How to Build State Space Models for Nonstationary Time Series and How to Measure Random Walk Components

by

Masanao Aoki
University of California, Los Angeles
USA

UCLA Working Paper #438

March 1987

How to Build State Space Models for Nonstationary Time Series and How to Measure Random Walk Components †

by

Masanao Aoki *
University of California, Los Angeles
USA

Abstract

A two-step procedure of building state space models for vector-valued time series with trends is described. Then a new measure of random walk components in time series having unit root components is proposed to correct some undesirable features of those used in the literature. The quarterly US real GNP from 1947.1 to 1986.2 is found to contain less than 10% of random walk component.

---

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

† The results are supported in part by a grant from the National Science Foundation.

1. Introduction

Recently, modeling of nonstationary time series without removing deterministic trends has gained favors among time series modelers as a way of preserving information on longer-run dynamic structure in time series. A growing body of literature now exists both on finding and measuring contents of unit root or random walk components in a given set of time series, and on finding a co-integrating vector for them, i.e., a vector ν such that $\nu' y_t$ becomes weakly stationary even when y_t is not.

This paper presents a two-step procedure for constructing state space (dynamic factor) models of a set of nonstationary time series by extending the procedure in Aoki (1983, 1987). The procedure will determine co-integrating factors, when some of the components of the vector-valued time series contain common trends. 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 and applied in Aoki and Havenner (1986). This matrix is a Hankel matrix in structure when the vectors are stacked as shown below:

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

where the stacked vectors are constructed as

$$y_{t+1}^{+\prime} = [y_{t+1}' \, y_{t+2}' \cdots y_{t+K}']$$

and

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

When applied to a weakly stationary process, the method constructs a dynamic factor model

$$x_{t+1} = Ax_t + Be_t$$

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 (1987, chapter 9).

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 (1987, Chapter 5).

When y_t series contain 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^{-3} or less, when the maximum lag length K is small, 2 or 3, for example. This fact suggests a two-step modeling procedure discussed next.

2. Two-step Procedure

Build a small (one or two-dimensional) model which capture common random trends of the components of the data series as above, even though $\{y_t\}$ is not weakly stationary,

$$\tau_{t+1} = \rho\tau_t + g'u_t,$$

and

$$y_t = h\tau_t + u_t. \tag{1}$$

The variable τ_t is the common trend term if it is a scalar. When it is two-dimensional, for example, then there are two dynamic factors that explain common trends of all the components of the data vector. Since the residuals are highly correlated, unlike the modeling situations for weakly stationary series, u_t is not innovation vector. The singular value decomposition for the Hankel matrix constructed from the residuals is used next to build

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

and

$$u_t = Cx_t + e_t,$$

where e_t is the innovation vector for u_t in the wide sense. When jointly written, one sees that (1) and (2) display a recursive dynamic structure between the trend and cyclical dynamics. This structure has been shown in Aoki (1987, p.39) as a standard representation of dynamic systems containing some eigenvalues of unit magnitude. This representation is better than that in which each component of the data vector series is separately detrended (either by deterministic transformation such as taking differences of the logarithms or by some random detrending schemes such as Kitagawa's procedure, see Akaike (1985), because the interactions between the longer-run dynamics and shorter-run dynamics are not allowed in the latter representation.

Rewrite (1) as

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

where $\hat{\rho}$ is equal to $\rho - g'h$. If $\sum_{k=0}^{\infty} \hat{\rho}^k \lambda_{t+k}$ converges, where λ_l is the covariance matrix of the data vector, then the covariance matrix of τ_t is well-defined. *

Similarly, rewrite (2) as

$$x_{t+1} = (A - BC)x_t + Bu_t$$

Then, the covariance matrix $\pi = cov x_t$ is well defined, and subscript t is dropped from π because x process is weakly stationary, if all the eigenvalues of $A-BC$ lie strictly inside the unit disk. Let $F = A-BC$. The matrix F is the dynamic matrix in the Kalman filter (2). To see the importance of this condition, consider solving the Riccati equation for the matrix π

$$\pi = A\pi A' + B(\lambda_0 - C\pi C')^{-1}B',$$

where $cov\ e_t = \lambda_0 - C\ \Pi\ C'$ and $B = (M - A\ \Pi\ C')(cov\ e_t)^{-1}$ by an iterative procedure. Supposing that π exists, denote $\pi_k - \pi$ by z_k. Then

$$vec z_{k+1} = (F \otimes F) vec z_k,$$

Therefore, the equation converges as k is increased if and only if all the eigenvalues of F lie inside the unit disk as claimed.

Eq (1) shows that any vector v which is in the null space of h' is a co-integrating vector since $v'y_t = v'u_t$ which is weakly stationary.

The above consideration 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 ρ in (1) is one, provided $\hat{\rho}$ is less than one in magnitude. A sufficient condition is that the unit root is a controllable eigenvalue of the model. See Aoki (1986) on detained demonstration of this claim.

3. How to Measure Random Walk Components

The findings on the percentages of random walks in the US GNP fall into two classes: Nelson and Plosser (1982) and Campbell and Mankiw (1986) found large components, while Watson (1986), Clark (1986) and Cochrane (1986) found small percentages of random walk components. Some studies fit a model $y_t = \rho\, y_{t-1} + u_t$, where the disturbance term is mean zero and modeled by AR (p), rather rarely as ARMA (p, q) or using the Wold decomposition. Cochrane attributes these divergent results to certain errors in neglecting sums of large numbers of small correlation coefficients. According to Cochrane those using low values of p and q found larger random walk component than

---

* For example, if $\lambda_{t+k} = \lambda_t + k\sigma^2$, then $\Sigma_k k\hat{\rho}^k$ converges if the magnitude of $\hat{\rho}$ is less than one.

those with higher values of p.

This section shows that the measured used by Nelson and Plosser, and Cochrane's have some undersirable properties and proposes an alternative measure for the random walk components. The state space representation of Aoki (1987) used in this paper also avoids a source of errors emphasized by Cochrane.

Cochrane (1986) proposed the ratio

$$\kappa = \lim_{k \to \infty} [var(y_t - y_{t-k}) / k] / var(y_t - y_{t-1}) \tag{3}$$

as a measure of random walk components in $\{y_t\}$. The numerator approaches the variance of the random walk σ_e^2 in the decomposition of $y_t = y_{1t} + y_{2t}$ where $y_{1t} - y_{1t-1} = e_t$, and e_t is i.i.d., since the variance of the random walk components grow linearly with lag k.

This ratio can be larger than one if Δy_{1t} and Δy_{2t} are negatively correlated where Δ denotes first difference. This seems undesirable. Even when the limit stays below 1, this limit then overstates the random walk component. As an example consider

$$\begin{aligned} y_t - y_{t-1} &= e_t + a_1 e_{t-1} \\ &= (1 + a_1)e_t - a_1(e_t - e_{t-1}), \end{aligned} \tag{4}$$

where e_t is i.i.d. with mean zero and variance σ_e^2. As the second line above suggests, the random walk component is $y_{1t} = [(1 + a_1)] / (1 - L)]\, e_t$ and $y_{2t} = -a_1\, e_t$ is a weakly stationary component y. The limit of the numerator of (3) is equal to $(1 + a_1)^2\, \sigma_e^2$ while the denominator is equal to $(1 + a_1^2)\sigma_e^2$. The ratio is $\kappa = (1 + a_1)^2 / (1 + a_1^2) > 1$ if $a_1 > 0$.

To remedy this undesirable feature, we first decompose y_t as $y_{1t} + y_{2t}$ where y_{1t} is a pure random walk component and y_{2t} is weakly stationary and define

$$\theta = var\, \Delta\, y_{1t}\ /\ [var\, \Delta\, y_{1t} + var\, \Delta\, \tilde{y}_{2t}]$$

where

$$var\, \Delta\, y_{1t} = lim\ var(y_t - y_{t-k}) / k$$

and

$$\Delta\, \tilde{y}_{2t} = \Delta\, y_{2t} - \hat{E}\ (\Delta\, y_{2t} \mid \Delta\, y_{1t})$$

where $\hat{E}$ denotes the orthogonal projection. In the example (4), $var\, \Delta\, y_{1t} = (1 + a_1)^2\sigma_e^2$ and $var\, \Delta\, \tilde{y}_{2t} = a_1^2\, \sigma_e^2$ and hence $\theta = (1 + a_1)^2 / (1 + a_1)^2 + a_1^2 < 1$. This ratio remains less than 1 even when $a_1 > 0$.

More generally, consider $y_t - y_{t-1} = d_t$ where $d_t = A(L)e_t$, e_t is mean zero, i.i.d., and

$$A(L) = \Sigma_{j=o}^{\infty} a_j L^j \ , \quad a_1 = 1 \ , \quad \Sigma_0^{\infty} a_j^2 < \infty \ . \tag{5}$$

Clearly $var\,\Delta y_t = var\, d_t = \Sigma_j a_j^2 < \infty$. By writing $A(L)$ as $A(1) + A(L) - A(1)$, we can identify y_{1t} by $\Delta y_{1t} = A(1)e_t$ and $\Delta y_{2t} = [A(L) - A(1)]e_t$.

From now on we consider only y_t with rational spectral density functions. Let $\mu_0 = var\, d_t$ and $\mu_l = E(d_t d_{t-l})$. Then

$$S(z) = \Sigma_{-\infty}^{\infty} \mu_l z^{-l} = \mu_0 + 2\Sigma \mu_l z^{-l}$$

is the spectral density function of d_t because $\mu_{-l} = \mu_l$ when d_t is scalar-valued. Because d_t has a rational spectral density function, it has a finite parametrization (Kronecker's *lemma*)

$$x_{t+1} = Fx_t + ge_t ,$$

$$d_t = h'x_t + e_t ,$$

with some vector x_t, which is the minimal realization of d_t. (See Aoki (1987) or Lindquist and Picci (1979)). Cochrane evaluates $S(1)$ by the infinite sum $\mu_0 + 2\sum_l \mu_l$.

By the spectral decomposition theorem (Whittle (1963, p. 26) or Aoki (1987, p. 69)),

$$S(z) = W(z)\, W'(1/z)\sigma_e^2$$

where $W(z) = I + h'(zI - F)^{-1}g$. Thus

$$S(1) = W(1)^2\sigma_e^2,$$

since $a_j = c'F^{j-1}g$, $j \geq 1$, i.e., $W(1) = A(1)$. Writing d_t as

$$d_t = W(1)e_t + h'[(zI - F)^{-1} - (I - F)^{-1}]ge_t$$

$$= W(1)e_t - (z - 1)H(z)e_t$$

where $H(z) = h'(zI - F)^{-1}(I - F)^{-1}ge_t$, we see that $\Delta y_t = \Delta y_{1t} + \Delta y_{2t}$ where $\Delta y_{1t} = W(1)e_t$ is the pure random walk component and $y_{2t} = -H(z)e_t$. The component y_{2t} has a state space representation

$$s_{t+1} = Fs_t + \psi e_t ,$$

$$y_{2t} = h' s_t + e_t,$$

where

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

Hence

$$\Delta \tilde{y}_{2t} = c' F s_{t-1} + (c' \psi - 1) e_{t-1}$$

where s_t and e_t are chosen to be uncorrelated. Thus

$$\theta = W(1)^2 / [W(1)^2 + (h'\psi - 1)^2 + h' FSF' h]$$

where $S = FSF' + \psi\psi'$ or using $vec\, S = (I - F \otimes F)^{-1} (\psi \otimes \psi)$, we can write $h'FSF'h = (c'F \otimes h'F)(I - F \otimes F)^{-1} (\psi \otimes \psi)$.

Here, the indicated inverse exists, because the constraint $\Sigma_j a_j^2 < \infty$ implies that $\Sigma_j F^i gg'(F')^j < \infty$ which in turn implies that all eigenvalues of F lie strictly inside the unit disk. (Alternatively this sum can be expressed as the positive definite solution of $X = FXF' + gg'$, which exists if and only if $I - F \otimes F$ is invertible.)

Nelson and Plosser (1982) suggests the ratio $\sigma_e^2 / \sigma_\delta^2$ in the model $y_t = u_t + v_t$, where $\Delta u_t = \mu + A(L) e_t$, and $v_t = B(L)\delta_t$. Here

$$\Delta y_t = A(L) e_t + \Delta u_t$$

$$= A(1) e_t + (1 - L) B(L) \delta_t + [A(L) - A(1)] e_t .$$

If v_t is weakly stationary, $\Delta y_{1t} = A(1) e_t$. Let $d_t = A(L) e_t$ as before. Now

$$\Delta \tilde{y}_{2t} = c' \phi x_{t-1} + h' F s_{t-1} + \delta_t + (c'b - 1)\delta_{t-1} - h'\psi e_{t-1}$$

where $A(z) = 1 + h'(zI - F)^{-1} g$ and $B(z) = 1 + c'(zI - \phi)^{-1} b$. The ratio $\mu = \sigma_\delta^2 / \sigma_e^2$ appears in θ as

$$\theta = A(1)^2 / \{A(1)^2 + [1 + (c'b - 1)^2]\, \mu + (h'\psi)^2 + c'\Phi c + h' S h \mu\}$$

where $\Phi = F\, \Phi\, F' + \psi\psi'$ and $S = \phi\, S\, \phi' + bb'$. The same ratio μ gives rise to different θ depending on particular $A(\cdot)$ and $B(\cdot)$. The same objection was raised by Cochrane as well.

Finally consider a more general vector-valued y_t with scalar unit root component, i.e., y_t has a single trend term. This is a special case considered in Aoki (1987), p. 39) and has a standard representation *

$$\begin{bmatrix}\tau_{t+1}\\ x_{t+1}\end{bmatrix} = \begin{bmatrix}1 & g'C\\ 0 & F\end{bmatrix}\begin{bmatrix}\tau_t\\ x_t\end{bmatrix} + \begin{bmatrix}g'\\ B\end{bmatrix} e_t\,,\quad y_t = (h \quad C)\begin{bmatrix}\tau_t\\ s_t\end{bmatrix} + e_t\,. \tag{6}$$

The transfer function is $[I + hg'/(z-1)]\,[I + C(zI - F)^{-1}B]$, so that $y_t - y_{t-1}$ has the transfer function

$$\Theta(z) = ((z-1)L + hg')\,(I + C(zI - F)^{-1}B).$$

The random walk component is then $\Delta y_{1t} = \Theta(1)e_t$ and

$$y_{2t} = I + c(zI - F)^{-1}B - hg'C(zI - F)^{-1}(I - F)^{-1}B$$

$$= I + [(I - hg')C - CF]\,(zI - F)^{-1}B$$

which has the state space representation

$$f_{t+1} = Ff_t + (I - F)^{-1}Be_t\,,\quad y_{2t} = [(I - hg')c - cF]\,f_t + e_t\,.$$

Thus

$$var\,\Delta\,\tilde{y}_{2t} = tr\,[\{I - \nu(I - F)^{-1}B\}\,\Delta\,\{I - \nu\,(I - F)^{-1}B\}' + \nu F\,\Phi F'\nu']$$

where

$$\nu = (I - hg')C - CF\,,$$

and

$$\Phi = F\,\Phi\,F' + (I - F)^{-1}B\,\Delta\,B'(I - F)^{-1}\,,\quad \Delta = cov\,e_t\,.$$

Here

$$\nu F\,\Phi\,F'\nu' = (\nu F \otimes \nu F)\,(I - F \otimes F)^{-1}[(I - F)^{-1}B \otimes (I - F)^{-1}B]\,vec\,\Delta\,.$$

Example: The US quarterly real GNP from the first quarter 1947 has been modeled by (6). In total 158 data points are used. After fitting one-dimensional trend dynamics, several models have been tied to model the residual series. Denote the ratio by $\theta(K, n)$ to indicate the model with K number of lags and the n dimensional state vec-

---

* The fact that the same e_t appears in both equations is no restriction. The covariance matrix, with disturbances ξ_t in the first equation and η_t in the second, can be factored as

$$cov\begin{bmatrix}\xi\\ \eta\end{bmatrix} = \begin{bmatrix}H\\ I\end{bmatrix}\Delta\,(H' \quad I)$$

where $\Delta = cov\,e_t$ where $H = \begin{bmatrix}g'\\ B\end{bmatrix}$. See Aoki (1987, p.67) tor is used. It is found that $\theta_{6,4} = 4.4\%$, $\theta_{3,4} = 5.5\%$. Thus we conclude about 10% of the random walk components are contained in the US real GNP fluctuations from 1947 to 1986.

4. Examples

Two examples of this procedure, one on the univariate quarterly real US GNP from the first quarter of 1947 to the second quarter of 1986, and the second on the bivariate monthly series for the US M1 and CPI from January 1974 to January 1986 are described in a separate paper. Figure 1 plots the residual series of the GNP and Figure 2 shows those for the M1 and CPI.

5. Conclusions

The two-step procedure of this paper improved on the separate detrending of individual series. The method fails only when there is a unit root which is not controllable. When this happens, it indicateds that the dimension of the common trend vector is too large. By reducing it, the minimal state space representation (3) is achieved.

As in the individual detrending scheme of Kitagawa, there is room for some trade-off between smoother trends v.s. more complex short run dynamics. If the lag length K in the first step is chosen large covering one year or more of the data span, say, then the trend dynamics tends to become complex, leaving smoother residuals, while a small K and scalar dynamics for the trend leaves more complex residual dynamics. A large K in the first step also reduces the ratio of the largest to the second largest singular values.

References

Akaike, H., et al., *TIMSAC-84, Part 1,* Inst. Stat. Math., Tokyo, 1985

Aoki, M., *Notes on Economic Time Series Analysis: System Theoretic Approach,* Springer-Verlag, Heidelberg, 1983

Aoki, M., "On Unit Root in Time Series Modeling", unpublished MS

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

Aoki, M. and Havenner, "Approximate State Space Modeling of Vector-Valued Macroeconomic Time Series: Some Cross-Country Comparisons", *J. E. D. C., 10,* p. 145-155, 1986

Campbell, J. Y., and N. G. Mankiw, "Are Output Fluctuations Transitory", NBER Meeting on Economics 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

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), pp. 135-162

Watson, M. W., "Univariate Detrending Methods with Stochastic Trends" *Journal of Monetary Economics, 18,* p. 45-75, 1986

Whittle, P., *Prediction and Regulation,* Van Nostrand, 1963

residuals of ln rgnp by m2.1 for 1982 constant dollar data

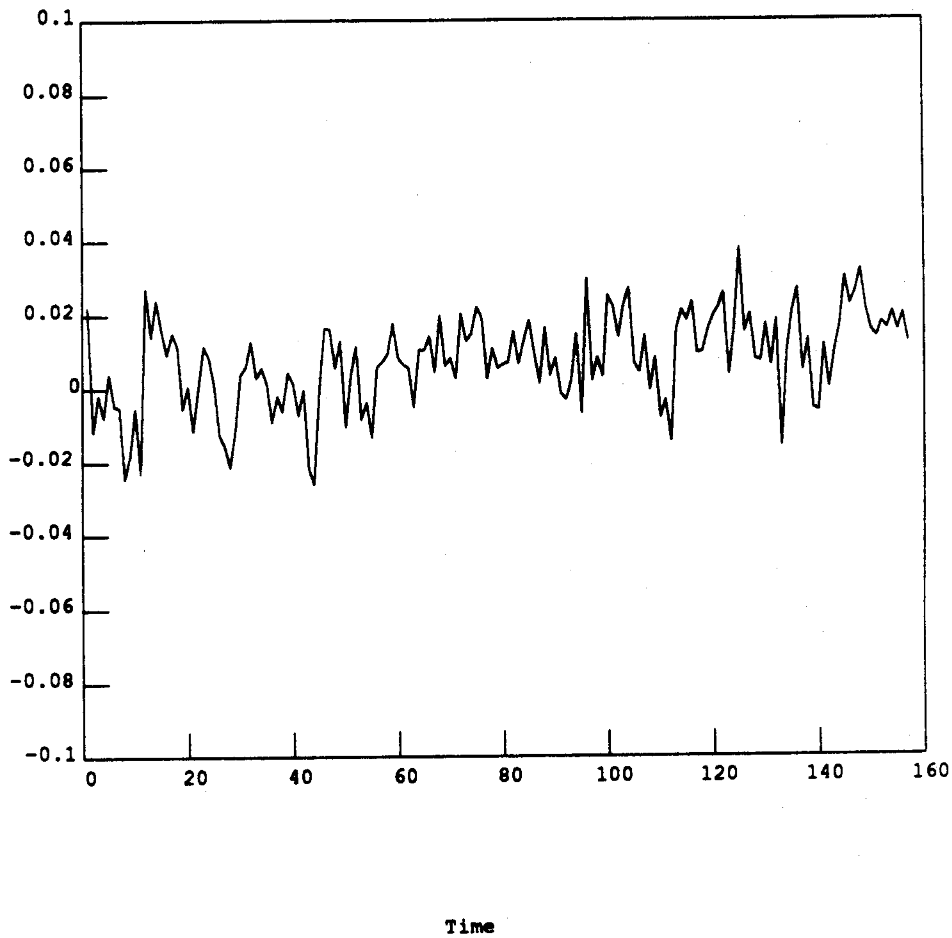

Figure 1

Residuals of ln GNP at 1982 constant dollar by a scalar dynamics (K = 2, n = 1)

residuals of ln m1 cpi by mcpi1-2

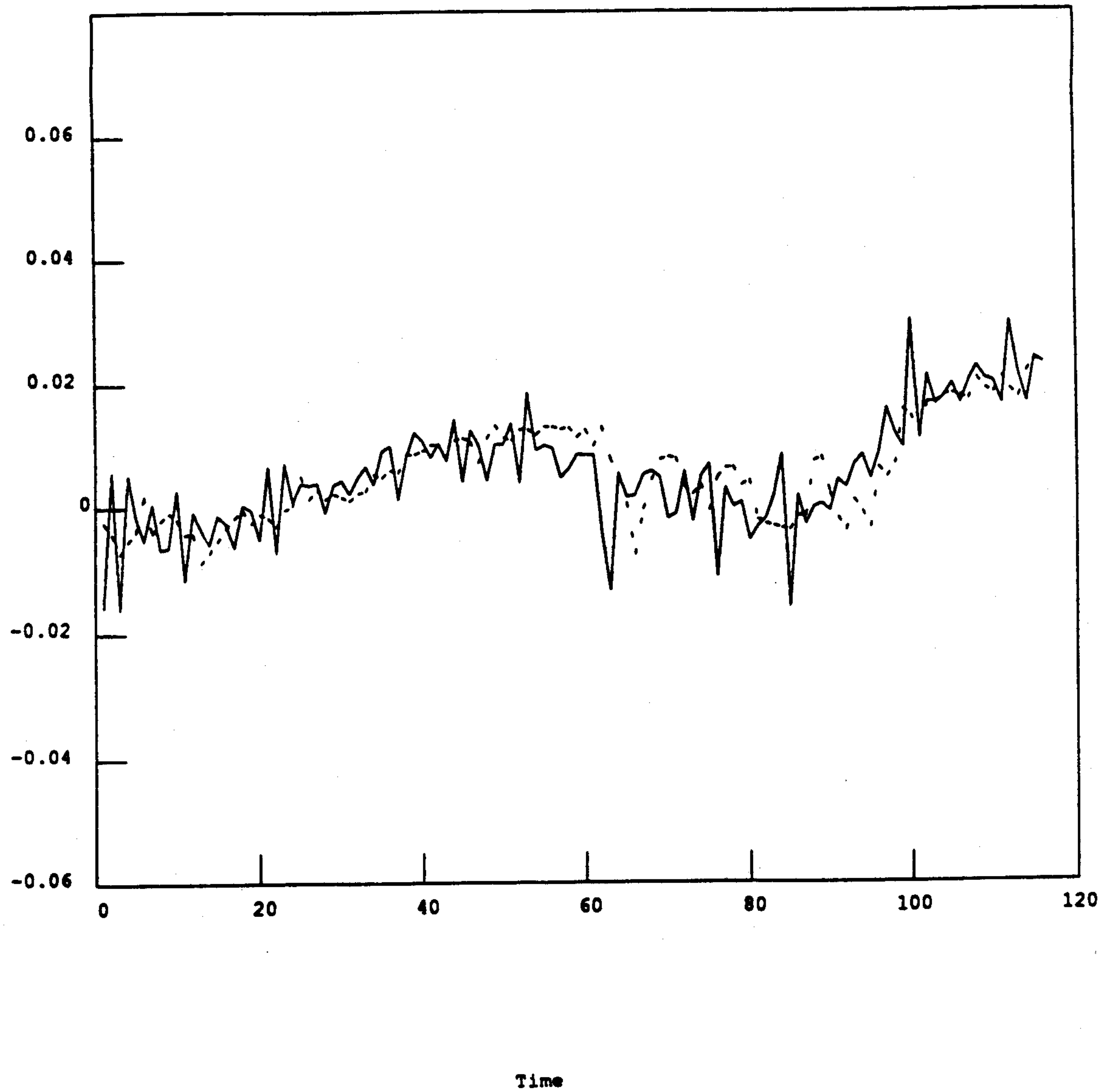

Figure 2

Residuals of ln M1 and ln CPI by 2-dim trend dynamics. (K = 1, n = 2)