A STATE SPACE TIME SERIES MODELING METHOD WITHOUT PRIOR DETRENDING

by

Masanao Aoki
University of California, Los Angeles
USA

UCLA Working Paper No. 465
February, 1988

A State Space Time Series Modeling

Method without Prior Detrending †

by

Masanao Aoki *
University of California, Los Angeles
USA

Abstract

A state space method for building time series models without detrending each component of data vectors is presented. The method uses the recent algorithm based on the singular value decomposition of the Hankel matrix and a two step sequential procedure suggested by the notion of dynamic aggregation.

* 4731 Boelter Hall, University of California, Los Angeles, California 90024-1600, USA

† Research supported in part by a grant from the National Science Foundation.

1. Introduction

Time series are usually decomposed into trends and the remainders (consisting of cyclical, seasonal, and residual components, or simply cyclical and residual) because trends convey information distinct from that to be culled from cyclical components. In macroeconomic time series, for example, policy-makers may be primarily interested in their trend behavior, while those concerned with business cycles are more interested in cyclical components, such as phases of business cycles than trends. Another reason for singling out trends is that they may have simpler dynamic structure than cyclical components in the sense that a small number of "common" factors are responsible for a larger number of trend components, as in macroeconomic time series where there are reasons to suspect or expect from economic theory for a set of macroeconomic variables to behave generally in the same way, at least in longer-run time horizon, i.e., aside from short-run (individual) variations. Here again one needs to separate "trend" components, and seek a set of a small number of common factors that may "cause" a larger number of macroeconomic variables to change, and to extract "common" trend components from these macroeconomic time series. Granger's notion of co-integration (1983) is one way to formalize this idea of common factors.

Time series are often transformed to render them weakly stationary for a technical reason that currently available modeling methods can more efficiently handle weakly stationary time series than nonstationary ones. One transformation takes differences of the logarithms of data series. A serious drawback of this common practice is that longer-run information of time series is lost in the process of rendering them weakly stationary. Recent interests in modeling economic time series without prior detrending is sparked by the seminal works of Beveridge and Nelson (1981) and Nelson and Plosser (1982), who posited a model of time series with separate and explicit equations for random trends. Harvey (1984) also used models with explicit random trend dynamics. Random trends are provided for by specifying that the first difference is weakly stationary. In other words, random trend dynamics are hypothesized to have a unit root. By now a number of studies is available which examines the question of unit roots in the economic time series, such as the US real GNP series, See Clark (1987), Perron and Phillips (1987), for example. In multivariate time series, however, this approach posits the same number of unit roots as the number of component series with "trends", which often results in too many unit roots. A transformation which extracts a smaller number of common trends than this approach is needed.

This paper proposes an alternative modeling procedure for separating out trend dynamics from those for the cyclical and residual components without constraining the components of time series to have unit roots from the beginning, and thus allow for easy determination of the presence (and the number) of common factors. The idea is based on the notion of dynamic aggregation which was originally suggested as a way for building simplified dynamic models for control purposes, see Aoki (1968). We build time series models in two sequential steps. In the first step state space models for trends are built followed by a second step in which state space models for the residuals of the first step are constructed. In each of the two steps, state space modeling algorithms recently proposed by Aoki (1983, 1987a) is employed. Aoki (1987b) has recently pointed out that Granger's co-integration and the idea of error correction mechanism, originally proposed by Sargan (1964) are derivable from the common notion of aggregation of

dynamic models.

The procedure will not require prior detrending as in Stock and Watson (1986), and will determine co-integrating factors, when some of the components of the vector-valued time series contain common trends. This paper also discusses why this two-step procedure may be superior to a single state modeling strategy, especially when trend components contain random walk components. Section 2 is a brief description of the dynamic aggregation procedure originally employed in Aoki (1968). Section 3 describes how to construct state space models in two stages following the suggested scheme in Section 2. Section 4 clarifies the differences in the extraction of trends in Beveridge Nelson and the state space models. Some examples are presented in Section 5 and the concluding Section 6 elaborates on the reasons why one might wish to employ the suggested two stage procedure.

2. Dynamic Aggregation

The dynamic aggregation procedure in Aoki (1968) starts by classifying dynamic modes of a model $x_{t+1} = Ax_t + v_t$, i.e., eigenvalues of the transition matrix A into two classes and transforming the coordinates to put A into a block triangular representation. Although many dichotomized classifications are possible, here put all eigenvalues with magnitude greater than some critical number into class C_1 and the rest in class C_2. Thus C_1 contains unit roots and those roots of the characteristic polynomial near the unit circle in the complex plane. Suppose that A is $n \times n$ and that there are k eigenvalues in C_1 (counting multiplicities). Let $n \times k$ matrix T form a basis for the right invariant subspace of linearly independent columns of A associated with the eigenvalues in C_1. If $At_i = t_i\lambda_i$, $i = 1 \ldots k$, then $T = [\, t_1 \ldots t_k \,]$ and $\Lambda =$ diag $(\lambda_1 \ldots \lambda_k)$, for example. Let S be an $n \times (n-k)$ matrix of linearly independent columns forming a left invariant subspace of A associated with class C_2. They satisfy,

$$\begin{aligned} AT &= T\Lambda \\ S'A &= NS' \end{aligned} \tag{1}$$

and we normalize T and S by $T'T = I_k$ and $S'S = I_{n-k}$. These two matrices are orthogonal

$$S'T = 0.$$

because $S'AT = S'T\Lambda$ and $S'AT = NS'T$ implies $0 = S'T\Lambda - NS'T$ and Λ and N have no eigenvalues in common. See Householder (1965, p. 168). Express the state vector x_t using the columns of P and S as basis vectors, i.e., let

$$x_t = Sz_t + T\tau_t$$

in the model $x_{t+1} = Ax_t + u_t$ i.e., $Sz_{t+1} + T\tau_{t+1} = T\Lambda\tau_t + ASz_t + v_t$, where the first equation in (1) is used. The vector τ_t is the set of new coordinates related to slower dynamic modes and z_t refers to the coordinates representing faster dynamic modes. Multiply this relation from the left by T' to see that

$$\tau_{t+1} = \Lambda\tau_t + T'ASz_t + T'v_t. \tag{2}$$

Multiplication from the left by S' yields

$$z_{t+1} = Nz_t + S'v_t. \tag{3}$$

The matrix N has eigenvalues of class C_2 only, i.e., they are all asymptotically stable eigenvalues by choice. Jointly written, the state space model has the recursive structure

$$\begin{bmatrix} \tau_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} \Lambda & T'AS \\ 0 & N \end{bmatrix} \begin{bmatrix} \tau_t \\ z_t \end{bmatrix} + \begin{bmatrix} T' \\ S' \end{bmatrix} v_t \tag{4}$$

Note that the term $T'AS$ explicitly shows how the state vector for short-run dynamics affect longer run dynamics. The model specification is completed by specifying that the data vector y_t is related to x_t by $y_t = Cx_t + v_t$. The data vector y_t is related to the new vectors τ_t and z_t by

$$y_t = CT\tau_t + CSz_t + v_t. \tag{5}$$

Eq. (5) shows how the data is decomposed into slower modes, i.e., trend (-like) movements $CT\tau_t$ and the rest $CSz_t + e_t$, i.e., cyclical component plus innovations on observations.

3. Modeling Procedure †

The previous section suggests a procedure to construct a model with a block triangular transition matrix. Since the dynamic matrix of time series in unknown, we do not know how many eigenvalues are in C_1. The data determines the dimension of the vector τ_t. In the algorithm of Aoki (1987a), the ratio of singular values of certain Hankel matrix is one important indication of the size of n. First, trend models is estimated

$$\begin{cases} \tau_{t+1} = \Lambda\tau_t + Gu_t \\ y_\tau = D\tau_t + u_t \end{cases} \tag{6*}$$

where u_t stands for $CSz_t + v_t$ in (5), followed by a model for short-run behavior

$$\begin{cases} z_{t+1} = Fz_t + Je_t \\ u_t = Hz_t + e_t. \end{cases} \tag{7}$$

Note that u_t is weakly stationary since the dynamics for z_t are stable by construction. In (6) u_t is usually (highly) serially correlated but e_t in (7) are not serially correlated.

When τ_t is chosen to be scalar, (6) is

---

† For completeness the model matrix estimation method in Aoki (1983, 1987a) is outlined in the Appendix.

* A seemingly more general model $x_{t+1} = Ax_t + v_t$, $y_t = Cx_t + w_t$, where u_t and w_t are serially uncorrelated, can be put in the assumed form where cov $\begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} G \\ I \end{bmatrix} \Delta(G' \; I)$, and $\Delta = \text{cov } e$ by spectral decomposition. See Aoki (1987, p. 67).

$$\begin{aligned} \tau_{t+1} &= \lambda \tau_t + g' u_t \\ y_t &= d \tau_t + u_t \end{aligned} \tag{8}$$

where g and d are p-dimensional column vectors where $p = \dim y_t$. The connection with Granger's notion of co-integration is now clearly seen from (8). Any vector ν orthogonal to d will nullify the dynamic mode with eigenvalue λ since $\nu' y_t = \nu' u_t$ is governed by the dynamics (7) and has no eigenvalue in C_1, i.e., $\nu' y_t$ is weakly stationary, even when some components of y_t have unit roots.

If the dimension 2 is tried, then the matrix Λ in (6) is 2×2. When it has two real eigenvalues one can decide then whether the trend dynamics has one dominant eigenvalue or two. When the data y_t contains a single common trend variable, this fact becomes apparent when the eigenvalues of the matrix Λ is calculated. The matrix Λ can be put into Schur form to exhibit this fact explicitly as

$$\Lambda = U \Theta U', \quad U'U = I_2$$

where

$$\Theta = \begin{bmatrix} \lambda_1 & * \\ 0 & \lambda_2 \end{bmatrix}, \quad \lambda_1 > \lambda_2.$$

If λ_1 is judged to be significantly larger than λ_2, then there is one trend factor. If λ_1 and λ_2 are sufficiently close to each other, then there are two common factors. Define $\mu_t = U' \tau_t$. Model (6) is transformed into

$$\begin{aligned} \mu_{t+1} &= \begin{bmatrix} \lambda_1 & * \\ 0 & \lambda_2 \end{bmatrix} \mu_t + U'G \, u_t \\ y_t &= DU \, \mu_t + u_t. \end{aligned}$$

The first column vector of DU now corresponds to the vector d of (8) since it shows how the first component of μ_t is distributed among the components of the vector y. Usually dim y_t is larger than dim μ_t. Thus DU is a disaggregation matrix which distributes the effects of μ_t among y_t. To allow for the possibility of a real eigenvalue and a pair of complex eigenvalues in the trend dynamics dim $\tau_t = 3$ should also be considered. Too high an initial choice of the dimension of τ_t causes no harm since the Schur decomposition tells us if the eigenvalues are all equally large and the column vectors of DU tells us if the components of the vector τ_t are equally important in y_t. If not, some of the modes in (6) can be easily lumped together with (7). We return to this point in Section 5 where an example is discussed.

Eq. (6) and (7) imply that the transfer function from e_t to y_t can be factored as

$$\begin{aligned} \hat{y}_t &= [\, I + D(qI - \Lambda)^{-1} G \,] \, \hat{u} \\ &= [\, I + D(qI - \Lambda)^{-1} G \,] \, [\, I + H(qI - F)^{-1} J \,] \, \hat{e} \end{aligned} \tag{9}$$

where q^{-1} is the lag operator $q^{-1} y_t = y_{t-1}$. The first factor of this factorization corresponds to the slower dynamics, i.e., lower frequency factor, the second to the factor dynamics, i.e., fast fre-

quency factor. This modeling method in effects factors the transfer function into a low frequency factor and a high frequency one as shown.

Since the residuals in (6) are usually correlated, unlike the modeling situations for weakly stationary, u_t is not an innovation vector. To show that the Riccati equation used in the algorithm of Aoki (1987a) is well-defined, consider the model with a scalar τ_t as an example.

Rewrite (8) as

$$\tau_{t+1} = \hat{\lambda}\tau_t + g'y_t$$

by substituting u_t out from the first equation, where $\hat{\lambda}$ is equal to $\lambda - g'd$. If $\sum_{k=0}^{\infty} \hat{\lambda}^k R_{t+k,t}$ converges for all t, where $R_{t+k,t} = E(y_{t+k}y'_t)$ is the covariance matrix of the data vector, then the covariance matrix of τ_t is well-defined and Aoki's algorithm (1987a) can be applied to estimate λ and g'. *

Similarly, rewrite (7) as

$$z_{t+1} = (F - JH)z_t + Ju_t.$$

Then, the covariance matrix $\Pi = \text{cov } z_t$ is well defined, and the subscript t is dropped from Π because z process is weakly stationary if all the eigenvalues of $F - JH$ lie strictly inside the unit disk. Let $\hat{F} = F - JH$. The matrix $\hat{F}$ is the dynamic matrix in the Kalman filter for (7). It is known that if (7) is observable $\hat{F}$ is asymptotically stable. To see the importance of this condition, consider solving the Riccati equation for Π by an iterative procedure, where $\Pi = \hat{F}\Pi\hat{F} + JU_0J'$ and where $U_0 = \text{cov } u = H\Pi H' + \Delta$, and where $\Delta = \text{cov } e_t$. Supposing that Π exists, denote $\Pi_k - \Pi$ by P_k. Then $P_{k+1} = \hat{F}P_k\hat{F}$ or

$$\text{vec } P_{m+1} = (\hat{F} \otimes \hat{F})\text{vec } P_m, 1, 2, \cdots$$

Therefore, the equation converges as m is increased if and only if all the eigenvalues of $\hat{F}$ lie inside the unit disk as claimed.

The above description shows that the proposed two-step procedure will construct state space models even when the data vector contain unit root components, i.e., even when λ in (7) is one, provided $\hat{\lambda}$ is less than one in magnitude. A sufficient condition is that the unit root is an observable mode of the model.

4. Decomposition into Trends and Cyclical Components

Beveridge and Nelson (1981) posit that a univariate y_t is governed by

---

* For example, if y_t is a pure random walk, then $R_{t+k,t} = R_{t,t} + k\sigma^2$. The sum $\Sigma_k k\hat{\lambda}^k$ converges if the magnitude of $\hat{\lambda}$ is less than one.

$$y_t = y_{t-1} + A(q^{-1})e_t \tag{10}$$

where $q^{-1}e_t = e_{t-1}$ and

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots$$

such that $\sum_0^\infty a_i^2 < \infty$. The coefficients in the Wold decomposition representation (10) are the impulse (dynamic multiplier) responses. For example a_3 in $\Delta y_t = A(q^{-1})e_t$ tells us how much the shock three period earlier, e_{t-3}, still affects Δy_t. This class of models has been proposed by Beveridge and Nelson (1981) and used by Nelson and Plosser (1982), Cochrane (1986) and others. This section demonstrates the difference in the decomposition of time series by this and state space methods for a univariate $\{y_t\}$.

By rewriting (10) as

$$y_t - y_{t-1} = A(1)e_t + [A(q^{-1}) - A(1)]e_t ,$$

where

$$A(1) = \Sigma_0^\infty a_i < \infty$$

is assumed, one can integrate this equation. Decompose y_t into $y_{1t} + y_{2t}$ where

$$\Delta y_{1t} = A(1)e_t \tag{11}$$

and

$$\Delta y_{2t} = [A(q^{-1}) - A(1)]e_t , \tag{12}$$

where $\Delta y_{it} = y_{it} - y_{it-1}$, $i = 1, 2$. Eq. (11) immediately shows that y_{1t} is a random walk because

$$y_{1t} - y_{1t-1} = (1-q^{-1})y_{1t} = A(1)e_t ,$$

or

$$y_{1t} = \frac{A(1)}{1-q^{-1}} e_t$$

$$= A(1)(e_t + e_{t-1} + e_{t-2} + \cdots). \tag{13}$$

It is the integral of past disturbances. This term represents the random trend in Beveridge-Nelson decomposition.

Solve y_{2t} from (12) as

$$y_{2t} = \frac{A(q^{-1}) - A(1)}{1-q^{-1}} e_t . \tag{14}$$

When the spectral density function of Δy_t is rational, Δy_t can be regarded as an output of a finite-dimensional state space model driven by a white noise sequence. The impulse responses $\{a_i\}$ are then characterized by finite parameter combinations. When $a_i = h'F^{i-1}g$, $i \geq 1$, $a_0 = 1$, we use

$$A(q^{-1}) = I + h'(qI - F)^{-1}g.$$

Then

$$\frac{A(q^{-1}) - A(1)}{1 - q^{-1}} = -H(q^{-1}) \tag{15}$$

where $H(q^{-1}) = h'(I - Fq^{-1})^{-1}(I - F)^{-1}g$. To see this directly let $A(q^{-1}) = \Sigma_0^{\infty} a_i q^{-1}$ and express

$$A(q^{-1}) - A(1) = a_1(q^{-1} - 1) + a_2(q^{-1} - 1) + a_3(q^{-1} - 1) +$$

$$= (q^{-1} - 1)\,[\,a_1 + a_2(q^{-1} + 1) + a_3(q^{-1} + q^{-1} + 1) + \ldots\,]$$

since $\{a_i\}$ is absolutely convergent, and note that

$$a_1 + a_2 + a_3 + \ldots = h'(I + F + F^2 + \ldots)g = h'(I - F)^{-1}g\ .$$

All other the term are similarly expressed in closed forms as

$$a_2 + a_3 + a_4 + \cdots = h'F(I - F)^{-1}g\ ,$$

and

$$a_3 + a_4 + \cdots = h'F^2(I - F)^{-1}g \quad etc.$$

Thus

$$A(q^{-1}) - A(1) = (q^{-1} - 1)h'(I + F + F^2 + \cdots)(I - Fq^{-1})^{-1}g$$

$$= (q^{-1} - 1)h'(I - F)^{-1}(I - Fz^{-1})^{-1}g$$

$$= (q^{-1} - 1)h'(I - Fq^{-1})^{-1}\psi$$

where

$$\psi = (I - F)^{-1}g.$$

From (14) and (15)

$$y_{2t} = -H(q^{-1})e_t = -(a_1 e_t + a_2 e_{t-1} + \cdots)$$

Since cov $y_{2t} = \Sigma_1^{\infty} a_i^2 < \infty$, y_{2t} is weakly stationary. Note that e_t appears both in y_{1t} and in y_{2t} in the Beveridge-Nelson decomposition.

The state space model decomposes y_t differently into two components. In the state representation, the eigenvalue $\lambda = 1$ is not imposed in the model. When (random) trends have one-dimensional dynamics, the model takes the form of (7) and (8) ore written jointly as

$$\begin{bmatrix} \tau \\ z \end{bmatrix}_t = \begin{bmatrix} \lambda & g'H \\ 0 & F \end{bmatrix} \begin{bmatrix} \tau \\ z \end{bmatrix}_{t-1} + \begin{bmatrix} g' \\ J \end{bmatrix} e_{t-1}\ ,$$

$$y_t = c\,\tau_t + H'z_t + e_t\ , \tag{16}$$

where e_t is serially uncorrelated. Eq. (16) shows that $y_{t|t-1}$, the predictable component of y_t at

time $t-1$, consists of trend part

$$y_{1t|t-1} = c\,\tau_{t|t-1}$$

and cyclical component part

$$y_{2t|t-1} = H'z_{t|t-1}\,.$$

By defining $y_{1t} = y_{1t|t-1}$ and $y_{2t} = y_{2t|t-1} + e_t$, we decompose y_t as the sum of y_{1t} and y_{2t} where the former is a predictable process. To relate this to the lag-transform (9), which in this case is equal to

$$\hat{y} = [\,1 + c(I - \lambda q^{-1})^{-1} g' q^{-1}\,]\,[\,1 + H'(I - Fq^{-1})^{-1} Jq^{-1}\,]\,\hat{e},$$

define

$$\hat{y}_1 = c(I - \lambda q^{-1})^{-1} g' q^{-1}\,[\,1 + H'(I - Fq^{-1})^{-1} Jq^{-1}\,]\,\hat{e}$$

and

$$\hat{y}_2 = [\,1 + H'(I - Fq^{-1})^{-1} Jq^{-1}\,]\,\hat{e}.$$

To compare with the Beveridge-Nelson procedure, suppose that $\lambda = 1$. Then

$$\hat{y}_1 = \frac{cg'q^{-1}}{1 - q^{-1}}\,[\,1 + H'(I - Fq^{-1})^{-1} Jq^{-1}\,]\,\hat{e}$$

or

$$y_{1t} - y_{1t-1} = cg'\,[\,e_{t-1} + H'Je_{t-2} + H'FGe_{t-3} + \ldots\,]$$

and

$$y_{2t} = e_t + H'Je_{t-1} + H'FJe_{t-2} + \cdots$$

Note the difference in the timing of e. The latest innovation is assigned only to y_{2t}. In other words, the trend component of the state space model y_{1t} lies in the subspace spanned by $e_{t-1}, \ldots$, i.e., is a predictable process in the sense of Doob decomposition of an arbitrary stochastic process [Kopp (1984, p. 66)] and y_{2t} is a martingale process. Beveridge-Nelson decomposition does not have this "canonical" property.

5. Examples

The Ex Ante Real Interest Rate in the USA and West Germany

Ex Ante real interest rates are calculated as

$$r_t = i_t - \pi_{t|t-1}$$

where

$$\pi_{t|t-1} = p_{t|t-1} - p_{t-1}$$

and where p_t is the logarithm of the cost of living index or consumer price index for the month t, and $p_{t|t-1}$ is the forecast of p_t formed at time $t-1$. In this paper only the history of $p_{t-1}, p_{t-2} \ldots$ is used in forming the conditional estimate $p_{t|t-1}$.

Figure 1 is the plot of the r_t series for the USA and West Germany from March 1974 to September 1984. We note that the real interest rate in Germany started to increase about March 1979 and stayed high. The US counterpart also increased in the same time span.

A trend the model with $n = 2$ has the eigenvalues .981 and .570. This is a clear indication that there is one eigenvalue responsible for possibly nonstationary characteristics of the two series. Although the residuals are not too highly correlated, they are still significant. The first two lag covariance matrices of $\{u_t\}$ are:

$$\text{lag } 0 \begin{bmatrix} 1.60 & \\ -.085 & .105 \end{bmatrix},$$

$$\text{lag } 1 \begin{bmatrix} .122 & \\ .079 & .031 \end{bmatrix},$$

$$\text{lag } 2 \begin{bmatrix} -.071 & \\ .054 & .025 \end{bmatrix}.$$

The long-run (trend) dynamics are modeled by

$$\tau_{t+1} = \Lambda\tau_t + Gu_t$$
$$y_t = D\tau_t + u_t$$

where

$$\Lambda = \begin{bmatrix} .9765 & -.0549 \\ -.0349 & .5750 \end{bmatrix} = UTU'$$

where

$$U = \begin{bmatrix} .9963 & .0856 \\ -.0856 & .9963 \end{bmatrix}, \quad T = \begin{bmatrix} .9812 & -.0200 \\ 0 & .5703 \end{bmatrix},$$

$$G = \begin{bmatrix} .043 & .238 \\ .471 & -.408 \end{bmatrix}$$

and

$$DU = \begin{bmatrix} .2910 & 1.1825 \\ .3788 & 1.1702 \end{bmatrix}.$$

Only one of the two dynamic factors is associated with the slow dynamic eigenvalue $\lambda_1 = .9812$. After some sensitivity experiments, the residual series $\{u_t\}$ is further modeled by a second order dynamics.

$$F = \begin{bmatrix} -.301 & .312 \\ .510 & .511 \end{bmatrix}$$

with eigenvalues $-.469$ and .674. The e_t series has an excellent residual characteristic, as

shown by the first three residual sample covariance matrices:

$$\text{lag } 0 \begin{bmatrix} 1.57 & \\ -.100 & .076 \end{bmatrix},$$

$$\text{lag } 1 \begin{bmatrix} .025 & \\ -.009 & .003 \end{bmatrix},$$

$$\text{lag } 2 \begin{bmatrix} -.032 & \\ .008 & .0002 \end{bmatrix}.$$

The covariance matrix of the innovation vector,

$$cov\, e = \begin{bmatrix} 1.590 & \\ -.093 & .081 \end{bmatrix}$$

shows that the US innovation is much larger than that of Germany. Note that e_{US} and e_G is correlated with the correlation coefficient of about $-.26$.

The US M1 and CPI

The monthly bivariate series for logarithms of the US M1 and CPI (consumer price index) is used to illustrate the procedure. From January 1975 on, with 117 data points, when a two-dimensional trend dimension is chosen, the method estimates the dynamic matrix A to be

$$A = \begin{bmatrix} .977 & -.030 \\ -.044 & .953 \end{bmatrix}$$

which has two real eigenvalues $\lambda_1 = .981$ and $\lambda_2 = .948$. When A is put into Schur form, the trend model becomes

$$\mu_{t+1} = \begin{bmatrix} .981 & -.025 \\ 0 & .948 \end{bmatrix} \mu_t + \ldots$$

$$y_t = \begin{bmatrix} .193 & .048 \\ .227 & .015 \end{bmatrix} \mu_t + \ldots$$

First, we note that λ_1 is sufficiently close to 1 (statistically indistinguishable from 1) so that M1 and CPI has a common random walk factor for all practical purposes. Note that $(.981)^{120} = .10$, i.e., after 10 years still 10% of the initial effects remain. The residuals are modeled by a two-dimensional dynamics. The total model is specified as in (4) to have the dynamic matrix

$$\begin{bmatrix} \Lambda & GH \\ 0 & F \end{bmatrix} = \begin{bmatrix} .977 & -.030 & .032 & .009 \\ -.004 & .953 & .012 & .012 \\ 0 & 0 & .943 & -.125 \\ 0 & 0 & .066 & -.075 \end{bmatrix}$$

$$\begin{bmatrix} G \\ J \end{bmatrix} = \begin{bmatrix} 3.05 & 1.56 \\ 39.77 & -35.96 \\ 42.30 & 34.29 \\ -917.07 & 49.55 \end{bmatrix}$$

and

$$[D,\ 1-1] = \begin{bmatrix} .197 & .020 & .007 & .002 \\ .227 & -.017 & .007 & .002 \end{bmatrix}.$$

Second, the contribution of the second component of μ_t is an order of magnitude smaller than that of the first component. Third, λ_2 cannot be identified with 1 because the off-diagonal elements of A have the same sign. This implies that A has two real eigenvalues, one of them is strictly greater than the (1, 1) element of A and the other strictly less than the (2, 2) element of A. Thus, A can't have two unit eigenvalues, even when $\lambda_1 = \lambda_2$ The bivariate series can have at most one unit root. Any vector ν with $\nu_t / \nu_2 = -.227 / .193 - 1.18$ renders $\nu' y_t$ stationary.

An alternative model is a scalar trend model which is

$$\tau_{t+1} = .977\tau_t + g' u_t$$

$$y_t = \begin{bmatrix} .193 \\ .227 \end{bmatrix} \tau_t + u_t$$

Note that .977 is the (1, 1) element of A and the first column vector of C as the disaggregation matrix in this one-dimensional model. This is because the models are put in balanced form. See Aoki (1987, Section 5.6) for the reasons. Scalar- and two-dimensional trend models mean different residual sequences to be fitted in the second stage. Figure 2 and 3 plot residuals with $n = 1$ and 2. The choice between the two must be made on the performances of the joint dynamic model (6) and (7). The model that produces $\{e_t\}$ most closely resembling the white noise sequence should be chosen, which is the one shown above.

6. Concluding Remarks

One may naturally wonder why the same model can't be obtained in one step using the same algorithm. One reason that this does not yield "good" models lies in the empirical fact that model building algorithms need to deal with quantities with large disparate magnitudes when data series contain random trend components. The ratio of the largest to the next largest singular values of Hankel matrices may exceed 10^3 when these are significant random trend components. In such situations trend components completely dominate small cyclical components. Because the algorithm relies on the relative sizes of singular values of the covariance matrix between the stacked future realization and stacked past data (Hankel matrix), models tend to ignore those components of the data vectors with small variances when they show large discrepancies in the variances (such as 10^3 to 1). For these two reasons, it is desirable to have a two-step procedure in which the residuals from the first step may be rescaled before applying the

second step of the algorithm.

The first stage of the proposed procedure may be interpreted as a new transformation or a new detrending scheme which better preserves trend information than taking the first differences of the logarithms because it is not constrained to produce the same detrending factors as the number of components of the data series.

References

Aoki, M., *State Space Modeling of Time Series,* Springer-Verlag, Heidelberg, 1987 a

Aoki, M., "Co-integration, Error Correction, and Aggregation in Dynamic Models: A Comment", submitted to *Oxford Bulletin of Economic Statistics*, 1987c

Aoki, M., *Notes on Economic Time Series: System Theoretic Perspectives, Lecture Notes in Economics and Mathematical Systems No. 220* Springer-Verlag, Heidelberg and New York, 1983

Beveridge, S., and C. R. Nelson, "A New Approach to Decomposition of Economics Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'", J.M.E. 7, p. 151-174, 1981

Campbell, J. R. and N. G. Mankiw, "Are Output Fluctuations Transitory", NBER Meeting on Economic Fluctuations, March 1986

Clark, P. K., "The Cyclical Component of US Economic Activity" *Stanford Graduate School of Business W. P., No. 875,* 1986

Cochrane, J. H., "How Big is the Random Walk in GNP", *Working Paper,* University of Chicago, April 1986

Granger, C. W. J., "Co-Integrated Variables and Error-Correcting Models", UCSD Discussion Paper, 83-13a, 1983

Harvey, A. C., "Trends and Cycles in Macroeconomic Time Series", *Journal of Business Economic Statistics 3*, p. 216-227, 1985

Havenner, A. and M. Aoki, "Econometrics and Linear System Theory in Multivariate Time Series Analysis", unpublished MS, 1987

Householder, A. S., *The Theory of Matrices in Numerical Analysis* Blaisdell Pub. Co. 1964

Kopp, P. E., *Martingales and Stochastic Integrals* Cambridge U P London 1984

Lindquist, A. and G. Picci, "On the Stochastic Realization Problem", *Siam Journal of Control and Optimization, 17*, p. 365-389, 1979

Nelson, C.R., and C.I. Plosser, "Trends and Random Walks in Economic Time Series", *Journal of Monetary Economics 10*, (1982), p. 135-162

Perron, P., and P. C. B. Phillips, "Does GNP have a Unit Root? A Revelation" *Cahier 2686*, C. R. D. E., 1986

Sargan, J. D., "Wages and prices in the United Kingdom: a Study in Econometric Methodology" in *Econometric Analysis for National Economic Planning*, P. E. Hurt et al (eds.), p. 28-63, Butterworths, London, 1964

Stock, J. H., and H. W. Watson, "Does GNP have a Unit Root?", *Economics Letters 22*, p. 147-151, 1986

Appendix

This appendix summarizes the model building procedure in Aoki (1987)

The singular value decomposition of the covariance matrix between a finite segment of the future realization and the past data of the time series is the basis of the recently developed method by Aoki. This matrix is a Hankel matrix in structure when the vectors are stacked as shown below:

$$H_K = E(y_t^+ y_{t-1}^{-\prime}) \tag{1}$$

where the stacked vectors are constructed as

$$y_t^{+\prime} = [y_t' \, y_{t+1}' \cdots y_{t+K-1}']$$

and

$$y_{t-1}^{-\prime} = [y_{t-1}' \, y_{t-1}' \cdots y_{t-K}']$$

When applied to a weakly stationary process, the method constructs a dynamic factor (state space innovation) model

$$x_{t+1} = Ax_t + Be_t \tag{2}$$

and

$$y_t = Cx_t + e_t,$$

where all the eigenvalues of the matrix $A-BC$ lie strictly inside the unit disk in the complex plane by construction, and e_t is weakly stationary innovation vector of the data, i.e., serially uncorrelated, $e_t = y_t - \hat{E}(y_t \mid y_{t-1}^-)$, where $\hat{E}$ denotes orthogonal projection, Aoki (1987a, chapter 9).

To be self-contained we summarize the procedure used to estimate A, B and C of (2) from the data set. See Aoki (1987a) for fuller account, and Havenner and Aoki (1987) for the relation of the estimation method to two-stage least squares, the instrumental variable method, and the (limited information) maximum likelihood method. The singular value decomposition is used to approximate the Hankel matrix,

$$H_K \cong U_n \Sigma_n V'_n$$

where n is the estimated numerical rank of H_K which is also the dimension of the state vector of the model (2). The ratios of the singular values $\sigma_i / \sigma_1 \quad i = 1, 2, \cdots$ determines n. Typically in series with significant trend (random walk) component, n is 1 or 2.

Relative magnitudes of the first few (counting from the largest) singular values are used, among other things, in selecting the dimension of the state vector of the model, n. In no case the dimension should be so large as to render the model non-minimal dimensional, i.e., the model should be observable and the dynamic matrix invertible, or the model should be both observable and controllable. See Aoki (1987a, Chapter 5). When the y_t series contains random walk components, the ratios of the second largest to the largest singular values and the third largest to the

first are usually very small, of the order of 10^{-2} or less, when the maximum lag length K is small so that Kp is about the same as n or slightly larger in the first step of the process, 2 or 3, for example.

In the innovation model (2), we note that y_t^+ is related to the state vector x_t by $y_t^+ = O\ x_t + K\ e_t^+$ where e_t^+ is uncorrelated with y_{t-1}^-, where O stacks $C, AC, A^2C \cdots$ and is called the observability matrix in the system theory. The matrix K is of no concern here. Therefore the Hankel matrix is equal to $H_K = O\ E(x_t\ y_{t-1}^{-\prime})$.

Its first p rows give the equation which involve the matrix C as

$$E(y_t\ y_{t-1}^{-\prime}) = C\ E\ (x_t\ y_{t-1}^{-\prime}). \tag{3}$$

Note that the left side is a $p \times Kp$ matrix made up of the sample covariance matrices $\Lambda_1, \Lambda_2, \cdots, \Lambda_k$, where $\Lambda_k = E(y_{t+k}\ y'_t)$, $k \geq 1$. Also the matrix A appears in

$$H^A = E\ (y_{t+1}^+\ y_{t-1}^{-\prime}) = OAE\ (x_t\ y_{t-1}^{-\prime}). \tag{4}$$

where the matrix H^A is the matrix H_K shifted left by p columns and the right-side p columns filled in by sample covariance matrices $\Lambda_{K+1}, \ldots \Lambda_{2K}$.

Fix the coordinate system by

$$O = U_n\ \Sigma_n^{1/2}.$$

(This puts the model (2) in the so-called balanced form. See Aoki (1987a, Sec 5.6). Assume that rank $O = dim\ x_t$ (This is the condition that (2) is observable and reachable, i.e., is the minimal realization. See Lindquist and Picci (1979) or Aoki (1987a, Sec 4.3) for further detail). By equating two factored expression of the Hankel matrix

$$H_K = U_n\ \Sigma_n^{1/2} \cdot \Sigma_n^{1/2} V'_n$$

$$= O\ E(x_t\ y_{t-1}^{-\prime}),$$

we can solve for $E(x_t\ y_{t-1}^{-\prime}) = \Omega = \Sigma_n^{1/2} V'_n$ in this coordinate system. Eq.(3) and (4) can now be solved for C and A to obtain their unique estimates by

$$\hat{C} = E(y_t\ y_{t-1}^{-\prime})\ \Omega^+ = E(y_t\ y_{t-1}^{-\prime})\ V_n\ \Sigma_n^{-1/2}$$

and

$$\hat{A} = O^+\ E(y_{t+1}^+\ y_{t-1}^{-\prime})\ \Omega^+ = \Sigma_n^{-1/2}\ U'_n\ E(y_{t+1}^+\ y_{t-1}^{-1})\ V_n\ \Sigma_n^{-1/2}.$$

Here + denotes pseudo-inverse. Associated with the model (2) is the Kalman filter with $z_{t|t-1} = \hat{E}(x_t\,|y_{t-1}^-)$ as its state vector, where $\hat{E}$ denotes the orthogonal projection. Since

$$\hat{E}(x_t\,|y_{t-1}^-) = E(x_t\ y_{t-1}^{-\prime})\ R^{-1}\ y_{t-1}^-$$

where

$$R = E(y_{t-1}^{-}\ y_{t-1}^{-}{}'),$$

the Kalman filter state vector summarizes the data vector y_{t-1}^{-} by

$$z_{t|t-1} = S\ y_{t-1}^{-}$$

where

$$S = E(x_t\ y_{t-1}^{-}{}')\ R^{-1} = \Omega\ R^{-1}.$$

From (2) and above $B\ e_t = \sigma_t$ where

$$\sigma_t = z_{t+1|t} - A\ z_{t|t-1}$$

$$= S\ y_t^{-} - A\ S\ y_{t-1}^{-}.$$

Finally B and $\Delta = cov\ e_t$ are estimated by noting the relation

$$B\ \Delta = E\ \sigma_t\ e'_t = \Gamma,$$

where $e_t = y_t - C\ z_{t|t-1}$. Finally the matrix Δ is related to the data covariance matrix by

$$\Delta = \Lambda_0 - C\ \Pi\ C', \tag{5}$$

where

$$\Pi = E(z_{t|t-1}\ z'_{t|t-1})$$

$$= A\ \Pi A' + B\ \Delta B'$$

$$= A\ \Pi A' + \Gamma(\Lambda_0 - C\ \Pi\ C')^{-1}\ \Gamma'. \tag{6}$$

Eq.(6) is solved for matrix Π using C and A of (3) and (4). Eq.(5) determines Δ and B is estimated as $G\ \Delta^{-1}$. (Alternatively $\Pi = \Omega\ R^{-1}\ \Omega'$, $\hat{B} = (\hat{M} - \hat{a}\ \Pi\ \hat{C}')\ \Delta^{-1}$ where $\hat{M} = \Sigma_n^{\frac{1}{2}}\ U_n'\ H^M$ and where H^M is the first block column $(Kp \times p)$ matrix of H_K.)