier l'interdépendance des pays en construisant des modèles pour des séries bien choisies des pays concernés et en utilisant un modèle espace-état qui incorpore des composantes moyennes mobiles en plus des composantes autorégressives. La décomposition en valeurs singulières est appliquée à la matrice de covariance entre les données et les réalisations futures du vecteur des données. Une étape cruciale est le choix d'un système de coordonnées qui résout l'identifiabilité et conduit à des estimations des matrices du système satisfaisant des propriétés d'emboîtement et d'orthogonalité. Le papier discute ensuite la présence de racines unité dans les séries et développe une représentation canonique des séries ayant une racine unité. Les racines unités sont supprimées en ôtant la tendance (l'auteur a développé depuis une méthode en deux étapes qui évite cet extraction préliminaire, voir AOKI (1987a)). La méthode est appliquée ensuite à l'interaction entre les indices de la production industrielle et les taux d'intérêt réels *ex-ante* aux USA et au Japon.

* M. AOKI: University of California, 4731 Bœlter Hall, University of California, Los Angeles, California 90024, USA. The research reported in this paper was supported in part by the NSF grant SES 85 08920. I thank two anonymous references of this journal and A. Havenner for very constructive comments.

1 Introduction

Traditionally, economic structure of a country is modeled by structural econometric models. Comparison of economic structure of several countries requires major econometric efforts by this approach. Less detailed but computationally far less demanding alternative way of capturing gross dynamic characteristics of economic structure is to construct models for a few key macroeconomic time series of several countries and to examine the models for evidence of interdependence. We describe a new state space modeling procedure for vector-valued time series as a superior alternative to vector autoregressive models. As will become evident, the proposed modeling procedure is very easy to implement computationally, and is ideal as a preliminary screening tool for more detailed structural modeling. Although theoretically fully equivalent to constructing ARMA(X) or VAR models, state space models differ from these models in numerical and statistical properties. The most significant advantage is that constructed models are asymptotically stable.

State space approach is useful because models often provide supplementary information on gross characterization of dynamic interdependence, as will be illustrated in this paper. A gradual but steady increase is observed in the literature of a class of models variously called as dynamic factor models, (unobserved) index models and the like. This class of models is a subset of state space models. In the literature, however, this connection and the essential nature of state space seems not sufficiently realized. For example, models with a few factors or indices, certainly less than the dimension of the data vector, seem to be favored for no compelling reason. [1]

The key of state space modeling is the notion of state vector. State vectors may be thought of informally as dynamic factors. More appropriately, however, they are (approximate) conditional sufficient statistics that summarizes the information contained in data. This set of sufficient statistics is updated as more data points are generated with passing time. It is known that state spaces are the space of conditionally sufficient statistics in the sense that the amount of information contained in the past realization of data is equivalently contained in the state vector, i. e., they span the same subspace in the Hilbert space of centered random vectors in the case of centered Gaussian random vectors, under some technical conditions. [2].

It is also known that if a weakly stationary (centered) vector-valued stochastic process possesses a rational spectral matrix, then finite-dimensional state space models exist. (If a stochastic process possesses irrational spectral density matrix, no finite dimensional state space models can produce the density matrix of the process. One then is faced with the problem of approximating irrational spectra with rational spectra. (We do not discuss this problem here.) When the process possesses a rational but complex density matrices, i. e., elements of the transfer matrices are ratios of high order polynomials, how to approximate them with lower dimensional ones is an importance practical problem. Here a consequence of alternative definitions of state vectors as (approximate) sufficient summaries of information in the data set and alternative choices of dimensions of the state vector

of approximating models can be evaluated in terms of certain upper and lower bounds of the norms of errors in approximating a time transfer function by an approximate one. [3] Singular values of some Hankel matrices we construct in Section 2 are used to define a norm for transfer matrices used in this method. (Adoption of mutual information as an alternative norm leads to the well-known Akaike model.) To compare dynamic structures of state space models is to compare mechanisms for updating sufficient statistics, i. e., how new information is incorporated into sufficient statistics. Building state space models is a useful way of structural comparisons not easily obtainable from econometric structural or other models.

AKAIKE [1974] was first to adopt a state space framework for time series modeling. The way of implementing state space models adopted in this paper, however, differs from Akaike's in one significant aspect: From an equivalent class of state space representations, we choose a specific coordinate system to produce the unique representation corresponding to a given sequence of data covariance matrices. This representation produces dynamic matrix A and the observation matrix C of the innovation model (described in Section 2) which are nested in the sense that system matrices corresponding to lower dimensional state vectors are appropriate submatrices of the higher dimensional models. Put differently, to produce larger-dimensional model, original system matrices are left unaltered, much in the way of the coefficients of expansion by a system of orthogonal function are invariant with respect to the number of terms in expansion approximating a function, and only additional submatrices to augment the existing system matrices are newly computed as discussed in Section 2. Only the associated Riccati equation need to be resolved when the state vector dimension is altered. This property can then be used to evaluate incremental contribution of individual sufficient statistic (i. e., individual state vector components) in the (one-step-ahead) prediction of future data points.

Section 2 is a quick summary of the modeling method used in this paper. It is worth repeating that the models are by construction asymptotically stable; a property not guaranteed in VAR models. For detailed exposition of this and other potentially advantageous properties of the method, see AOKI [1987]. Section 3 discusses the unit roots of time series models in the context of state space models. Section 4 describes the results of numerical experiments to construct models for interactions between Japan and the USA through the industrial production indices and *ex ante* real interest rates. The paper concludes with Section 5.

---

1. GEWEKE [1977] calls his model dynamic factor models. SARGENT and SIMS [1977] use the term index model. QUAH [1986] also calls his model index model. Engle and Watson call their model dynamic multiple indicator-multiple cause model (DYMIMIC) which is a mixed observable-unobervable index model in Sargent and Sims terminology. SINGLETON [1980] singled out two factors out of five data series. Both ENGLE and WATSON [1981] and QUAH [1986] use single index models. ENGLE and WATSON [1981, p. 775] state that their models are special cases of the state space model without mentioning the conditionally sufficient statistics aspect of such models.

2. The notion of the conditionally sufficient statistics is called the minimal splitting subspace by PICCI [1976] and LINDQUIST, PICCI and RUCKEBUSCH [1979]. Ruckebusch proved that the state space of an observable model with an invertible dynamic matrix is a minimal splitting subspace. See also Ibragimv and Rozanov on minimal σ-algebra.

3. See AOKI [1987], and GLOVER [1984] for example.

2 The Method and its Properties

We assume that time series are mean zero and weakly stationary and that data generating processes belong to the class of finite dimensional dynamics. Without loss of generality [4] we can use

$$z_{t+1} = A z_t + B e_t, \qquad \text{and} \qquad y_t = C z_t + e_t, \tag{1}$$

where e_t is uncorrelated with past y's as the representation of the data generating process.

We say that the model is minimal when (1) is both observable and reachable. The state space of the model is the same as the minimal splitting subspace of the Hilbert space spanned by all future and all past data vectors. See LINDQUIST, PICCI and RUCKEBUSH [1979] or IBRAGIMOV and ROZANOV [1978] on minimal splitting subspaces. When the matrix A is invertible, then obervability is necessary and sufficient for minimal realization.

To summarize the method of this paper for constructing the system matrices A, B, and C from the data set, it is convenient to start with the two infinite-dimensional vectors of stacked future and past data vectors

$$y_t^+ = \begin{bmatrix} y_t \\ y_{t+1} \\ \cdots \end{bmatrix} \qquad \text{and} \qquad y_t^- = \begin{bmatrix} y_t \\ y_{t-1} \\ \cdots \end{bmatrix},$$

and consider forecasting y_t^+ from y_{t-1}^- by orthogonal projection, [5]

$$\hat{E}(y_t^+ \mid y_{t-1}^-) = H R^{-1} y_{t-1}^-$$

where H is the covariance matrix $E(y_t^+ y_{t-1}^{-\prime})$. This matrix has submatrices $\Lambda_l = E(y_{t+l} y_t')$, $l = 1, 2, \ldots$, arranged into couter-diagonal arrays, and is a Hankel matrix. The matrix R arranges Λ_0 and Λ_l in the diagonal fashion and is a Tœplitz matrix.

The model (1) implies that the covariance matrices have the structure

$$\Lambda_l = C A^{l-1} M, \qquad l = 1, 2, \ldots \tag{2}$$

where the matrix M is defined to be

$$M = A \Pi C' + B \Delta,$$

where $\Pi = \mathrm{cov}(z_t)$, and $\Delta = \mathrm{cov}(e_t)$. We use two factorizations of the Hankel matrix. One is the well-known singular value decomposition, and will be introduced shortly. The other is obtained by noting the relation (2). Because of the structure given by (2) the Hankel matrix can be factored as

$$H = O \Omega \tag{3}$$

where the matrix O, called the observability matrix in the system literature, has an infinite number of rows

$$O=\begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \\ \cdots \end{bmatrix},$$

and the matrix Ω has an infinite number of columns

$$\Omega=[M\ AM\ A^2M\ A^3M\ \ldots]$$

The model representation (1) is observable when the rank of the matrix O is the same as the dimension of the state vector.

The product expression $G_0=O'O$ is called the observability gramian in the systems literature. [6] This matrix is well-defined and solves a matrix Lyapunov equation

$$A'XA-X=-C'C \tag{4}$$

when A is asymptotically stable. Note that G_0 is positive definite when (1) is observable. Another matrix product $G_c=\Omega'\Omega$ is called the reachability gramian. It is the solution of a Lyapunov equation

$$AXA'-X=-MM'. \tag{5}$$

We mention these gramians because they will be used in selecting a certain coordinate system we use later.

In practice we use finite sections of the stacked y vectors. For example we may be interested in forecasting $y_t, \ldots y_{t+J-1}$ from a finite data set $y_{t-1}, \ldots, y_{t-K}$. Instead doubly of infinite dimensional Hankel matrix, we have $Jp\times Kp$ submatrix $H_{J,K}$ and its factorization involves also finite segments O_J and Ω_K,

$$H_{J,K}=\begin{bmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_2 & \Lambda_3 \\ \cdot & \cdot \\ \Lambda_J & \Lambda_{J+1} \end{bmatrix} \vdots \begin{bmatrix} \Lambda_K \\ \Lambda_{K+1} \\ \cdot \\ \Lambda_{J+K-1} \end{bmatrix} : Jp\times Kp$$

Note that the covariance matrix between y^+_{t+1} and y^-_{t-1} is

$$H^A=OA\,\Omega. \tag{6}$$

4. Model representations in which different noise processes appear in the dynamic and observation equations can be put into the form of (1) by a suitable factorization of spectral density matrices. See AOKI [1983, 1987].

5. If one uses, instead, a "normalized" stacked data vector, then the orthogonal projection produces a normalized Hankel matrix rather than H itself leading to the canonical correlation method by AKAIKE [1974] for example.

6. This is somewhat of a misnomer. Gramians usually denote matrices made up of inner products of a set of vectors.

the first submatrix row of H, denoted by H^C, is

$$H^C = C\Omega \tag{7}$$

and the first submatrix column of H, denoted by H^M, is related to O by

$$H^M = OM. \tag{8}$$

The system matrices A, C, and M must be chosen to satisfy (6) through (8). We show how to determine B later. These equations can be solved easily by using the singular value decomposition of the Hankel matrix $H_{J,K}$. For expositional ease, we assume that $J=K$ from now on and drop the subscript. Compare the singular value decomposition

$$H = U\Sigma V'$$

with the factorization in terms of O and Ω of (3), where Σ is a diagonal matrix with singular values in decreasing order as the diagonal elements, and U and V are orthogonal matrices of compatible dimension. Theoretically, the number of positive singular values is equal to the theoretical rank of the Hankel matrix. Numerical errors tend to cause the Hankel matrix to have full rank. Some selection criteria must be used to judge if some of the singular values are effectively zero or statistically insignificantly different from zero. For the brevity of the exposition of our method, we do not discuss this aspect in this paper. See AOKI [1987] for example where several possible criteria for deciding on the effective rank are discussed. Now, choose a coordinate system in which

$$O = U\Sigma^{1/2}, \qquad \text{and} \qquad \Omega = \Sigma^{1/2} V'.$$

Note that the gramians are equal and diagonal in this coordinate system

$$O'O = \Sigma = \Omega\Omega'.$$

In other words, we choose the coordinate system in which the two Lyapunov equations have the same diagonal matrix as the solutions. (See AOKI [1987, Section 5.6] for the actual coordinate transformation to accomplish this.)

In this coordinate system (6) through (8) have the solutions

$$C = H^C V \Sigma^{-1/2}, \qquad M = \Sigma^{-1/2} U' H^M, \tag{9}$$

and

$$A = \Sigma^{-1/2} U' H^A V \Sigma^{-1/2}. \tag{10}$$

For the reasons explained below, model (1) with these system matrices is called balanced. Putting models in balanced representation or balanced form uniquely determine the system matrices, i. e., the model is uniquely identified. The notion of balanced representation is due to MOORE [1978]. The word "balanced" in the above description comes from the fact that the representation balances two sources of errors: one related to the observability, i. e., error in reconstructing past state vector values from future observations, and the other errors related to reachability, i. e., the error due to past prediction errors.

To see that (9) and (10) solve (6) through (8) uniquely, consider a matrix equation

$$Q\,vec\,X = vec\,Y$$

for any matrices X, Y, and Q. Its solution is expressible as

$$vec\,X = Q^{+}\,vec\,Y + (I - Q^{+}\,Q)\,vec\,P$$

where *vec* P is arbitrary, and Q^{+} is the Moore-Penrose inverse of Q, if and only if $QQ^{+}\,vec\,Y = vec\,Y$. Apply the above to (6) as an example. It is the same as

$$(\Omega' \times I)\,vec\,C = vec\,H^{C}.$$

Here $V\Sigma^{1/2} \times I$ is the matrix Q. This has the unique solution because $I - (\Sigma^{-1/2}\,V' \times I)\,(V\Sigma^{1/2} \times I)$ is zero. The other two equations can be similarly analyzed.

The balanced model representation uses the system matrices computed in (9) and (10). It has several useful properties. One is that system matrices A, C and M of any lower-dimensional balanced model than a given one is obtained by merely taking a leading principal submatrix of the original dynamic matrix A and the corresponding submatrices C and M. We call this as the nested property or orthogonality property for the reason we explain soon. Another property is that these lower-dimensional models are all asymptotically stable if the original matrix A is asymptotically stable. (See PERNEBO and SILVERMAN [1982] for proof). To see that the nested or orthogonality property holds, partition the factors in the singular value decomposition into $(U_1\,U_2)$, $(\Sigma_1\,\Sigma_2)$, and $(V_1\,V_2)$, where Σ_1 contains the first n^* significant singular values, where n^* is less than the effective rank of the Hankel matrix. Take the matrix C as an example. The above partitioned singular value decomposition produces the corresponding partition of the matrix C as $(C_1\,C_2)$, where C_i is given by $H^{C}\,V_i\,\Sigma_i^{-1/2}$, $i = 1, 2$. This consideration shows that system matrices for the state vector of dimension less than the effective rank of the Hankel matrix are embedded in the larger original system matrices. In other words, once a model of state vector dimension 7 is constructed, system matrices of models of dimension one through 6 can be read off from those of the corresponding system matrices of the model of dimension 7 (except for the matrix B which require separate solutions of the corresponding Riccati equation as we show later). Furthermore, if a model of dimension 9 is desired, then the 8th and 9th singular values and the corresponding singular vectors are used to compute C_2 as $H^{c}\,V_2\,\Sigma_2^{-1/2}$ and appended to the original C_1 to form the system matrix C as $(C_1\,C_2)$, i. e., as $(C_1\,0) + (0\,C_2)$, and similarly for M and A. In this sense, the balanced representation has the important "orthogonality" of representation like the coefficients of expansion of a system of orthogonal polynomials in approximating a continous function, say.

Determination of the matrices B and Δ requires solving a certain matrix Riccati equation. From (1) note the relations

$$\Pi = A\,\Pi\,A' + B\,\Delta B', \qquad \text{and} \qquad \Lambda_0 = C\,\Pi\,C' + \Delta.$$

From the definition of the matrix M we have

$$B\Delta = M - A\Pi C'.$$

From these we obtain the equation for Π as

$$\Pi = A\Pi A' + (M - A\Pi C')(\Lambda_0 - C\Pi C')^{-1}(M - A\Pi C')'. \tag{11}$$

The smallest positive definite solution of (11), which exists under certain technical conditions, is used to determine B and Δ. Equation (11) can be solved either recursively or by converting it into a certain symplectic matrix and putting it into Schur decomposition form. See AOKI [1987, Chapter 7] for detail.

Before we finish this brief exposition of the proposed method, we touch on the idea of approximating a given model by a smaller dimensional one by retaining a subset of the singular values used in the original model. This idea can be expressed as the approximation problem of a rational transfer function matrix. The transfer matrix of the original model to be approximated by that of lower-dimensional model is of the form $C(zI - A)^{-1}B$ (the constant matrix I plays no role in the approximation) where z may be thought of as the inverse of the lag operator. By using the Hankel norm, which is defined to be the largest singular value of the associated Hankel matrix, one can justify the use of the ratio of the $(n+1)$-th singular value over the largest and the ratio of the sum of the neglected singular values over the largest singular value as bounds of the approximation by n-dimensional models. See AOKI [1987] for further detail on error and sensitivity analysis. GLOVER [1984] develops the transfer function approximation for continuous time dynamic systems using the Hankel norm.

To summarize, there are two choice variables; size of the Hankel matrix is determined by K, and the size of models is determined by the state vector dimension. Larger K values mean that more, but increasingly noisier, sample covariance matrix information is used in estimating system parameters. Larger dimensions mean better in sample fit but tendency for overfitting. Models must balance these two opposing tendencies. Reductions of the sums of squares of residuals by the model and either the trace or the determinants of the one-step-ahead error covariance matrix vis-à-vis Λ_0 may be used as measures of goodness of in- and out-of- sample fits of the constructed models. These figures need be examined as functions of both K and model dimension in finally selecting a model.

Kalman Filters

This section establishes that balanced models constructed by the method described above are Kalman filters driven by innovation sequences. In the representation (1), note that y_t^+ is related to x_t by

$$y_t^+ = Oz_t + K\mathbf{e}_t^+,$$

where K is some matrix of no concern here. Then, from this relation, observe that the orthogonal projection is

$$\hat{E}(y_t^+ \mid y_{t-1}^-) = O\hat{E}(z_t \mid y_{t-1}^-) \tag{12}$$

which is used to define a vector $x(t)$ by equating the right hand side to $Ox(t)$, i. e.,

$$x(t)=\hat{E}(z_t \mid y_{t-1}^-). \tag{13}$$

Noting that the left-hand side of (12) is equal to $HR^{-1}y_{t-1}^-$ and from (3), $x(t)$ can be put as

$$x(t)=S_H y_{t-1}^- \tag{14}$$

where

$$S_H=\Omega R^{-1}$$

is the matrix used in summarizing the past as the state vector when O is full rank. [7] Later, it is shown that $x(t)$ also evolves over time according to (1), and that e_t is the innovation sequence.

Next, the state equation for the vector $x(t)$ is derived that governs the time evolution of $x(t)$ following FAURRE *et al.*, [1979]. Advancing t by one unit in (14), write

$$x(t+1)=\Omega_{t+1}R_t^{-1}y_t^- \quad . \tag{15}$$

where subscripts are now introduced to be explicit about timing. Note that the matrices Ω_{t+1} and R_t can be written as

$$\Omega_{t+1}=[M,\ A\Omega_t], \tag{16}$$

$$R_t=\begin{bmatrix} \Lambda_0 & C\Omega_t \\ \Omega_t' C & R_{t-1} \end{bmatrix}. \tag{17}$$

Partition the stacked data vector also conformably as

$$y_t^-=\begin{bmatrix} y_t \\ y_{t-1}^- \end{bmatrix}. \tag{18}$$

The inverse matrix R_t^{-1} can be written as

$$R_t^{-1}=\begin{bmatrix} R^{11} & -R^{12} \\ -R^{21} & R^{22} \end{bmatrix}, \tag{19}$$

where

$$R^{11}=\Lambda_0-CP_tC', \qquad R^{12}=R^{11}C\Omega_t R_{t-1}^{-1}, \qquad R^{21}=(R^{12})'$$

and

$$R^{22}=R_{t-1}^{-1}+R_{t-1}^{-1}\Omega_t C' R^{11} C\Omega_t R_{t-1}^{-1}.$$

To simplify the expressions introduce an auxiliary matrix by

$$P_t=\Omega_t R_{t-1}^{-1}\Omega_t'.$$

7. An alternative summary method using canonical vectors is used by Akaike (who maximizes the mutual information in the sense of GEL'FAND and YAGLOM [1959]), and another by ARUN *et al.* [1983], (who use the predictive error of RAO [1964]). See Chapter 8 of AOKI [1987] for a unified account of these alternatives.

From (13) and (14), P_t is seen to be the covariance matrix of $x(t)$. Grouping terms in (19) appropriately, (15) is rewritten as

$$x(t+1)=A\,x(t)+\Gamma_t v_t, \tag{20}$$

where

$$v_t=y_t-C\,x(t)$$

is the innovation vector and

$$\Gamma_t=(M-AP_tC')(\Lambda_0-CP_tC')^{-1}.$$

The matrix P_t evolves with time according to

$$\begin{aligned} P_{t+1}&=\operatorname{cov} x(t+1)\\ &=(M,\ A\,\Omega_t)\,R_{t-1}^{-1}(M,\ A\,\Omega_t)'.\\ &=AP_tA'+(M-AP_tC')(\Lambda_0-CP_tC')^{-1}(M-AP_tC')'. \end{aligned} \tag{21}$$

From the last part of Section 2, we recognize the correspondence between v and e, Γ and B, and P and Π when the steady state values exist. The steady state covariance satisfies the algebraic Riccati equation in which P_t and P_{t+1} are replaced by Π^*. This Π^* van be shown to be positive definite and is the unique minimal element of the set of symmetric solutions of the Riccati equation when (1) is reachable and observable, and if the system is regular, i. e. $\Lambda_0-C\Pi^*C'>0$, see FAURRE *et al.* [1979].

The choice of the initial condition of the Kalman filter is not crucial when 100 or so data points are available because A^{100} is very small in general. One can refine the initial segments of in-sample fits by estimating the initial condition of the filter by a smoothing method over some initial subset of data point, if desired.

3 Unit Roots

Economic time series are usually nonstationary. Some forms of pre-processing of data, such as taking the first difference or logarithms, is customarily performed to produce weakly stationary series to which the method described above or some other methods are applied to construct ARMA(X), VAR or state space models. One of the important questions is the presence of unit roots or the presence of random walk components in the time series. The method of this paper produces an asymptotically stable models if the data series is weakly stationary. This fact follows from the way of solving the Riccati equation for the state vector covariance matrix Π. The matrix Π is obtained by solving the associated symplectic matrix which has eigenvalues λ, $\bar{\lambda}$ and their inverses. Thus, unless $|\lambda|=1$, exactly n eigenvalues has magnitude less than 1 and the remaining n eigenvalues have magnitude greater than 1 where n is the dimension of the

state vector. The asymptotically stable matrix A can then be constructed by selecting only those eigenvalues with magnitude less than 1 and the corresponding eigenvectors, see AOKI [1987]. The eigenvalues do not lie on the unit circle if and only if the innovation model is observable and reachable as shown in AOKI [1987].

Thus unit roots are present if and only if the dynamic mode(s) corresponding to the eigenvalue $|\lambda|=1$ is either not observable or not reachable. (Such roots are not detected in the transfer functions.) State space models may fail to have these properties if the residuals are not weakly stationary, if (random) detrending of original series fail to remove significant portion of random walk components, for example. This section shows an explicit construction which separates out unstable or unit roots by deflating the dynamic matrix of a state space model.

If one does not wish to decompose time series into deterministic trends plus disturbances but rather regard the time series to contain random growth components, one may wish to model the time series directly without preliminary processing such as differencing. Suppose that one considers a state space model of a vector-valued time series where its dynamic matrix has one or more unit roots or eigenvalues of magnitude greater than 1. A suitable redefinition of state vectors yields a lower-dimensional state space model. AOKI [1971] provides some link between deflation and aggregation of dynamic systems. See HOUSEHOLDER [1964, p. 168] on deflation. The process of deflating the dynamic matrix by removing unit eigenvalues is related to aggregation of the original state vector into a lower-dimensional one. The notion of co-integration introduced by GRANGER [1983] can also be regarded as a way of deflating the dynamic matrix.

The relation with aggregation of dynamic systems, as discussed in AOKI [1971], is easily established. Suppose that a dynamic matrix A has at least one unit eigenvalue, and define $F=I-A$. Because F is not full rank, it has a rank factorization representation

$$F=-HK$$

where H is n by k and K is k by n, both of rank k. Multiply the state equation from the left by K to obtain

$$\chi_{t+1}=(I+KH)\chi_t+KB e_t$$

where a new aggregated state vector is introduced by $\chi_t=K x_t$. Noting the matrix identity

$$\mathrm{Det}(\lambda I-A)=\det(\mu I-HK)=\mu^{n-k}\det(\mu I-KH)$$

where $\mu=\lambda-1$, the matrix $I+KH$ no longer has the unit roots present in the original matrix A. The expression for the new state vector is a general version of co-integration.

More generality, let the matrix P, consisting of k independent column vectors, be such that

$$AP=PM,$$

i. e., the column vectors form a basis for an invariant subspace of A where

M is the matrix with eigenvalues of unit magnitude or greater. There exists another matrix S of rank $n-k$ such that $S'A=NS'$ corresponding to eigenvalues with non-unit magnitudes, $|\lambda|<1$. Moreover $S'P=0$ holds, since the invariant subspaces belong to distinct eigenvalues of A. See HOUSEHOLDER [1964, p. 169]. Represent x_t as $S\zeta_t+P\eta_t$. Substitute this into the state equation and multiply it from left by S′ to produce

$$S'S\zeta_{t+1}=NS'S\zeta_t+S'Be_t$$

where the first term on the right, S′AS, is seen to be equal to NS′S. The vector $S'S\zeta_t$ is governed by the state equation with the dynamic matrix N. Define a vector

$$z_t=S'S\zeta_t.$$

It is governed by

$$z_{t+1}=Nz_t+S'Be_t. \tag{22}$$

Multiplication from the left by P′ instead of S's of the state equations yields

$$\eta_{t+1}=M\eta_t+(P'P)^{-1}AS(S'S)^{-1}z_t+(P'P)^{-1}P'Be_t. \tag{23}$$

This transformation decomposes the state vector x_t into two parts: (i) a part z_t which is stationary, and is governed by (22); and (ii) a part which is nonstationary which also depends on z_t as seen from (23). The data vector is then expressible as

$$y_t=CS(S'S)^{-1}z_t+CP\eta_t+e_t.$$

This is a general representation of the time series containing random walk or explosive components, unless CP vanishes. More detailed discussion of the implications of unit root on the transfer matrix (i. e., ARMA) representation are found in AOKI [1987*a*].

4 Examples: US-Japan Interactions

This section describes some examples of single and two-country models mostly for the industrial production indices and *ex ante* real interest rates computed by subtracting one-month ahead expected consumer price index inflation rates (calculated as a single series) from nominal interest rates which is 4-6 month commercial paper rates in the US and the call rates in Japan. The letters y and y^* refer to the industrial production indices and r and r^* to the real interest rates in Japan and US respectively. In some examples the series of monthly data for the US fiscal expenditure deflated by the US consumer price index is also used. The letter g is used to refer to the residual series of this data. To be explicit about the two choice variables mentioned in Section 2, a model constructed using K and dimension n is denoted by $m(K, n)$.

4.1. Single Country Examples

To illustrate the use of state space model to infer dynamic interactions, we modify Kitagawa's TIMSAC program to detrend time series and fit the residuals by state space models. [8] (The residuals are quite different from those from taking the first differences of the logarithms. Also, different residuals are extracted depending on the first difference or the second difference of the logarithms of IPI's are regarded as being driven by exogenous stationary noises. This paper uses the quadratic detrending.)

Example 1: Model m (5, 3) for (y, r).

This bivariate series provide better models for y than the univariate series consisting of IPI alone, because the one-step-ahead forecast error variances is reduced further by including $\{r_t\}$: Model m (5, 2) has the one-step-ahead forecast error variances of $\Delta_y = .92$ and $\Delta_r = 4.52$. m (5, 3) has the one-step-ahead prediction error variance $\Delta_y = .89$ and $\Delta_r = 3.58$. These improve slightly by increasing the model dimension to 4 as m (4, 4) processes $\Delta_y = .85$ and $\Delta_r = 3.54$. Here we use m (5, 3) for simpler illustration.

In m (5, 3), the system matrix A is nearly block upper triangular

$$A = \begin{bmatrix} .726 & -.029 & -.036 \\ .071 & .614 & -.749 \\ .165 & .644 & .193 \end{bmatrix}$$

$$\approx \begin{bmatrix} .726 & 0 & 0 \\ x & & A_2 \\ x & & \end{bmatrix}$$

i. e., the dynamics are approximately recursive and the matrix C also has a special structure

$$C = \begin{bmatrix} 1.418 & -.321 & .122 \\ -.044 & -.898 & -1.174 \end{bmatrix} \approx \begin{bmatrix} 1.418 & -.321 & .122 \\ 0 & -.898 & -1.174 \end{bmatrix}.$$

The first component of the state vector affects only y. The covariance matrix of the state vector

$$\Pi = \begin{bmatrix} .754 & & \\ -.117 & .646 & \\ .047 & .003 & .525 \end{bmatrix}$$

shows that all three state variables are uncorrelated. Even without further analysis, the matrix C and Π tells us the following: The contribution of x_3 to y is less than 10% of x_1, indicating why a two-dimensional model m (5, 2) does reasonably well in reducing a one-step-ahead forecast errors of IPI. This is confirmed by var $y = .92$ of m (5,2) is reduced only to var $y = .89$ by m (5. 3). The contribution of x_3, however, is essential in reducing the one-step-ahead forecast errors in the *ex ante* real interest rate because x_1 contributes little to r. (The element $c_{21} = -.04$ is much smaller than

8. Kitagawa's program fits AR models to residuals (see AOKI [1987*a*] for an alternative method).

c_{22} or c_{23} in magnitude.) This is confirmed by noting that the one-step-ahead error variance var $r=3.58$ of m (5, 3) deteriorate to var $r=4.52$ when a two-dimensional model m (5, 2) is used. Furthermore, the dynamics for the first state vector component is not affected by the other two variables.

The approximate dynamics are characterized by these two features. These features indicate that y can be decomposed into two components

$$y_t = y_{1t} + y_{2t}$$

where

$$y_{1t} = 1.418\, x_{1t}$$

and

$$y_{2t} = -.321\, x_{2t} + .122\, x_{3t} + \varepsilon_{y_t},$$

where

$$x_{1t+1} = .726\, x_{1t} + .573\, \varepsilon_{y_t} + .093\, \varepsilon_{r_t}.$$

because the matrix B is

$$\mathrm{B} = \begin{bmatrix} .573 & .093 \\ -.289 & .144 \\ .267 & -.245 \end{bmatrix}.$$

Thus the component y_{1t} evolves with time as

$$\begin{aligned} y_{1t+1} &= 1.42\, x_{1t+1} \\ &= .73(1.42\, y_{1t}) + 1.42(.57\, \varepsilon_{y_t} + .10\, \varepsilon_{r_t}) \\ &= .73\, y_{1t} + .81\, \varepsilon_{y_t} + .13\, \varepsilon_{r_t}, \end{aligned}$$

while y_{2t} is governed jointly with r_t by

$$\begin{bmatrix} y_{2t} \\ r_t \end{bmatrix} = \begin{bmatrix} -.32 & .12 \\ -.90 & -1.17 \end{bmatrix} \begin{bmatrix} x_{2t} \\ x_{3t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y_t} \\ \varepsilon_{r_t} \end{bmatrix},$$

where

$$\begin{bmatrix} x_2 \\ x_3 \end{bmatrix}_{t+1} \approx \mathrm{A}_2 \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}_t + \begin{bmatrix} 0 \\ .165 \end{bmatrix} x_{1t} + \begin{bmatrix} -.29 & .14 \\ -.27 & -.24 \end{bmatrix} \begin{bmatrix} \varepsilon_{y_t} \\ \varepsilon_{r_t} \end{bmatrix}.$$

Because $\{y_t\}$ has the variance 2.60 and

$$\sigma_{y_1}^2 = 1.42^2\, \sigma_{x_1}^2 = 1.516,$$

we conclude that about 58.3% of the variation associated with the Japanese IPI is autonomous. The component y_1 thus represents the portion of japanese IPI which is unresponsive to changes in domestic real interest rate, such as export components. It is interesting to observe no such structural feature is found in the US data.

In the ARMA representation, the dynamic structure is approximately given by a three-dimensional ARMA (1, 1) (obtained by eliminating x_2 and x_3 from the above, see Example 2 for the detailed procedure.)

$$y_{1t+1} = .726\, y_{1t} + .81\, \varepsilon_{y_t} + .13\, \varepsilon_{r_t}$$

$$y_{2t+1} = .775\, y_{2t} - .145\, r_t + .014 y_{1t} - .076\, \varepsilon_{r_t} + \varepsilon_{y_{t+1}}$$

$$r_{t+1} = 3.976\, y_{2t} + .032\, r_t - .22\, y_{1t} - 4.03\, \varepsilon_{y_t} + .127\, \varepsilon_{r_t} + \varepsilon_{r_{t+1}}$$

Note that r_{t+1} is much more influenced by the non-autonomous part of y_t than the autonomous one. The autonomous component of y decays a little faster than the non-autonomous component. For example $(.726)^6 = .146$ and $(.775)^6 = .217$.

Example 2: Model m (7, 3) for (y^*, g).

Model m (7, 2) has the one-step-ahead error variances of var $y^* = .684$ and var $g = 4.961$, while m (7,3) improves the latter to var $y^* = .688$, even though var $y^* = .688$ is slightly worse. [y^* modeled as a univariate series produces var $y^* = .71$ by m (7, 2).]

The system matrix C of m (7, 3)

$$C = \begin{bmatrix} 1.430 & 1.022 & -.0209 \\ -.107 & -.104 & -1.324 \end{bmatrix}$$

shows that x_1 and x_2 primarily explains the movements in y and x_3 contributes most of movements in g. The error covariance matrix

$$\Pi = \begin{bmatrix} 1.14 & & \\ .02 & .55 & \\ -.003 & .05 & .36 \end{bmatrix}$$

shows that x_3 is uncorrelated with x_1 and only slightly correlated with x_2, since $.05/\sqrt{.55 \times .36} = .11$. Thus, movements in g is nearly uncorrelated with those of y.

Reducing the model dimension to two results in a greater increase of the forecasting error variance of g for this reason. Another way to state this is to compare the in-sample fit of m (7, 3) and m (7, 2). Reduction of the sum of the squares of the residuals of g by the model is 26.6% by m (7, 3) but only 0.3% by m (7,2), while 85.9% and 84.7% of the sums of the squares of the residuals of y^* are reduced by m (7, 3) and m (7, 2), i.e., as far as the y^*-series is concerned m (7, 2) is almost as good as m (7, 3).

The ARMA model of m (7,2) is

$$\begin{cases} y^*_{t+1} = 2.41\, y^*_t + 20.76\, g_t - 1.07\, \varepsilon_{y_t} - 20.7\, \varepsilon_{g_t} + \varepsilon_{y_{t+1}} \\ g_{t+1} = -.12\, y^*_t - .7\, g_t + .01\, \varepsilon_{y_t} + .71\, \varepsilon_{g_t} + \varepsilon_{g_{t+1}}. \end{cases}$$

A very large coefficient on g_t is an indication of poor representation for g-series by this model.

To deduce the corresponding ARMA for $m(7,3)$, let

$$\mathrm{H}=\begin{bmatrix} \mathrm{C} \\ c_1\,\mathrm{A} \end{bmatrix}.$$

Then

$$\begin{bmatrix} y_t^* \\ g_t \\ y_{t+1}^* \end{bmatrix}=\mathrm{H}\,x_t+\begin{bmatrix} e_t \\ c_1\,\mathrm{B}\,e_t \end{bmatrix}+\begin{bmatrix} 0 \\ e_{t+1} \end{bmatrix}$$

where c_1 is the first row of the matrix C.

Eliminate x_t by inverting H (the row vectors of C and c_1 A are linearly independent. The existence of one additional linearly independent row vector is guaranteed by the observability of the model), we obtain

$$\begin{bmatrix} y_{t+1}^* \\ g_{t+1} \\ y_{t+2}^* \end{bmatrix}=\mathrm{HAH}^{-1}\begin{bmatrix} y_t^* \\ g_t \\ y_{t+1}^* \end{bmatrix}+\begin{bmatrix} e_{t+1} \\ \varepsilon_{y_{t+2}} \end{bmatrix}+\begin{bmatrix} 0 \\ e_{t+2} \end{bmatrix}$$
$$+\mathrm{HB}\,e_t-\mathrm{HAH}^{-1}\begin{bmatrix} \mathrm{I} \\ c_1\,\mathrm{B} \end{bmatrix}e_t$$
$$-\mathrm{HAH}^{-1}\begin{bmatrix} 0 \\ e_{t+1} \end{bmatrix}.$$

Noting that

$$[\mathrm{HAH}^{-1}]=\begin{bmatrix} 0 & 0 & 1 \\ & \varphi_1 & \varphi_2 \end{bmatrix}$$

where φ_1 is 2×2. we see that the first row of this relation is an identity; $y_{t+1}^*=y_{t+1}^*$.

The second and the third row now yield

$$g_{t+1}=-.238\,y_t^*-.375\,g_t+.223\,y_{t+1}^*+\text{noise}$$

and

$$y_{t+2}^*=-.789\,y_t^*-.0806\,g_t+1.73\,y_{t+1}^*+\text{noise}.$$

Replacing $t+1$ by t from this second equation, and eliminating y_{t+1}^* from the equation for g_{t+1}, we obtain an ARMA (2, 2) model :

$$g_{t+1}=.146\,y_t^*-.546\,g_t-.02_{t-1}+\text{noise},$$
$$y_{t+1}^*=1.73\,y_t^*-.789\,y_{t-1}^*-.081\,g_{t-1}+\text{noise}.$$

(The detail on the noise is omitted.)

4.2. Two-Country Examples

Industrial Production Index

The *a-d* framework is used by defining $y_a = y + y^*$ and $y_d = y - y^*$.[9] Here y and y^* are 10^2 times residuals of the IPI after detrended by TIMSAC program.

Model $m(5, 3)$ reduces

$$\Lambda_0 = \begin{bmatrix} 5.913 & \\ -1.047 & 6.586 \end{bmatrix}$$

to

$$\Delta_{5,3} = \begin{bmatrix} 1.799 & \\ .369 & 1.617 \end{bmatrix}.$$

In terms of the original variables, one-step-ahead forecast error variances of y and y^* are 1.04 and .67 respectively.

Augment the matrix C by c_2 A to obtain

$$\begin{bmatrix} y_t^d \\ y_t^a \\ y_{t+1}^a \end{bmatrix} = \begin{bmatrix} \mathrm{C} \\ c_2 \mathrm{A} \end{bmatrix} x_t + \text{noise}$$

Proceeding as in the previous example, one obtains a two-dimensional ARMA (2, 2) model representation

$$y_{t+1} = .98 y_t - .847 y_t^* - .672 y_{t-1} + .47 y_{t-1}^* + \text{noise}.$$

and

$$y_{t+1}^* = -.43 y_t + 1.39 y_t + .54 y_{t-1} - .50 y_{t-1}^* + \text{noise}.$$

Table 1 lists the responses of the Japanese and the US IPI to a common shock and country specific impulse disturbances by model $m(4,3)$. Table 1′ does the same for $m(5,3)$. Three noteworthy features are observed in both tables : (i) In response to a common shock, the initial response in the US is larger than that in Japan, but dissipates more quickly in the US turning negative after five months, while a positive effects persist much longer in

9. This framework has certain theoretical advantages in modeling two countries with similar macroeconomic characteristic as shown by AOKI [1981]. Numerical experiences favor this framework since models in the $a-c$ framework tends to produce models with smaller one-step-ahead error variances, even though Japan and the USA do not have similar macroeconomic characteristics, when a series not in the subspace spanned by y and y^*, such as the fiscal expenditure, is included in the data series.

Japan. (ii) A positive US shock also has expansionary effects in Japan. The initial expansionary effects on the US economy turn contractionary after 6 months in the US and last for the next 10 months or so, while the extansionary effects in Japan also lasts about 10 months. (iii) A positive Japanese shock produces contradictionary effects in the US after a positive initial impact effects.

TABLE 1

IPI Dynamic Multipliers of m(4, 3)

Japan Shock		Common Shock		US Shock	
Japan	USA	Japan	USA	Japan	USA
.584	.180	.933	1.519	.349	1.339
.483	−.045	.793	.910	.310	.954
.393	−.189	.704	.444	.311	.633
.307	−.269	.632	.095	.325	.363
.225	−.297	.559	−.154	.334	.143
.147	−.287	.476	−.318	.329	−.025
.079	−.250	.388	−.478	.309	−.157

Very similar responses are obtained by $m(5, 3)$ shown in Table 1′ even though the model looks quite different.

TABLE 1′

IPI Dynamic Multipliers of m(5, 3)

Japan Shock		Common Shock		US Shock	
Japan	US	Japan	US	Japan	US
.577	.179	.929	1.488	.352	1.309
.477	−.042	.781	.897	.304	.939
.387	−.184	.690	.441	.304	.625
.304	−.265	.624	−.094	.304	.358
.225	−.203	.558	−.158	.334	.137
.150	−.287	.483	−.326	.330	−.039
.082	−.252	.398	−.421	.315	−.169

Ex Ante *Real Interest Rate Interactions*

A four-dimensional model, $m(4, 4)$ has been selected. The model reduces the data covariance matrix

$$\Lambda_0 = \begin{bmatrix} 8.426 & \\ 2.138 & 6.599 \end{bmatrix}$$

to

$$\Delta_{4,4} = \begin{bmatrix} 6.180 & \\ 2.202 & 4.065 \end{bmatrix}.$$

In the original variables, the one-step-ahead forecast error variances are var $r = 3.66$ and var $r^* = 146$. From this four-dimensional model, the state vector can be eliminated as follows : stack d_t and d_{t+1} together as

$$\begin{bmatrix} d_t \\ d_{t+1} \end{bmatrix} = \begin{bmatrix} \mathrm{C} \\ \mathrm{CA} \end{bmatrix} x_t + \begin{bmatrix} \mathrm{I} \\ \mathrm{CB} \end{bmatrix} e_t + \begin{bmatrix} 0 \\ e_{t+1} \end{bmatrix}, \qquad \text{where} \quad d_t = \begin{bmatrix} r_t^a \\ r_t^d \end{bmatrix}.$$

This relation can be solved for x_t since $\begin{bmatrix} \mathrm{C} \\ \mathrm{CA} \end{bmatrix}$ is a 4×4 nonsingular matrix. (The nonsingularity is guaranteed by the model observability)

$$x_t = \begin{bmatrix} \mathrm{C} \\ \mathrm{CA} \end{bmatrix}^{-1} \left[\begin{pmatrix} d_t \\ d_{t+1} \end{pmatrix} - \begin{pmatrix} \mathrm{I} \\ \mathrm{CB} \end{pmatrix} e_t - \begin{pmatrix} 0 \\ e_{t+1} \end{pmatrix} \right].$$

Since x_t is governed by the dynamics

$$x_{t+1} = \mathrm{A}\, x_t + \mathrm{B}\, e_t$$

and

$$\begin{aligned} d_{t+1} &= \mathrm{C}\, x_{t+1} + e_{t+1} \\ &= \mathrm{CA}\, x_t + \mathrm{CB}\, e_t + e_{t+1} \end{aligned}$$

we obtain

$$\begin{pmatrix} d_{t+1} \\ d_{t+2} \end{pmatrix} = \begin{pmatrix} \mathrm{C} \\ \mathrm{CA} \end{pmatrix} \mathrm{A} \begin{pmatrix} \mathrm{C} \\ \mathrm{CA} \end{pmatrix}^{-1} \left[\begin{pmatrix} d_t \\ d_{t+1} \end{pmatrix} - \begin{pmatrix} \mathrm{I} \\ \mathrm{CB} \end{pmatrix} e_t - \begin{pmatrix} 0 \\ e_{t+1} \end{pmatrix} \right] + \begin{pmatrix} \mathrm{C} \\ \mathrm{CA} \end{pmatrix} \mathrm{B}\, e_t + \begin{pmatrix} \mathrm{I} \\ \mathrm{CB} \end{pmatrix} e_{t+1} + \begin{pmatrix} 0 \\ e_{t+2} \end{pmatrix}.$$

Noting the relation

$$\begin{pmatrix} \mathrm{C} \\ \mathrm{CA} \end{pmatrix} \mathrm{A} \begin{pmatrix} \mathrm{C} \\ \mathrm{CA} \end{pmatrix}^{-1} = \begin{pmatrix} 0 & \mathrm{I} \\ \varphi_2 & \varphi_1 \end{pmatrix},$$

we obtain

$$\begin{aligned} d_{t+2} = {} & \varphi_1\, d_{t+1} + \varphi_2\, d_t + (\mathrm{CAB} - \varphi_1\, \mathrm{CB} - \varphi_2)\, e_t \\ & + (\mathrm{CB} - \mathrm{I})\, e_{t+1} + e_{t+2} \end{aligned}$$

which is a two-dimensional vector ARMA (2, 2).

The system matrices of $m(4, 4)$ produce

$$\varphi_1 = \begin{bmatrix} -.670 & .849 \\ -.385 & .835 \end{bmatrix} \quad \text{and} \quad \varphi_2 = \begin{bmatrix} 1.484 & -1.389 \\ 1.000 & -.428 \end{bmatrix},$$

i. e.,

$$r_{t+2} = .33\, r_{t+1} + .33\, r_t - 2.15\, r_{t+1}^* + 1.37\, r_t^* + u_{1t+2},$$

and

$$r_{t+2}^* = .72\, r_{t+1}^* - .15\, r_t^* + .24\, r_{t+1} + .13\, r_t + u_{2t+2},$$

where

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_{t+2} = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}_{t+2} + \begin{pmatrix} -.069 & 3.097 \\ -.348 & .018 \end{pmatrix} e_{t+1} + \begin{bmatrix} -.1.445 & -.074 \\ -.110 & .075 \end{bmatrix} e_t.$$

Here the covariance matrix of the innovation is the same as $\Delta_{4,4}$ given above.

The models' dynamic multipliers for the first 7 months are listed in Table 2.

A common shock produces more positive *ex ante* real interest rates in the US than in Japan. The response to country specific shocks are also clearly discernible from Table 2. A Japan shock lowers ex ante real interest rates in the US by small amount. A shock in the US tends to raise the real interest rates in both countries.

TABLE 2

Ex Ante *Real Interest Rate Multipliers*

Japan Shock		Common Shock		US Shock	
Japan	USA	Japan	USA	Japan	USA
.247	−.145	.522	.379	.275	.525
−.310	.015	−.233	.477	.077	.462
−.257	−.008	−.421	.304	−.163	.311
−.146	−.013	−.212	.212	−.066	.225
.101	−.149	.181	−.142	.081	.007
.155	−.087	.320	−.091	.164	−.004
.065	.010	.164	.056	.099	.046

Ex Ante *Real Interest Rates and Industrial Production Index*

Model $m(5, 3)$ is constructed. It's forecasting performance is

$$\Delta_r = 3.32, \quad \Delta_{r*} = 1.07 \quad \Delta_y = .911 \quad \text{and} \quad \Delta_{y*} = .73,$$

which is better than $\Delta_r = 3.66$ and $\Delta_{r*} = 1.04$ of $m(4, 4)$ from the bivariate series (r, r^*), and better than $\Delta_y = 1.04$ of $m(4, 3)$ from the bivariate series (y, y^*) but worse than $\Delta_{y*} = .65$ of the same model $m(4, 3)$. Model $m(3, 5)$ for the four-dimensional series (r, r^*, y, y^*) does better jobs of forecasting r, r^*, and y but worse for y^*.

Since the system matrix C is 4 by 5, we augment it by c_1A where c_1 is the first row of C to construct a nonsingular matrix H

$$\mathrm{H} = \begin{pmatrix} \mathrm{C} \\ c_1 \mathrm{A} \end{pmatrix},$$

where

$$\begin{pmatrix} d_t \\ r_{t+1} \end{pmatrix} = \mathrm{H} x_t + \begin{pmatrix} e_t \\ \varepsilon_{t+1} \end{pmatrix}.$$

FIGURE 1

IPI Responses to a r Shock in US

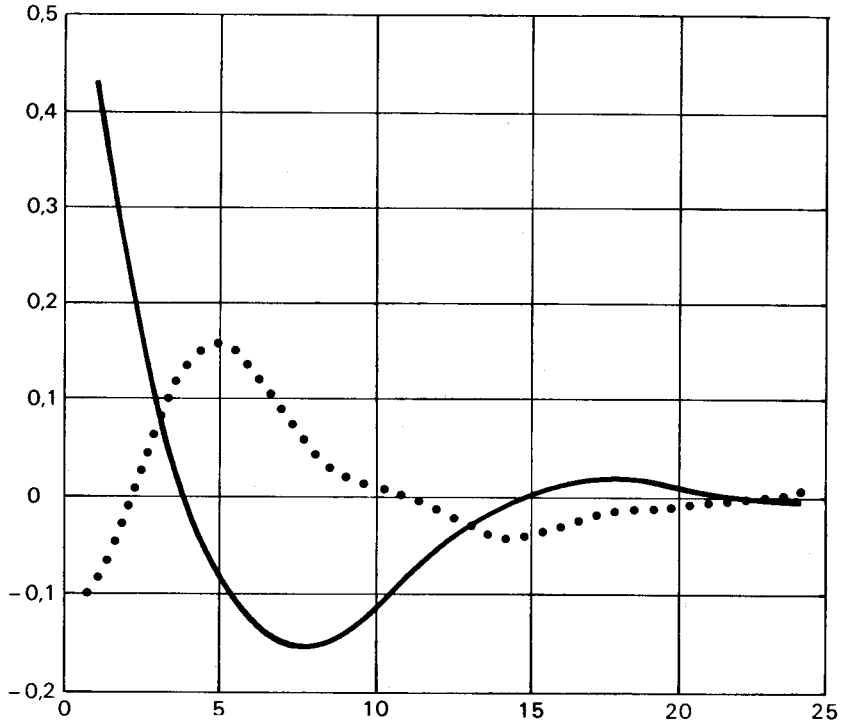
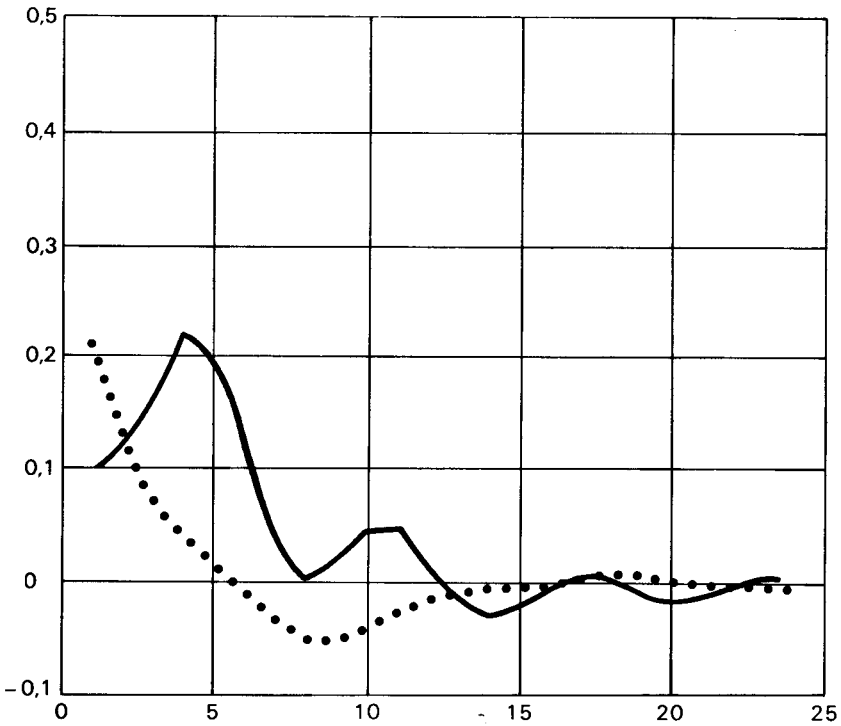


FIGURE 2

IPI Responses to a r Shock in Japan



From this and the state transition equation

$$x_{t+1} = \mathrm{A}\, x_t + \mathrm{B}\, e_t$$

$$\begin{pmatrix} d_{t+1} \\ r_{t+1} \end{pmatrix} = \mathrm{HAH}^{-1} \begin{pmatrix} d_t \\ r_{t+1} \end{pmatrix} + \mathrm{HB}\, e_t - \mathrm{HAH}^{-1} \begin{pmatrix} e_t \\ \varepsilon^r_{t+1} \end{pmatrix} + \begin{pmatrix} e_{t+1} \\ e^r_{t+2} \end{pmatrix},$$

or the four-dimensional ARMA (1, 1) model results :

$$\begin{aligned} y_{t+1} &= .707\, y_t + .026\, y^*_t - .266\, r_t + .131\, r^*_t + \text{noise} \\ y^*_{t+1} &= -.419\, y_t + .762\, y^*_t - .091\, r_t - .045\, r^*_t + \text{noise} \\ r_{t+1} &= -.862\, r_t + .962\, y^*_t - .203\, y_t + .132\, y^*_t + \text{noise} \\ r^*_{t+1} &= .158\, r_t + .874\, r^*_t + .034\, y_t + .219\, y^*_t + \text{noise}. \end{aligned}$$

Figure 1 and 2 plot the IPI responses to a US r shock and to a Japanese r shock. Solid lines show own country IPI responses, and broken lines are for the other country responses. The US IPI response to a common r shock is similar in shape with slightly smaller magnitude to a US r shock. The Japanese IPI responds to a common r shock as to a US r shock except for the absence of initial negative effects. Figure 1 shows that the effects on the US IPI of a real interest rate shock in the US quickly reverse the sign and remain negative for the next 10 months. The initial negative effects on the Japanese IPI become positive after a few months and remain so for the next 10 months. Figure 2 shows that a r shock in Japan is expansionary in both countries for the first five months or so, after which the US IPI responses turn slightly negative.

5 Conclusion

The paper shows how some information on dynamic structure is easily obtained by state space models than other (VAR or ARMA) approaches. Examples of US-Japan interactions of the industrial production and *ex ante* real interest rates show that these two economies respond quite differently to a common shock and their spill-over dynamic effects are also quite different. The real interest rate tends to be more negative in Japan than in the US to any type of real interest rate shocks except at impacts of common shocks and shocks specific to Japan. Real interest rates remain positive in the US in response to a positive shock to its own *ex ante* real interest rate. Real interest rates become negative in both countries to Japanese shock after one month. Responses to shocks to the industrial production index tend to be more magnified in the US than in Japan. The US IPI response tends to become negative sooner than in Japan in response to a common shock and in response to a Japanese shock. A US shock has an expansionary effects to Japan, on the other hand. Thus, during the period covering the second quarter 1975 to the end of 1981, an expansionary move in Japan would have reduced the industrial production index in the USA and a move to reduce the ex ante real interest rate in Japan would have raised the US real interest rate after one quarter or so.

Joint modeling of the IPI and the *ex ante* real interest rate does not alter the qualitative feature of the dynamic multiplier time profile although the magnitudes of the initial responses are altered, sometimes by 25 % or so.

Inclusion of the US fiscal expenditure in the data series magnifies the initial portion of the dynamic multiplier time profiles, i. e., the magnitude of the spill-over effects to country specific shocks and the responses to common shocks both become larger. Details are reported elsewhere.

Finally, the method used in this paper is useful because it reveals the presence of unit roots when the residual series are not weakly stationary, thus avoiding the danger of unwittingly building models when such models do not exist by the nature of the residuals.

APPENDIX 1

Data and Detrending

Japan

The logarithms of the monthly Japanese industrial production index (IPI) has been decomposed into a random quadratic trend with the var $n = .416 \times 10^{-2}$ and the residual with variance 2.63. Thus, the random walk component is negligible here.

A series for *ex ante* real interest rate has been constructed by subtracting a one-step- ahead forecast of the CPI inflation rate from the call rate. It has been decomposed by a random quadratic polynomial.

USA

The logarithms of the monthly US IPI are detrended by a random quadratic polynomial.

The monthly federal government expenditure was divided by the consumer price index, the logarithms of which is then detrended by a random linear trend.

APPENDIX 2

Methodological Background

We use the procedure as described in AOKI [1983, 1985, 1987] and in AOKI and HAVENNER [1986], to construct

$$s_{t+1} = \mathrm{A}\, s_t + \Gamma e_t$$
$$d_t = \mathrm{C}\, s_t + e_t$$

where d_t is a p-dimensional (suitably detrended) data vector. See Appendix on the (random) detrending one must perform to obtain $\{s_t\}$ which is weakly stationary. The vector s_t is the Kalman filter output of the state vector, i. e., $s_t = x_{t|t-1}$ where $x_{t|t-1}$ is the orthogonal projection of the state vector x_t onto the subspace spanned by past data vector $d_{t-1}, d_{t-2}, \ldots$. The state vector x_t is a vector of conditionally sufficient statistics for $d_t, d_{t-1} \ldots$. It may also be interpreted as a vector of dynamic factors or other unobservables.

The one-step-ahead forecast of d_{t+1} given $d_t, d_{t-1}, \ldots$, denoted by $d_{t+1|t}$, is then related to the state vector by

$$d_{t+1|t} = \mathrm{C}\, s_{t+1}$$

where

$$s_{t+1} = \mathrm{A}\, s_t + \Gamma (d_t - \mathrm{C}\, s_t) = \hat{\mathrm{A}}\, s_t + \Gamma d_t$$

where $\hat{\mathrm{A}} = \mathrm{A} - \Gamma\, \mathrm{C}$.

By eliminating $\hat{\mathrm{A}}$ by means of the Cayley-Hamilton theorem, an equivalent ARMA form for the one-step-ahead forecast error update equation can be derived. To be concrete, suppose that $\hat{\mathrm{A}}$ is two by two with the characteristic polynomial $\lambda^2 - \alpha\ \lambda - \beta$. Then multiply the first equation below by $-\beta$ and the second by $-\alpha$ and add them to the third

$$d_{t-1|t-2} = \mathrm{C}\, s_{t-1}$$
$$d_{t|t-1} = \mathrm{C}(\hat{\mathrm{A}}\, s_{t-1} + \Gamma d_{t-1})$$
$$d_{t+1|t} = \mathrm{C}\hat{\mathrm{A}}\,(\hat{\mathrm{A}}\, s_{t-1} + \Gamma d_{t-1}) + \mathrm{C}\,\Gamma d_t$$

to produce the forecast updating equation

$$d_{t+1|t} = \alpha d_{t|t-1} + \beta d_{t-1|t-2} + \mathrm{C}\,\Gamma d_t + \mathrm{C}(\hat{\mathrm{A}} - d\,\mathrm{I})\,\Gamma d_{t-1}.$$

Noting that $d_{t+1|t} = d_{t+1} - e_{t+1}$ etc., this equation is the same as the ARMA (2, 2) model

$$d_{t+1} - (\alpha \mathrm{I} + \mathrm{C}\,\Gamma)\, d_t - [\beta \mathrm{I} + \mathrm{C}(\hat{\mathrm{A}} - \alpha \mathrm{I})\,\Gamma]\, d_{t-1} = e_{t+1} - \alpha e_t - \beta e_{t-1}.$$

Conversion to ARMA forms

State space models are converted into ARMA forms by eliminating state vectors. A basic way for this is to append to the Cayley-Hamilton theorem,

as explained in AOKI [1976, 1983, 1987]. When the matrix C is square, then it is necessarily nonsingular because the model is observable by construction. Suppose that C is square, then use $x_t = C^{-1}(y_t - e_t)$, together with the state transition equation to produce a vector-valued ARMA(1, 1) :

$$\begin{aligned} y_{t+1} &= C x_{t+1} + e_{t+1} \\ &= CAC^{-1}(y_t - e_t) + CB e_t + e_{t+1}. \end{aligned}$$

When $p < n$, then it is not true that all the row vectors of the matrices C, CA, etc are linearly independent. The observability condition guarantees, however, that n-p additional linearly independent row vectors can be selected from the observability matrix and appended to the matrix C to construct a nonsingular matrix.

To convey the basic idea of this alternative procedure, suppose that C is 1 by 2. Then C and CA are two linearly independent row vectors and hence

$$\begin{pmatrix} y_t \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} C \\ CA \end{pmatrix} x_t + \begin{pmatrix} I \\ CB \end{pmatrix} e_t + \begin{pmatrix} 0 \\ e_{t+1} \end{pmatrix},$$

where the linear independence of the row vectors of C and CA is guaranteed by the observability. Treating y_t and y_{1t+1} together as a new data vector, the state vector x_t is eliminated to produce

$$\begin{aligned} \begin{pmatrix} y_{t+1} \\ y_{t+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \varphi_2 & \varphi_1 \end{pmatrix} \left[\begin{pmatrix} y_t \\ y_{t+1} \end{pmatrix} - \begin{pmatrix} 1 \\ CB \end{pmatrix} e_t - \begin{pmatrix} 0 \\ e_{t+1} \end{pmatrix} \right] \\ + \begin{pmatrix} C \\ CA \end{pmatrix} B e_t + \begin{pmatrix} e_{t+1} \\ CB e_{t+1} \end{pmatrix} + \begin{pmatrix} 0 \\ e_{t+2} \end{pmatrix} \end{aligned}$$

where

$$\varphi_1 = CA^2 d_2 \qquad \text{and} \qquad \varphi_2 = CA^2 d_1$$

where

$$\begin{pmatrix} C \\ CA \end{pmatrix}^{-1} = (d_1, d_2).$$

Note that the first row of this equation is an identity and the second row produces an ARM(2, 2) relation for y_t :

$$\begin{aligned} y_{t+2} = \varphi_1 y_{t+1} + \varphi_2 y_t + e_{t+2} + (CB - \varphi_1) e_{t+1} \\ + (CAB - \varphi_1 CB - \varphi_2) e_t. \end{aligned}$$

In general when the matrix C contains several row vectors, care must be exercised to use the row vector which produces linearly independent row vectors when multiplied by some powers of A. Several such different row vectors of C could conceivably be needed to construct a nonsingular matrix.

• References

AKAIKE, H. (1974). – "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes," *Annals of the Institute of Statistical Mathematics*, 26, p. 363-387.

AOKI, M. (1971). – "Aggregation," Chapter 5 in *Optimization Methods for Large-Scale Systems*, WISMER D. A. Ed., McGraw Hill, New York.

AOKI, M. (1981). – *Dynamic Analysis of Open Economies*, Academic Press., New York.

AOKI, M. (1983). – "Notes on Economic Time Series Analysis : System Theoretic Perspectives", No. 220 *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag Heidelberg.

AOKI, M. (1987). – *State Space Modeling of Time Series*, Springer-Verlag, Heidelberg.

AOKI, M. (1987*a*). – "How to Build State Space Models for Nonstationary Time Series", *UCLA working paper*, n° 439, March 1987, also to be presented at the Sixth International Conference on Mathematical Modeling, St-Louis, USA.

ARUN, K. S., RAO B., and KUNG S. Y., (1983). – "A New Prediction Efficiency Criterion for Approximate Stochastic Realization", 1983 Conference Revision and Control Preprint, p. 1353-55.

ENGLE, R and WATSON M. (1981). – "A One-Factor Multivariante Time Series Model of Metropolitan Wage Rates", *JASA*, 76, p. 774-781.

FAURRE, P. L., CLERGET M. and GERMAIN F. (1979). – *Operateurs rationels positifs*, Dunod, Paris.

GEL'FAUD I. M. and YAGLOM, A. M. (1959). – "Calculation of the Amount of Information about Random Function Contained in Another Such Function", *American Mathematical Society Translation.* Series 2, 12, p. 191-198.

GEWEKE, J. F. (1977). – "The Dynamic Factor Analysis of Economic Time Series" in AIGNER D. J. and GOLDBERGER, A. S. Eds. *Latent Variables in Socioeconomic Models*, North Holland, Amsterdam, p. 365-383.

GLOVER, K. (1984). – "All Optimal Hankel-Norm Approximations of Linear Multivariable Systems and Their L^∞ Bounds", *International Journal of Control* 39, p. 1115-1193.

GRANGER, C. W. J. (1983). – "Co-Integrated Variables and Error-Correcting Models", *Working Paper*, Dept. of Economics, UC San Diego.

HOUSEHOLDER, A. S. (1964). – *The Theory of Matrices in Numerical Analysis*, Blaisdell Pub. Co., New York.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). – *Gaussian Random Processes*, Springer-Verlag, New York.

KITAGAWA, G. (1985). – Decomposition Program in TIMSAC84. Akaike *et al.*, Inst. Stat. Math., Tokyo.

LINDQUIST, A., PICCI G. and RUCKEBUSCH, G. (1979). – "On Minimal Splitting Subspace and Markovian Representation", *Mathematical System Theory*, 12, 271-279.

MOORE, B. C. (1978). – "Singular value analysis in linear systems, I and II", Preprint IEEE Conference Decision and Control, p. 66-73.

PERNEBO, L and SILVERMAN L. (1982). – "Model Reduction via Balanced State Representations", IEEE AC-27, p. 382-387.

PICCI, G. (1976). – "Stochastic Realization of Gaussian Processes", IEEE Proceedings, 64, 112-122.

QUAH, D. (1986). – "How are International Business Cycles Also Alike", *unpublished manuscript*.

RAO, R. C. (1964). – "The Use and Interpretation of Principal Component Analysis in Applied Research" *Sankya*, Series A, 26.

SARGENT, T. J., and SIMS C. A. (1977). – "Business Cycle Modeling Without Pretending To Have Too Much a Priori Economic Theory" in *Proceedings of New Methods in Business Cycles*, Research Federal Reserve Bank of Minneapolis, MN.

SINGLETON, K. J. (1980). – "A Latent Time Series Model of the Cyclical Behavior of Interest Rates", *IER*, 21, p. 559-575.