Modeling Time-Varying Dynamical Systems

CARLO GRILLENZONI*

A global methodology of identification, estimation and forecasting of transfer function (Box–Jenkins) models with deterministically varying parameters is provided. First, properties of stability and forecasting algorithms are investigated by means of Markovian representations and methods of solution of nonstationary difference equations. Next, the degree of the polynomials of the system is specified with typical off-line methods, and the shape of the coefficients (parameter functions) is identified by means of recursive (on-line) algorithms. Finally, the identified parameter functions are inserted in the model and their coefficients are estimated (off-line) on the original data by means of pseudolinear regression techniques.

KEY WORDS: Nonstationary forecasting; Pseudolinear regression; Recursive estimation; Transfer functions; Varying parameters.

1. INTRODUCTION

Nonstationary time series models with deterministically varying paremeters were discussed by Farley, Hinich, and McGuire (1975), Mèlard and Kihem (1981), Bittanti (1986), and others. With respect to the random-coefficient approach, the advantage of the deterministic one is that the resulting schemes use the same inferential framework as the stationary models; moreover, their building procedures can avoid the need for a priori specifications. On the other hand, the idea of "nonstochastic evolution" is synonymous with slow and systematic change, so the fields of major application are social–economic data and quasi-deterministic processes [typical examples are communication systems with periodic parameters; see Bittanti (1986)].

Consistent with the data-based philosophy of Box–Jenkins models, this article attempts to provide a complete procedure of identification, estimation, and forecasting of transfer-function models with deterministically varying parameters (VTF). Given the complexity of the task and the kind of applications to which the method is addressed, some structural restrictions are needed, such as fixed orders for the model and continuous functions for the parameters. This article is particularly concerned with cubic polynomials of time. The global model-building strategy is summarized by Figure 1.

I now describe the crucial phases of the strategy. Step 1 identifies the orders of the system with typical off-line methods (Box and Jenkins 1970; Poskitt 1989), which measure average tendencies on the sample. Step 2 identifies the shape of the parameters in two stages: In step 2A the raw-time path is observed by estimating the model recursively (Ljung and Söderström 1983), and in step 2B the functional form is recovered by interpolating the recursive estimates with time functions. Finally, in step 3 the identified parameter functions are included in the model, and their coefficients are reestimated (off-line) on original data. Notice that at every step the choice between alternative techniques (in italics) is possible; this makes the procedure open. In particular, it includes the stochastic-coefficient (Bayesian forecasting) approach, which simply results from stopping the model building at step 2A.

The need to proceed beyond step 2A arises because optimal design and statistical analysis of recursive estimators are not possible without precise a priori information on the evolution of the system (see Benveniste 1987). Instead, as we see, step 3 can actually free the modeling from dependence on the particular on-line algorithm used and the arbitrary priors it needs. Moreover, forecasting with step 5 has the advantage of making predictions on the parameters as well, but without changing the inferential framework of the stationary models. Step 4 closes the entire scheme, allowing for revisions to previous errors and nonoptimal specifications.

Methodologically, the outlined strategy requires the examination of the following issues: (a) solution and analysis of the stability of stochastic difference equations with rational polynomials and time-varying coefficients; (b) analysis of the tracking ability (in a statistical sense) of recursive estimators under the assumption of deterministic evolution; and (c) derivation of algorithms able to solve highly nonlinear estimation problems. These points are developed analytically in Sections 2, 3, and 4; throughout, an example based on real economic data is developed to illustrate the procedure.

2. REPRESENTATION

In this section I investigate the structure of the transfer-function model with deterministically varying parameters (VTF), its solution, the related properties of stability, and the forecasting algorithms. The dynamical system considered is the following:

$$Y_t = \frac{\omega_{0_t} + \omega_{1_t}B + \cdots + \omega_{s_t}B^s}{1 - \delta_{1_t}B - \cdots - \delta_{r_t}B^r} X_{t-b} + (1 - B)^{-d}\frac{1 + \theta_{1_t}B + \cdots + \theta_{q_t}B^q}{1 - \phi_{1_t}B - \cdots - \phi_{p_t}B^p} a_t \sim IN(0, \sigma^2), \tag{2.1}$$

where $\{Y_t\}$ is the system output, $\{X_t\}$ is the exogenous input, and $\{a_t\}$ is the casual disturbance. With (B, b, d)

* Carlo Grillenzoni is Lecturer in Statistics, Department of Economics, University of Modena, Via Giardini 454, 41100 Modena, Italy. The author is grateful to the associate editor and a referee for the constructive reviewing process.

Reprinted from: © 1990 American Statistical Association
Journal of the American Statistical Association
June 1990, Vol. 85, No. 410, Theory and Methods

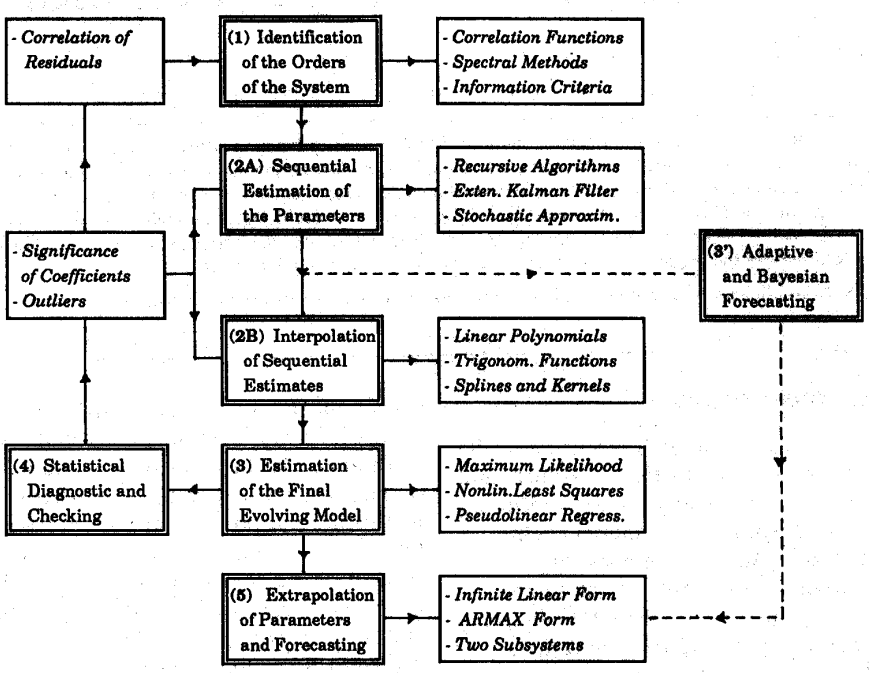

Figure 1. Summary of the Global Model-Building Strategy.

we denote the back-shift operator, the delay and the degree of nonstationarity in mean of input and output [for details, see Box and Jenkins (1970, part III)]. Nonstationarity in covariance is induced by the time-dependent coefficients.

The class of VTF models is very large and practically intractable. I emphasize the features of the aforementioned representation, which enables the development of a strategy of modeling. The orders (r, s, b), (p, d, q) are fixed, and the number of parameters $(r + s + 1 + p + q)$ is minimum. The parameters are continuous polynomial functions for t real (this article is concerned with $\beta_{i_t} = \sum_{j=0}^{3} \alpha_{ij} t^j$, for all i). The process $\{a_t\}$ is white noise and stationary. The first condition permits the identification of the orders with off-line methods, the second involves the consistent application of recursive algorithms in identifying the shape of the parameters, and the third means that the resulting model must represent all of the possible information.

Stability. The definition of properties of stability is crucial for estimation and forecasting purposes. It cannot be developed inside the stationary framework, however. To simplify the analysis, we substitute $(1 - B)^d Y_t = y_t$ and $(1 - B)^d X_t = x_t$ in (2.1) and split the resulting model into two subsystems $y_t = m_t + n_t$, with $m_t = \omega_t(B)x_{t-b}/\delta_t(B)$ and $n_t = \theta_t(B)a_t/\phi_t(B)$. Now, using a vector notation, we rewrite (2.1) in pseudolinear form:

$$m_t = \boldsymbol{\delta}_t' \mathbf{m}_{t-1} + \boldsymbol{\omega}_t' \mathbf{x}_{t-b},$$

$$n_t = \boldsymbol{\phi}_t' \mathbf{n}_{t-1} + \boldsymbol{\theta}_t' \mathbf{a}_{t-1} + a_t, \tag{2.2}$$

$$y_t = \boldsymbol{\beta}_t' \mathbf{z}_t(\boldsymbol{\beta}) + a_t, \tag{2.3}$$

and

$$\boldsymbol{\beta}_t' = [\boldsymbol{\delta}_t', \boldsymbol{\omega}_t', \boldsymbol{\phi}_t', \boldsymbol{\theta}_t'],$$

$$\mathbf{z}_t'(\boldsymbol{\beta}) = [\mathbf{m}_{t-1}', \mathbf{x}_{t-b}', \mathbf{n}_{t-1}', \mathbf{a}_{t-1}'],$$

where $\boldsymbol{\delta}_t' = (\delta_{1_t} \cdots \delta_{r_t})$, $\mathbf{m}_{t-1}' = (m_{t-1} \cdots m_{t-r})$, and so forth; $\boldsymbol{\beta}_t$ is the vector of time-varying parameters, and $\mathbf{z}_t(\cdot)$ is the vector of pseudolinear regressors (Solo 1978).

The stability of the VTF model then coincides with the stability of the two subsystems $m_t(x_t)$ and $n_t(a_t)$. To simplify the analysis further, we henceforth work with their equivalent Markovian representations. For $\{m_t\}$ this means

$$\begin{bmatrix} m_t \\ m_{t-1} \\ \vdots \\ m_{t-r+1} \end{bmatrix} = \begin{bmatrix} \delta_{1_t} \cdots \delta_{r_t} \\ \mathbf{I}_{r-1} \end{bmatrix} \begin{bmatrix} m_{t-1} \\ m_{t-2} \\ \vdots \\ m_{t-r} \end{bmatrix} + \begin{bmatrix} \omega_{0_t} \cdots \omega_{s_t} \\ \mathbf{O}_s \end{bmatrix} \begin{bmatrix} x_{t-b} \\ x_{t-b-1} \\ \vdots \\ x_{t-b-s} \end{bmatrix};$$

that is, $\mathbf{m}_t = \boldsymbol{\Delta}_t \mathbf{m}_{t-1} + \boldsymbol{\Omega}_t \mathbf{x}_{t-b}$. Now, solving recursively for $\mathbf{m}_t$ $(t > 0)$, with $b = 0$ and initial condition $\mathbf{m}_0 = \mathbf{0}$, we get

$$\mathbf{m}_1 = \boldsymbol{\Omega}_1 \mathbf{x}_1,$$

$$\mathbf{m}_2 = \boldsymbol{\Delta}_2 \boldsymbol{\Omega}_1 \mathbf{x}_1 + \boldsymbol{\Omega}_2 \mathbf{x}_2,$$

and

$$\mathbf{m}_3 = \boldsymbol{\Delta}_3 \boldsymbol{\Delta}_2 \boldsymbol{\Omega}_1 \mathbf{x}_1 + \boldsymbol{\Delta}_3 \boldsymbol{\Omega}_2 \mathbf{x}_2 + \boldsymbol{\Omega}_3 \mathbf{x}_3,$$

etc.; so the limiting expression depends on infinite products of lagged $\boldsymbol{\Delta}_t$ terms and one $\boldsymbol{\Omega}_t$. Sufficient conditions for the asymptotic stability of the solution $\mathbf{m}_t = f(\mathbf{x}_t)$ are then given by the following.

Condition 1. $\{\omega_t(z)\}$ has bounded coefficients everywhere $t > 0$.

Condition 2. $\{\delta_t(z)\}$ has roots outside the unit circle nearly everywhere (i.e., with the possible exception of a finite number of $t > 0$).

The statement holds, since the eigenvalues $\{\lambda_{i_t}\}$ of $\{\Delta_t\}$ are the inverse of the roots of $\{\delta_t(z)\}$ (see Fuller 1976); so

$$\left|\prod_{\tau=2}^{t} \Delta_\tau\right| = \prod_{\tau=2}^{t} |\mathbf{P}_\tau \Lambda_\tau \mathbf{P}_\tau^{-1}| = \prod_{\tau=2}^{t} \prod_{i=1}^{r} \lambda_{i_\tau} \xrightarrow[t]{} 0.$$

Now, assuming bounded roots (i.e., nonzero eigenvalues), the terms Δ_τ are nonsingular and the aforementioned is true only if each element of the matrix $\prod_\tau^t \Delta_\tau$ tends to 0.

With reference to the second subsystem $\{n_t\}$ [equivalent to an autoregressive moving average (ARMA) process], the associated Markovian representation coincides with a vector ARMA (1, 1) model

$$\begin{bmatrix} n_t \\ n_{t-1} \\ \vdots \\ n_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_{1_t} \cdots \phi_{p_t} \\ \mathbf{I}_{p-1} \end{bmatrix} \begin{bmatrix} n_{t-1} \\ n_{t-2} \\ \vdots \\ n_{t-p} \end{bmatrix} + \begin{bmatrix} \theta_{1_t} \cdots \theta_{q_t} \\ -\mathbf{I}_{q-1} \end{bmatrix} \begin{bmatrix} a_{t-1} \\ a_{t-2} \\ \vdots \\ a_{t-q} \end{bmatrix} + \begin{bmatrix} a_t \\ a_{t-1} \\ \vdots \\ a_{t-q+1} \end{bmatrix};$$

that is, $\mathbf{n}_t = \mathbf{\Phi}_t \mathbf{n}_{t-1} + \mathbf{\Theta}_t \mathbf{a}_{t-1} + \mathbf{a}_t$. Again, solving recursively for $\mathbf{n}_t$, $\mathbf{a}_t$ $(t > 0)$ with initial conditions $\mathbf{n}_0 = \mathbf{a}_0 = \mathbf{0}$, we have

$$\mathbf{n}_1 = \mathbf{a}_1,$$

$$\mathbf{n}_2 = (\mathbf{\Phi}_2 + \mathbf{\Theta}_2)\mathbf{a}_1 + \mathbf{a}_2,$$

and

$$\mathbf{a}_3 = -\mathbf{\Theta}_3(\mathbf{\Theta}_2 + \mathbf{\Phi}_2)\mathbf{n}_1 - (\mathbf{\Theta}_3 + \mathbf{\Phi}_3)\mathbf{n}_2 + \mathbf{n}_3.$$

Reiterating the previous reasoning, we may conclude that the sufficient conditions for the global stability of the system are completed by the following requirement.

Condition 3. $\{\phi_t(z)\}$ and $\{\theta_t(z)\}$ have roots outside the unit circle nearly everywhere. It is easy to show that in the case of nonperiodic parameter functions these are necessary.

Under the aforementioned conditions (1–3), the asymptotic Markovian solutions of the subsystems are bounded and independent of any initiation $\mathbf{m}_0$, $\mathbf{n}_0 < \infty$:

$$\mathbf{m}_t = \sum_{i=0}^{\infty} \mathbf{V}_{i_t} \mathbf{x}_{t-b-i}, \qquad \mathbf{n}_t = \sum_{j=1}^{\infty} \mathbf{\Psi}_{j_t} \mathbf{a}_{t-j} + \mathbf{a}_t, \quad (2.4)$$

where $\mathbf{V}_{0_t} = \mathbf{\Omega}_t$, $\mathbf{V}_{1_t} = \Delta_t \mathbf{\Omega}_{t-1}$, $\mathbf{V}_{2_t} = \Delta_t \Delta_{t-1} \mathbf{\Omega}_{t-2}$, $\mathbf{V}_{3_t} = \Delta_t \Delta_{t-1} \Delta_{t-2} \mathbf{\Omega}_{t-3}$; $\mathbf{\Psi}_{0_t} = \mathbf{I}$, $\mathbf{\Psi}_{1_t} = (\mathbf{\Phi}_t + \mathbf{\Theta}_t)$, $\mathbf{\Psi}_{2_t} = \mathbf{\Phi}_t(\mathbf{\Phi}_{t-1} + \mathbf{\Theta}_{t-1})$, and $\mathbf{\Psi}_{3_t} = \mathbf{\Phi}_t \mathbf{\Phi}_{t-1}(\mathbf{\Phi}_{t-2} + \mathbf{\Theta}_{t-2})$. For every time instant there are different sequences of matrix weights, and thus a saving of calculations is in order when t change; here, simple one-step-ahead updating formulas are $\mathbf{V}_{i_{t+1}} = \Delta_{t+1} \mathbf{V}_{i-1_t}$ and $\mathbf{\Psi}_{jt+1} = \mathbf{\Phi}_{t+1} \mathbf{\Psi}_{j-1_t}$. Notice that when the orders of the system are different, $r \neq s + 1$ and $p \neq q$, the products $\Delta_i \mathbf{\Omega}_j$ and the sums $\mathbf{\Phi}_t + \mathbf{\Theta}_t$ can still be carried out by adjusting the dimension of the matrices with rows and columns of zeros on the bottom and the right.

As a result of the analysis conducted so far, we finally recover from (2.4) the scalar solution for the model (2.1), namely

$$y_t = \sum_{i=0}^{\infty} v_{i_t} x_{t-b-i} + \sum_{j=1}^{\infty} \psi_{j_t} a_{t-j} + a_t \quad (2.5)$$

and

$$v_{i_t} = \mathrm{tr}[\mathbf{V}_{i_t}], \qquad \psi_{j_t} = \mathrm{tr}[\mathbf{\Psi}_{j_t}], \qquad j \geq 1 \quad \forall\, t. \quad (2.6)$$

The relationship (2.6) follows by induction as in the stationary case (Fuller, 1976), and it actually proves the stability of (2.5) under Conditions 1–3. Its computational and analytical importance for the evolving models is fundamental. Indeed, the sequences $\{v_{i_t}, \psi_{j_t}\}$ cannot be obtained, as in stationarity, by expanding in Taylor series the rational polynomials $\omega_t(B)/\delta_t(B)$ and $\theta_t(B)/\phi_t(B)$ for each t. These calculations serve at most as approximations in the case of smooth evolution and roots of $\delta_t(B)$ and $\phi_t(B)$ well outside the unit circle.

Forecasting. Previous results have a direct utilization in terms of forecasting algorithms. The methods proposed by Box and Jenkins (1970), based on the ARMAX form $\delta_t(B)\phi_t(B)y_t = \omega_t(B)\phi_t(B)x_{t-b} + \theta_t(B)\delta_t(B)a_t$ or the infinite linear form (2.5), present, in the nonstationary context, several problems of calculation. In fact, at every step it would be necessary to execute three products of polynomials or compute two sequences of matrix weights.

To avoid these complications, we may refer to the decomposition of the VTF into two independent subsystems. In prediction this means $\hat{y}_t(l) = \hat{m}_t(l) + \hat{n}_t(l)$, namely

$$\hat{y}_t(l) = E[y_{t+l} \mid x_{t-i}, a_{t-j}] = E[y_{t+l} \mid x_{t-i}] + E[y_{t+l} \mid a_{t-j}], \quad (2.7)$$

$$\hat{m}_t(l) = \boldsymbol{\delta}'_{t+l} \hat{\mathbf{m}}_{t+l-1} + \boldsymbol{\omega}'_{t+l} E[\mathbf{x}_{t+l-b} \mid x_{t-i}; i \geq 0], \quad (2.8)$$

and

$$\hat{n}_t(l) = \boldsymbol{\phi}'_{t+l} \hat{\mathbf{n}}_{t+l-1} + \boldsymbol{\theta}'_{t+l} E[\mathbf{a}_{t+l-1} \mid a_{t-j}; j \geq 0]. \quad (2.9)$$

(2.8) and (2.9) are derived from (2.2), and together provide a very practical algorithm in forecasting. Once the VTF model has been estimated, the vector $\boldsymbol{\beta}_{t+l}$ must be replaced by its deterministic extrapolation $\hat{\boldsymbol{\beta}}_{t+l}$, and $\hat{\mathbf{m}}_{t+l-1}$ and $\hat{\mathbf{n}}_{t+l-1}$ become vectors of lagged forecasts. A computational problem in (2.7)–(2.9) is that the predictor requires the series m_t, n_t, which are available only by filtering the whole sample $\{y_t, x_t\}_1^N$; as we shall see, however, these quantities are generated in the estimation phase.

What makes the Markovian calculations (2.4)–(2.6) fundamental is the expression of the variance of the l-steps-ahead prediction error $\hat{e}_t(l) = y_{t+l} - \hat{y}_t(l)$. This variance is necessary for the confidence intervals of $\hat{y}_t(l)$, and since in (2.5) the sequences $\{v_{i_t}, \psi_{j_t}\}$ are deterministic, it can be

obtained as by Box and Jenkins (1970, p. 405):

$$\hat{\sigma}_t^2(l) = E[\hat{y}_t(l) - y_{t+l}]^2 = \left(\sum_{i=0}^{l-b} v_{i_{t+l}}^2\right)\sigma_x^2 + \left(1 + \sum_{j=1}^{l-1} \psi_{j_{t+l}}^2\right)\sigma_a^2,$$

where for example, $v_{1_{t+l}} = \text{tr}[\boldsymbol{\Delta}_{t+l}\boldsymbol{\Omega}_{t+l-1}]$ and $\boldsymbol{\Delta}'_{t+l} = [\boldsymbol{\delta}_{t+l}: \mathbf{I}_{r-1}]$, $\boldsymbol{\omega}'_{t+l-1} = [\boldsymbol{\omega}_{t+l-1}: \mathbf{O}_s]$, and so forth. The advantage of forecasting with VTF systems is then twofold: One can predict the parameters $\boldsymbol{\beta}_t$, but in a deterministic fashion, that is, leaving unchanged the inferential framework of the stationary TF models. In particular, this means that the predictor (2.7)–(2.9) is asymptotically normal distributed. A similar result could not be established in adaptive forecasting, in which the last available recursive estimate $\hat{\boldsymbol{\beta}}(t)$ is utilized from $t + 1$ to $t + l$, and whose statistical properties are uncertain.

Example. The empirical example that begins here involves the relationship of causality between two economic processes. We define X = exchange rate £/\$, Y = index of wholesale prices, and t = January 1973–December 1985 ($N = 156$). Stationarity in the levels was reached with difference of order one; the resulting series are displayed in Figure 2, showing nonstationarity in covariance. Table 1 reports the sample correlation functions (CF's); since x_t is practically a white noise, the implied TF model is

$$(1 - B)Y_t = [(\omega_0 + \omega_1 B)/(1 - \delta_1 B)](1 - B)X_t + [1/(1 - \phi_1 B)]a_t.$$

Table 2 reports the parameter estimates obtained with two related methods: nonlinear least squares (NLS) and pseudolinear regression (PLR). Since the next sections are based on these algorithms, it is worth noting here that their performance on these data is identical. To make a preliminary guess on the nonstationarity of the system, Table 2 provides NLS estimates on three subsamples.

3. IDENTIFICATION

A natural way of identifying the functional form of the parameters is by estimating their value for each instant with sequential (on-line) estimators; the resulting series are then fitted with functions of time. Recursive estimators for TF models were provided by Solo (1978), Ljung and Söderström (1983), Young (1984), and Sherif and Liu (1987), using the PLR, NLS, refined instrumental variable (RIV), and extended Kalman filter (EKF) approaches respectively. Given its computational simplicity and its adaptability (we see later what this means in practice), I refer to the PLR approach.

Recursions. Heuristically, an *iterative* PLR (IPLR) algorithm for the TF model may be obtained as by Spliid (1983), by generating at each step the pseudolinear regressors $\mathbf{z}_t(\cdot)$ and then applying the ordinary least squares (OLS) method to the model (2.3):

$$\hat{\boldsymbol{\beta}}(k + 1) = \left[\sum_{t=1}^{N} \hat{\mathbf{z}}_t(k)\hat{\mathbf{z}}_t'(k)\right]^{-1} \sum_{t=1}^{N} \hat{\mathbf{z}}_t(k)y_t. \quad (3.1)$$

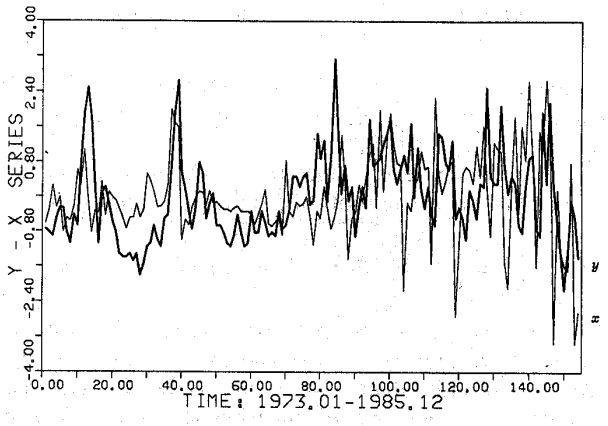

Figure 2. Series Resulting From Differencing and Standardizing Two Economic Processes.

The corresponding *recursive* PLR (RPLR) can easily be recovered by equating the number of iterations and the number of processed data ($k = N$) = t and proceeding as in the derivation of the recursive least squares algorithm (RLS; see Plackett 1950):

$$\begin{aligned} \hat{e}_t &= y_t - \hat{\boldsymbol{\beta}}(t-1)'\hat{\mathbf{z}}_t(t-1) \\ \mathbf{R}(t) &= \lambda \cdot \mathbf{R}(t-1) + \hat{\mathbf{z}}_t(t-1)\hat{\mathbf{z}}_t'(t-1) \\ \hat{\boldsymbol{\beta}}(t) &= \hat{\boldsymbol{\beta}}(t-1) + \mathbf{R}(t)^{-1}\hat{\mathbf{z}}_t(t-1)\hat{e}_t \end{aligned} \quad (3.2)$$

and

$$\begin{aligned} \hat{m}_t &= \sum_{i=1}^{r} \hat{\delta}_i(t)\hat{m}_{t-i} + \sum_{i=0}^{s} \hat{\omega}_i(t)x_{t-b-j} \\ \hat{n}_t &= y_t - \hat{m}_t \\ \hat{a}_t &= \hat{n}_t - \sum_{j=1}^{p} \hat{\phi}_j(t)\hat{n}_{t-j} - \sum_{j=1}^{q} \hat{\theta}_j(t)\hat{a}_{t-j}. \end{aligned} \quad (3.3)$$

(3.3) provides a proper dynamic adaptation of the three-step filtering of Box–Jenkins, and the quantities $\hat{m}_t$, $\hat{n}_t$, and $\hat{a}_t$ are used for updating the vector of regressors $\hat{\mathbf{z}}_t(\cdot)$. By discounting old observations in $\mathbf{R}(t) = \sum_\tau^t \lambda^{t-\tau}\hat{\mathbf{z}}_\tau\hat{\mathbf{z}}_\tau'$ and preventing $\mathbf{R}(t)^{-1}$ from vanishing, the parameter $0 < \lambda < 1$ actually enables the tracking of parameter changes $\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}$.

The properties of this algorithm have been investigated in depth only under the assumptions of stationarity, stability, and $\lambda = 1$. Since the gradient of the (optimal) NLS estimator of the TF model takes on the analytic expression

$$\begin{aligned} \boldsymbol{\xi}_t(\boldsymbol{\beta}) &= -\frac{\partial a_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{G}(B)\mathbf{z}_t(\boldsymbol{\beta}), \\ \mathbf{G}(B) &= \text{diag}\left[\frac{\phi(B)}{\theta(B)\delta(B)} \cdots \frac{1}{\theta(B)}\right], \end{aligned} \quad (3.4)$$

it has been established (Ljung and Söderström 1983; Solo 1978) that the approximation $\boldsymbol{\xi}_t \approx \mathbf{z}_t$, implicitly made by (3.3), does not affect the properties of convergence in probability only if $\mathbf{G}(B)$ behaves like a passive filter: $\Re[\mathbf{G}(z)] > 0$ $|z| = 1$. This feature, for second-order polynomials, is satisfied on about 90% of the stability region.

Table 1. Sample Correlation Functions

Function	Lags 0	1	2	3	4	5	6	7	8	9	10	11	12	Q(24)
ACR(y)	1	.62	.41	.28	.23	.20	.14	.11	.09	.00	−.03	.00	.08	47.3
ACR(x)	1	.24	.07	.05	.08	.12	.00	.01	.06	−.05	−.06	−.13	−.06	21.4
CCR(y, x)	.34	.45	.24	.23	.19	.10	.08	.07	.14	.04	−.07	−.10	−.14	38.1
CCR(x, y)	.34	.07	.08	.05	.06	.02	.01	.04	.06	−.01	.04	.10	.14	17.5

NOTE: ACR denotes autocorrelation. CCR denotes cross-correlation.

The nonoptimality of (3.2) is irrelevant at this stage, however, because we are assuming nonstationary parameters, and we are only interested in the tracking capability of the algorithm. In statistical terms, this problem has been well investigated in regression models with nonstochastic regressors and random-walk parameters. The conclusion was that the Kalman filter provides the best mean squared error (MSE) estimator if the probability distribution of parameters is known a priori. The systematic analysis of algorithms of type (3.2)–(3.3) has received attention from Benveniste (1987) and Solo (in press); the general framework developed by Benveniste seems particularly useful for dealing with the VTF system. In the sequel, I assume that the true parameter model $\boldsymbol{\beta}_0(\boldsymbol{\alpha}, t)$ is a continuous and bounded function of time, unknown in the coefficients $\boldsymbol{\alpha}$ and the structure.

First, we exclude that (3.2) may be a consistent estimator. In fact, the estimator of $\Delta_0(t) = \boldsymbol{\beta}_0(t) - \boldsymbol{\beta}_0(t-1)$ is $\mathbf{R}(t)^{-1}\hat{\mathbf{z}}_t\hat{e}_t = \hat{\Delta}(t)$; hence unless $\lambda = 1$ and $\mathbf{R}(t)^{-1} \to 0$ a.s., the estimator will be constantly subject to the random fluctuations of the prediction error e_t. On the other hand, for $\lambda = 1$ we have $\hat{\Delta}(t) \to 0$ a.s., but since $\Delta_0(t) \neq 0$ the estimator will be uniformly biased. This conclusion must not prevent us from analyzing other properties, such as minimum MSE and unbiasedness, suitable for characterizing the tracking capability of (3.2)–(3.3) in statistical terms.

Since $\boldsymbol{\beta}_0(t)$ is differentiable we must reduce $\hat{\boldsymbol{\beta}}(t)$ to the same nature. To this end, we rewrite (3.2) in stochastic approximation from (see Tsypkin 1971):

$$\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\beta}}(t-1) + \gamma\boldsymbol{\Gamma}H(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0, \mathbf{z}, t), \tag{3.5}$$

where γ is the stepsize, $\boldsymbol{\Gamma}$ is a nondecreasing gain matrix, and $H(\cdot)$ is the general search direction. (3.5) is fairly general, because in PLR with $\lambda = 1$ it means $\gamma = 1/t$ and $\boldsymbol{\Gamma} = \mathbf{R}(t)^{-1}t$, whereas in NLS we have $H = \boldsymbol{\xi}_t e_t = \mathbf{G}\mathbf{z}_t e_t$

Table 2. Estimated Transfer Functions

Estimates	Parameter ω_0	ω_1	δ_1	ϕ_1	Residual sum of squares	Q(24)
NLS	.0104	.0098	.458	.552	114.3	16.7
	(4.2)	(3.2)	(3.2)	(8.1)		
PLR	.0104	.0098	.465	.550	114.5	17.4
	(4.2)	(3.1)	(3.4)	(8.0)		
Subsamples						
t = 1–50	.025	.017	.66	.61	16.6	8.5
t = 51–100	.012	.001	.81	.73	35.3	14.7
t = 101–150	.007	.011	.37	.29	45.3	17.2

NOTE: t-statistics are in parentheses.

and so on. Now, assuming γ is fixed and $\boldsymbol{\Gamma}$ is constant, we may study the performance of (3.5) by means of its associated ordinary differential equation (ODE) (see Ljung and Söderström 1983, p. 147):

$$\partial\boldsymbol{\beta}(t)/\partial t = \boldsymbol{\Gamma}\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\beta}_0, t), \tag{3.6}$$
$$\mathbf{h}(\cdot) = E[H(\boldsymbol{\beta}, \boldsymbol{\beta}_0, \mathbf{z}, t)].$$

The meaning of this must be sought in the behavior of its solution $\boldsymbol{\beta}(t)$ with respect to the estimator $\hat{\boldsymbol{\beta}}(t)$: $\lim_{\gamma\to 0} p(\|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(\gamma t)\| > \varepsilon) = 0$. The situation in the evolving context is more complex, however, and requires that for a $\boldsymbol{\beta}_0(t)$ fixed, $\boldsymbol{\beta}(t) = \boldsymbol{\beta}_0(t)$ be a locally stable equilibrium of (3.6). This may be ensured by assuming that the algorithm correctly matches the true system [i.e., $\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\beta}_0, t) = 0 \Leftrightarrow \boldsymbol{\beta}(t) = \boldsymbol{\beta}_0(t)$], and that the matrix $\boldsymbol{\Gamma}$ is asymptotically stable:

$$\Re\{\text{eig}[\boldsymbol{\Gamma}\mathbf{h}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, t)\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}]\} < 0 \quad \forall\, t, \tag{3.7}$$
$$\mathbf{h}_{\boldsymbol{\beta}}(\cdot) = \frac{\partial\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\beta}_0, t)}{\partial\boldsymbol{\beta}(t)}.$$

The behavior of (3.6) can now be compared with its true counterpart $\partial\boldsymbol{\beta}_0(t)/\partial t$. Here, although $\boldsymbol{\beta}_0(t)$ is smooth, its associated ODE may not have a stable equilibrium [Benveniste (1987) called this situation a *nonzero drift* parameter model], so a permanent effort is required for the algorithm to keep the bias small enough. In other words, the speed of adaptation (γ) of (3.5) must be larger than the speed of variation of $\boldsymbol{\beta}_0(t)$, and (3.7) becomes necessary. If these conditions are met, then following Benveniste (1987, p. 14) we get the

$$E\|\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\|^2 \approx \left\|[\boldsymbol{\Gamma}\mathbf{h}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, t)]^{-1}\left[\frac{\partial\boldsymbol{\beta}_0(t)}{\partial t}\right]\right\|^2 + \text{tr}\,\mathbf{P}(\boldsymbol{\beta}_0, t), \tag{3.8}$$

where $\mathbf{P}(\cdot)$ is the dispersion of the estimator that satisfies the Lyapunov equation $(\boldsymbol{\Gamma}\mathbf{h}_{\boldsymbol{\beta}})\mathbf{P} + \mathbf{P}(\boldsymbol{\Gamma}\mathbf{h}_{\boldsymbol{\beta}})' + \boldsymbol{\Gamma}E(H_t H_t')\boldsymbol{\Gamma}' = \mathbf{0}$. Benveniste's results are more precise and general; they regard both deterministic and stochastic parameter models, with zero or nonzero drift. At this point, however, I wish to comment on the implications of the previously given results.

1. (3.8) provides the classical decomposition of the MSE in terms of bias^2 + variance; this means that recursive estimators like (3.5) are not only inconsistent for $\boldsymbol{\beta}_0(t)$, but persistently biased as well.

2. The optimal design of the gain $\boldsymbol{\Gamma}$ requires the exact

knowledge of the speed of variation of the true system; in particular, the scaling factor $\|\Gamma\|$ should be of order greater than $\|\partial\beta_0(t)/\partial t\|$. This leads to a time-varying gain.

3. Without a priori information on the parameters, only reasonable choices of Γ are possible; increasing gains (i.e., speed of adaptation) reduce the bias in (3.8), but also inflate $\mathbf{P}$ in the Lyapunov equation. Therefore, the optimal design of Γ should provide a compromise between fast tracking and high accuracy.

To this unsatisfactory situation, we must add the difficulty of finding the sampling distribution of recursive estimates. In the stationary linear model $y_t = \beta'\mathbf{x}_t + a_t$ with independent regressors, we have that $\hat{\beta}(t) \to N[\beta, \sigma^2 E(\mathbf{x}_t\mathbf{x}_t')^{-1}(1 - \lambda)/(1 + \lambda)]$, but the same is hard to prove for the system (2.1), even under stationarity.

When a priori information on the system is not available, reasonable recursions can only provide a crude idea of the underlying parameter models. This strengthens our proposal of proceeding further in the model building.

Interpolations. The question of fitting recursive estimates with time functions largely coincides with a problem of numerical analysis. Recall the Weierstrass theorem, which states that any continuous function can be approximated in a closed interval by a finite linear polynomial. A better performance is clearly guaranteed by orthogonal polynomials of Lagrange, Chebyshev, and so forth (see Farley et al. 1975).

Consistent with modern spline techniques that combine kernels instead of polynomials, a general interpolation strategy is represented by a linear combination of nonlinear local functions $\hat{\beta}_i(t) = \sum_j \alpha_{ij} f_{ij}(t)$. Here, *local* means functions that outside certain finite intervals are practically constant; typical examples are probability densities and distributions, spectral windows and their integrals, and so on. The parameters of location and slope of these functions are identifiable on the plots of $\hat{\beta}_i(t)$. In any case, before the interpolation a smoothing of $\hat{\beta}(t)$ with moving averages is necessary to show up the functional form of $\beta_0(t)$.

Example. Figure 3 displays RPLR estimates fitted with polynomials of third order ($\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$); these functions have been broadly used in traditional spline techniques. The sequential algorithm (3.2)–(3.3) has been initialized with $\hat{\beta}(0) = \hat{\beta}/2$ of the off-line estimation (Table 1) and the coefficients $\lambda = .97$ and $\mathbf{R}(0) = \text{diag}[1/.17]$, which empirically have yielded uniform and mild variability of estimates. Table 3 resumes the corresponding results of regression.

4. ESTIMATION

In this phase the identified parameter functions are inserted in the TF model, as initial values, and their coefficients are reestimated on the original data by optimization programs. The problems to be tackled are highly nonlinear, however, such that optimal estimation methods NLS and maximum likelihood—which require the analytical specification of complex likelihood or objective functions and the exact computation of their gradient—have serious numerical problems; often they do not converge or yield inadmissible results. Suboptimal solutions, based on off-line PLR techniques, can provide practicable alternatives. In fact, they avoid any analytical specification and are easily implementable on standard statistical software. Moreover, they treat the nonlinearity of the TF system separately from that of the parameter functions. These features permit the monitoring of the minimization process to enable the identification and control of the specific factors that cause numerical instability. To simplify the derivation and the analysis of these techniques, we refer without loss of generality to a simple moving average model MA(1).

Derivation. I begin with parameter functions given by linear combinations of time functions, generally nonlinear but observable from the identification phase: $z_t = \theta_t(B)a_t = (1 + \theta_t B)a_t$ and $\theta_t = \sum_i \theta_i f_i(t) = \boldsymbol{\theta}'\mathbf{f}(t)$ (in the previous example we had $f_i(t) = t^i$). Now, given the usual NLS estimator

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \left[\sum_t^N \hat{\boldsymbol{\xi}}_t(k)\hat{\boldsymbol{\xi}}_t'(k)\right]^{-1} \times \sum_t^N \hat{\boldsymbol{\xi}}_t(k)\hat{a}_t(k) \quad (4.1)$$

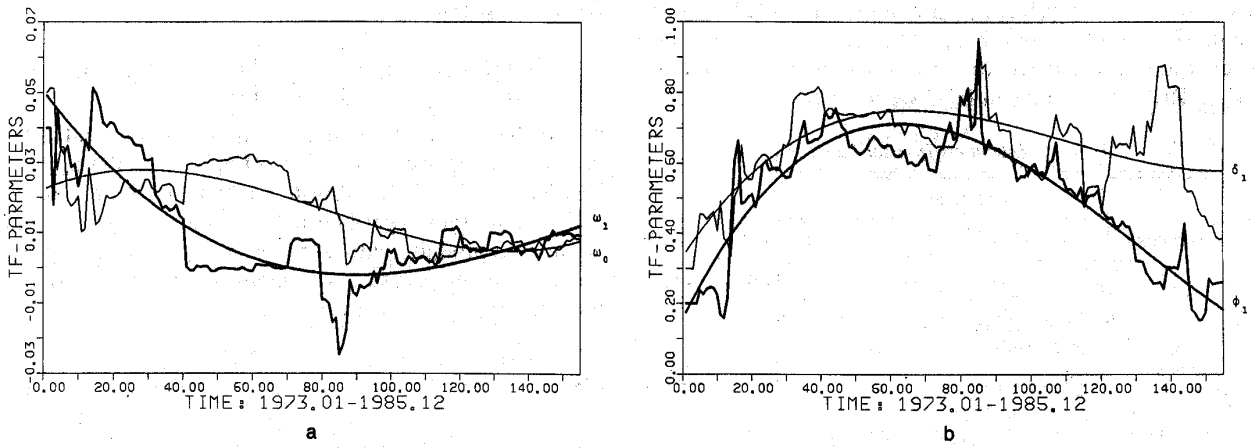

Figure 3. Recursive PLR Estimates and Their Interpolations: (a) ω_1 and ω_0; (b) δ_1 and ϕ_1.

Table 3. Identified Parameter Functions

Parameter	α_0	α_1	α_2	α_3	R^2
$\omega_0(t)$	.222E−1	.432E−3	−.936E−5	385E−7	.63
	(9.8)	(3.4)	(−5.0)	(4.9)	
$\omega_1(t)$	.505E−1	−.133E−2	101E−4	−.201E−7	.73
	(19.2)	(−9.1)	(4.7)	(−2.2)	
$\delta_1(t)$	−.333E−0	−.150E−1	.164E−3	−.500E−6	.48
	(−10.3)	(−8.4)	(6.2)	(−4.5)	
$\phi_1(t)$	−.155E−0	−.197E−1	.207E−3	−.529E−6	.84
	(−6.6)	(−15.3)	(10.8)	(−6.5)	

NOTE: t-statistics are in parentheses.

and

$$\xi_t(\theta) = -\frac{\partial a_t}{\partial \theta} = \frac{1}{\theta_t(B)}\mathbf{f}(t)a_{t-1},$$

a formal derivation of the PLR algorithm consists of approximating the gradient with the input–output quantities, that is, avoiding the filtering with $1/\theta_t(B)$. Indeed, setting $\hat{\xi}_t(k) \approx \mathbf{f}(t)\hat{a}_{t-1}(k) \equiv \hat{\mathbf{x}}_t(k)$ and $\hat{a}_t(k) = z_t - \hat{\theta}(k)'\hat{\mathbf{x}}_t(k)$ and substituting these in (4.1), we may obtain a compact algorithm similar to (3.1):

$$\hat{\theta}(k+1) = \left[\sum_t^N \hat{\mathbf{x}}_t(k)\hat{\mathbf{x}}_t'(k)\right]^{-1} \sum_t^N \hat{\mathbf{x}}_t(k)z_t. \quad (4.2)$$

This estimator is an iterative weighted PLR (WPLR), where *weighted* means that the evolution $\mathbf{f}(t)$ is inserted in the pseudoregressors a_{t-1}, yielding a transformation of stationarity. This feature is peculiar to the PLR method.

In the case of nonlinear parameter functions containing unknown coefficients $z_t = [1 + f(\theta, t)B]a_t$, the derivation of the WPLR algorithm is not easy, and compact expressions like (4.2) cannot generally be derived. In heuristic terms, however, the idea of PLR is still applicable in the following way: generate $\hat{a}_t(k) = z_t - \hat{f}(k, t)\hat{a}_{t-1}(k)$ with $\hat{a}_0(k) = 0$, then NLS estimate $z_t = f(\theta, t)\hat{a}_{t-1}(k) + e_t$. This yields the doubly iterative estimator

$$\hat{\theta}(h+1 \mid k) = \hat{\theta}(h \mid k) + \left[\sum_t^N \hat{\mathbf{y}}_t(h \mid k)\hat{\mathbf{y}}_t'(h \mid k)\right]^{-1} \times \sum_t^N \hat{\mathbf{y}}_t(h \mid k)\hat{e}_t(h \mid k) \quad (4.3)$$

and

$$\hat{\mathbf{y}}(h \mid k) = -\left.\frac{\partial e_t}{\partial \theta}\right|_{\theta=\hat{\theta}(h|k)} = \hat{\mathbf{g}}_t(h)\hat{a}_{t-1}(k),$$

$$\hat{e}_t(h \mid k) = [z_t - \hat{f}(h, t)\hat{a}_{t-1}(k)],$$

which disaggregates the nonlinearity of the parameters (h) from that inherent to the model (k). Equating ($k = h$), we may compare (4.3) with the full NLS estimator (4.1). Since the exact gradient is

$$\xi_t(\theta) = -\frac{\partial a_t}{\partial \theta} = \frac{1}{\theta_t(B)}\mathbf{g}_t(\theta)a_{t-1}, \qquad \mathbf{g}_t(\theta) = \frac{\partial f(\theta, t)}{\partial \theta},$$

the algorithm (4.3) may still be obtained from (4.1) in the formal way of deriving PLR estimators, that is, by avoiding the filtering with $1/\theta_t(B)$: $\hat{\xi}_t(k) \approx \hat{\mathbf{g}}_t(k)\hat{a}_{t-1}(k) \equiv \hat{\mathbf{y}}_t(k)$. The goodness of this approximation determines the statistical properties.

Analysis. Specific analyses of the IPLR algorithm (3.1) in the hypothesis of stationarity were provided by Stoica, Söderström, Ahlen, and Solbrand (1985) and Hannan and McDougall (1988). Their conclusions generalize those already reached in the analysis of the recursive version (3.2–3.3) by Solo (1978) and Ljung and Söderström (1983). That is, PLR's by approximating optimal estimators tend to be inefficient and may not converge. In the sequel, I prove these results for the WPLR method with informal but direct arguments.

From the analysis of stochastic approximation schemes (Tsypkin 1971, p. 56) we know that a sequential algorithm is consistent if the expectation (in θ_0) of the angle between the direction of the estimator (Δ) and the gradient of the objective function (∇) is positive. Now, for (4.2) and (4.3) with a quadratic objective function, we have

$$\hat{\Delta}(k) = [\hat{\theta}(k+1) - \hat{\theta}(k)] = \left[\sum_t^N \hat{\mathbf{y}}_t(k)\hat{\mathbf{y}}_t'(k)\right]^{-1} \sum_t^N \hat{\mathbf{y}}_t(k)\hat{a}_t(k)$$

and

$$\hat{\nabla}(k) = \left.\frac{\partial \sum_t^N a_t^2/2N}{\partial \theta}\right|_{\theta=\hat{\theta}(k)} = \frac{1}{N}\sum_t^N \frac{1}{\hat{\theta}_t(k, B)}\hat{\mathbf{y}}_t(k)\hat{a}_t(k),$$

so equating ($k = N$) $= t$, the condition of convergence becomes and implies

$$E[\Delta(t)'\nabla(t) \mid \theta_0] > 0 \Longleftrightarrow \Re[\theta_t(z)] > 0, \quad |z| = 1 \quad \forall\, t. \quad (4.4)$$

This condition directly extends to the evolving level of the conclusion reached in the analysis of PLR algorithms applied to stationary models. That is, the suboptimal WPLR scheme converges almost surely only if the polynomial $\{\theta_t(z)\}$ moves inside a subset of the invertibility region, in which it behaves like a passive filter. For general VTF models, (4.4) becomes $\Re[\mathbf{G}_t(z)] > 0$, $|z| = 1\ \forall\, t$, where $\mathbf{G}_t(\cdot)$ is defined as in (3.4).

To prove the inefficiency, we use the approach of Spliid (1983) based on the objective function $H_N(\theta) = \sum_t^N \mathbf{x}_t a_t/N$, concerning the correlation residuals-regressors in (4.2). Assume that (4.4) holds and expand the empirical $\hat{H}(t)$, with $t = (k = N)$, in θ_0:

$$[\hat{H}(t) - H_0] \approx \frac{1}{t}\sum_{\tau=1}^t \left[\frac{\partial \mathbf{x}_\tau}{\partial \theta'}a_\tau + \mathbf{x}_\tau\xi_\tau'\right] \times [\hat{\theta}(t) - \theta_0] \xrightarrow[t]{p} E[\mathbf{x}_t\xi_t'][\hat{\theta}(t) - \theta_0]$$

(the result $E[\partial \mathbf{x}_\tau/\partial \theta' \cdot a_\tau] = 0$ follows because $\mathbf{x}_\tau$ contains lagged pseudoregressors). Now, since $\lim tE[\hat{H}(t) - H_0]^2 = E[\mathbf{x}_t, \mathbf{x}_t']\sigma^2$ [$t = (k = N)$], the asymptotic dispersion becomes

$$\lim tE[\hat{\theta}(t) - \theta_0]^2 = E[\mathbf{x}_t, \xi_t']^{-1}E[\mathbf{x}_t, \mathbf{x}_t']\sigma^2E[\xi_t, \mathbf{x}_t']^{-1} > E[\xi_t\xi_t']^{-1}\sigma^2, \quad (4.5)$$

Table 4. Results of Iterative WPLR Estimation

Parameter	α_0	α_1	α_2	α_3	Statistic
$\omega_0(t)$	—	.140E−2	−.227E−4	.921E−7	RSS
		(3.5)	(−3.3)	(3.2)	93.7
$\omega_1(t)$	.800E−1	−.356E−2	428E−4	−.148E−6	
	(4.4)	(−4.0)	(3.7)	(−3.4)	R^2
$\delta_1(t)$	—	.378E−1	−.544E−3	.203E−5	.61
		(4.9)	(−4.0)	(3.3)	
$\phi_1(t)$	—	−.451E−1	−.647E−3	.240E−5	DW
		(5.3)	(−4.1)	(3.3)	1.95

NOTE: t-statistics are in parentheses. RSS denotes residual sum of squares, and DW denotes the Durbin–Watson statistic.

and the loss of efficiency in passing from (4.1) to (4.2) is thus well established.

Example. The VTF system resulting from inserting the third-order polynomials of Table 3 in the model of Table 2 was tentatively reestimated with the ZXSSQ routine of the IMSL library, but without success. The failure was probably due to the high number of parameters to be estimated (16) and the complexity of the quadratic objective function constructed with the filtering (3.3). Success was obtained with the WPLR algorithm implemented on the TSP package. Denoting the polynomials in vector form $\beta_i(t) = \mathbf{t}'\boldsymbol{\beta}_i$ [$t' = [1, t, t^2, t^3]$, $(i = 1, \ldots, 4)$], a typical iteration was performed as the following: generate $\hat{m}_t(k) = \mathbf{t}'\hat{\boldsymbol{\delta}}_1(k) \cdot \hat{m}_{t-1}(k) + \mathbf{t}'\hat{\boldsymbol{\omega}}_0(k) \cdot x_t + \mathbf{t}'\hat{\boldsymbol{\omega}}_1(k) \cdot x_{t-1}$, $\hat{m}_0(k) = 0$, and $\hat{n}_t(k) = y_t - \hat{m}_t(k)$ $(t = 1, 2, \ldots, N)$, then OLS estimate $y_t = [\mathbf{t}'\boldsymbol{\omega}_0]x_t + [\mathbf{t}'\boldsymbol{\omega}_1]x_{t-1} + [\mathbf{t}'\boldsymbol{\delta}_1]\hat{m}_{t-1}(k) + [\mathbf{t}'\boldsymbol{\phi}_1]\hat{n}_{t-1}(k) + e_t$. In the first iteration (using initial values of Table 3), all of the coefficients of the vectors $\boldsymbol{\omega}_0$, $\boldsymbol{\delta}_1$, and $\boldsymbol{\phi}_1$ had t ratios less than 1.2. The intercepts α_0, unimportant for the evolutive behavior, were then deleted, and surprisingly all of the coefficients turned out significant. Convergence was fairly quick (five iterations); estimation results are listed in Table 4 and the implied parameter functions are displayed in Figure 4.

Since the VTF system is deterministic in the parameters and linear in the variables, the classical inferential framework of the stationary models (e.g., see Pierce 1972) extends directly to it. In particular, comparisons with the static TF of Table 1 can be made with a simple F test: $\hat{F} = 3.5 > 2.5 = F_{1\%}(9, 142)$. This assumes the meaning of a test on the degree of deterministic evolution of the system.

5. CONCLUSIONS

The weakness of the proposed modeling clearly resides in the impossibility of having a general interpolating function for the parameters that is able to adapt to the various situations of evolution and in a parimonious manner. Recently, promising numerical–statistical techniques of deterministic interpolation, such as nonparametric spline smoothing (Silverman 1985) and projection pursuit regression (Friedman and Stuetzle 1981) have been proposed. Nevertheless, in the context of this work their implementation should be substantially reexamined.

External modelings of the parameters with solutions of stochastic type, such as fitting the recursive estimates with vector ARMA models, $\mathbf{A}(B)(\mathbf{I} - B)\hat{\boldsymbol{\beta}}(t) = \mathbf{C}(B)\mathbf{e}_t$, overcome the problem of the rigidity of deterministic time functions, but they remain conditioned on arbitrary priors and require the redefinition of the inference. These features actually raise questions on the real meaning of these solutions, since analysis and design of on-line estimators (PLR, NLS, RIV, EKF) in the presence of evolving parameters is still an open question.

The advantage of the internal modeling proposed in this article is that it is completely data-based; that is, the recursive estimation of the system [which requires as priors $\boldsymbol{\beta}(0)$, $\mathbf{R}(0)$, λ] only represents an intermediate phase. The possibility of inserting the identified parameter functions in the model and reestimating their coefficients on the original data may offset the nonoptimal specification of the priors. Finally, the classical inferential framework applies directly to it, with the possibility of making statistical comparisons with the static version of the system. The empirical results obtained in the article (a 17.5% reduction of the residual variance with the use of polynomials) are fairly encouraging and can be further improved.

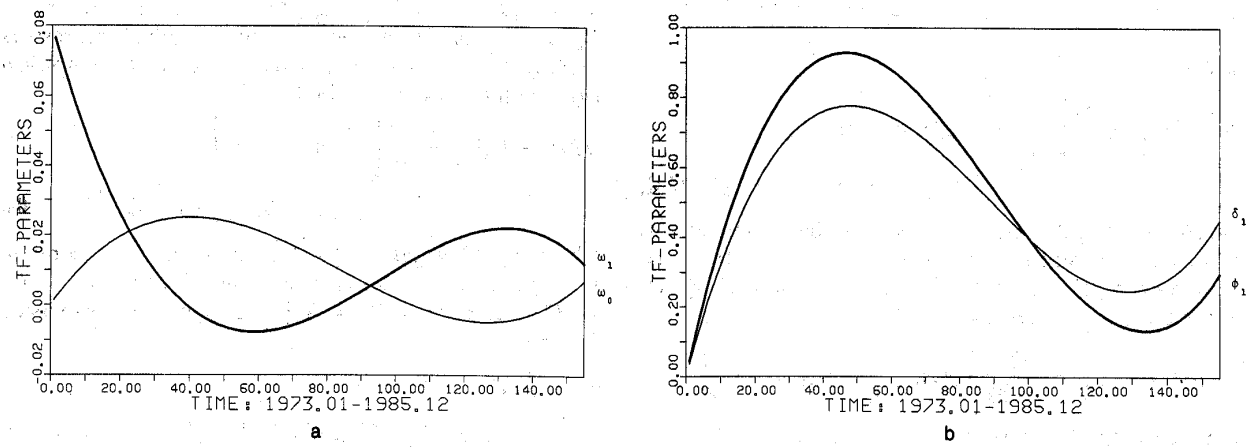

Figure 4. Parameter Functions Estimated on Original Data: (a) ω_1 and ω_0; (b) δ_1 and ϕ_1.

[*Received November 1986. Revised September 1989.*]

REFERENCES

Benveniste, A. (1987), "Design of Adaptive Algorithms for the Tracking of Time-Varying Systems," *International Journal of Adaptive Control and Signal Processing*, 1, 3–29.

Bittanti, S. (1986), "Deterministic and Stochastic Linear Periodic Systems," in *Time Series and Linear Systems*, ed. S. Bittanti, New York: Springer-Verlag, pp. 141–182.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

Farley, J. U., Hinich, M., and McGuire, T. W. (1975), "Some Comparisons of Tests for a Shift in the Slopes of a Multivariate Linear Time Series Model," *Journal of Econometrics*, 3, 297–318.

Friedman, J. H., and Stuetzle, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–824.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, New York: John Wiley.

Hannan, E. J., and McDougall, A. J. (1988), "Regression Procedures for ARMA Estimation," *Journal of the American Statistical Association*, 83, 490–498.

Ljung, L., and Söderström, T. (1983), *Theory and Practice of Recursive Identification*, Cambridge, MA: MIT Press.

Melard, G., and Kiehm, J. L. (1981), "ARIMA Models With Time Dependent Coefficients for Economic Time Series," in *Time Series Analysis*, eds. O. D. Anderson and M. R. Perryman, Amsterdam: North-Holland, pp. 355–363.

Pierce, D. A. (1972), "Least Squares Estimation in Dynamic Disturbance Time Series Models," *Biometrika*, 59, 73–78.

Plackett, R. L. (1950), "Some Theorems in Least Squares," *Biometrika*, 37, 149–157.

Poskitt, D. S. (1989), "A Method for the Estimation and Identification of Transfer Functions Models," *Journal of the Royal Statistical Society*, Ser. B, 51, 29–46.

Sherif, M. H., and Liu, L. M. (1987), "Estimation of Transfer Function Models Using Mixed Recursive and Nonrecursive Methods," in *Control and Dynamic Systems* (Vol. 26), ed. C. T. Leondes, Orlando, FL: Academic Press, pp. 157–185.

Silverman, B. W. (1985), "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting," *Journal of the Royal Statistical Society*, Ser. B, 47, 1–21.

Solo, V. (1978), "Time Series Recursions and Stochastic Approximation," unpublished Ph.D. dissertation, Australian National University, Canberra, Dept. of Statistics.

——— (in press), "The Limiting Behaviour of LMS," *IEEE Transactions on Acoustic, Speech and Signal Processing*, ASSP-37.

Spliid, H. (1983), "A Fast Estimation Method for Vector ARMAX Models," *Journal of the American Statistical Association*, 78, 843–849.

Stoica, P., Söderström, T., Ahlen, A., and Solbrand, G. (1985), "On the Convergence of Pseudolinear Regression Algorithms," *International Journal of Control*, 40, 1429–1444.

Tsypkin, Y. A. (1971), *Adaptation and Learning in Automatic Systems*, New York: Academic Press.

Young, P. (1984), *Recursive Estimation and Time Series Analysis: An Introduction*, New York: Springer-Verlag.