Optimal Recursive Estimation of Dynamic Models

Carlo GRILLENZONI*

This article checks, using both real and simulated data, the effectiveness of modern adaptive techniques to track the parameters of time-varying dynamic models. The real case studies concern a bone marrow transplant data set published by Tong, the gas furnace model of Box and Jenkins, and two series of West German interest rates. Simulation studies focus on ARX models with smoothly and suddenly changing parameters. The general approach is to compare the fitting–forecasting performance of classical and adaptive methods, holding fixed the order of the models. At the methodological level, the basic step is taken by unifying known estimators, such as recursive least squares and Kalman filter, into a general algorithm. Next, the problem of optimal design of the tracking coefficients (such as discounting factors and learning rates), is solved by optimizing a quadratic functional based on one-step-ahead prediction errors. All applications show that adaptive modeling, based on the design and the optimization of recursive algorithms, leads to significant improvements of the forecasting performance.

KEY WORDS: Bone marrow transplant data; Extended Kalman filter; Gas furnace data set; Recursive least squares; Transfer function models; West German interest rates.

1. INTRODUCTION

In time series analysis, system identification, and econometrics, there is a growing interest in statistical methods for nonstationary and nonlinear stochastic processes. Realizations of such processes, in the form of finite time series, may be encountered in many applied fields, including economics, industrial production, environmental phenomena, and medicine. Dynamic models developed by Box and Jenkins (1976) have worked satisfactorily in many situations under conditions of constant parameters and linearity in the variables; moreover, they provide useful starting points for new kind of representations. Indeed, bilinear autoregressive moving average (ARMA) models, threshold autoregressive models (see Tong 1990), are obtained by direct extension of standard time series schemes.

To properly link real information content of the data and hypotheses about the model structures, exploratory tools of data analysis must be used. With regard to nonstationarity and the detection of parameter changes, natural tools are given by recursive estimators with adaptive implementation (see Ljung and Söderström 1983). Except for the Kalman filter (KF) case, these methods do not assume specific laws of parameter evolution and as such are analogous to nonparametric regression techniques. The problem of designing the tracking coefficients of adaptive estimators—for example, the discounting factor in recursive least squares (RLS) and the learning rate in stochastic approximation schemes—is similar to the problem of choice of the window width in kernel type estimators (see Härdle, Hall, and Marron 1988). In the context of recursive algorithms, this problem may be solved efficiently by minimizing a quadratic loss function based on one-step-ahead prediction errors. This solution belongs to the conditional least squares (CLS) estimation for stochastic processes discussed by Hall and Heyde (1980, p. 172) and Tjostheim (1986) and extends the maximum likelihood approach used in Gaussian state-space systems (see Harvey 1989 and Pagan 1980).

The central purpose of this article is to check, on real and simulated data, the efficacy of modern adaptive methods in tracking the parameters of evolving dynamic models. At the methodological level, the basic steps are represented by the unification of existing algorithms into a general recursive estimator and the optimization of its coefficients on each data set by means of the CLS approach. The attempt to develop unified algorithms has been recently pursued by several authors in system identification (see Salgado, Goodwin, and Middleton 1988). Here, further developments are possible in the field of adaptive implementation, concerning variable tracking coefficients (see Bittanti, Bolzern, and Campi 1989) and on-line robustification. On the other hand, the idea of estimating the coefficients of adaptive algorithms is typical of econometrics, but it has been never applied to schemes other than the KF.

My case studies consist of medical, industrial, and economic applications. Section 2 focuses on the bone marrow transplant data set of Tong (1990); Section 3 investigates the gas furnace model of Box and Jenkins (1976), and Section 4 models interest rate series of West Germany. Finally, Section 5 presents, two simulation studies carried out on ARX systems with smoothly and suddenly changing parameters.

2. A MEDICAL APPLICATION

Tong (1990, p. 500) published a set of 55 observations on three variables—X = white blood cell count, Y = platelet count, and Z = hematocrit—from a patient affected by leukemia who received a bone marrow transplant. Platelet count in the initial period of posttransplant has been shown to be a good indicator of subsequent long-term survival; thus it represents the main factor to be explained. Whereas the process Z_t is stationary in mean, the variables X_t and Y_t exhibit a growth (see Fig. 1), which reflects the positive response of the patient's hematopoietic system. Owing to this common pattern, we have focused the analysis on these two variables. To make their dimension comparable, Y_t was divided by 10. Stationarity in mean was achieved with a difference of order one: $y_t = Y_t - Y_{t-1}$, $x_t = X_t - X_{t-1}$; the corresponding sample correlation functions $r(y_t x_{t-k})$ are reported in Table 1. It can be seen that y_t and x_t are mutually dependent, but the

* Carlo Grillenzoni is Associate Professor of Statistics at the IUAV: Architecture University Institute of Venice-Ca'Tron, 30135 Venice, Italy. The author would like to thank the editors and the referees for the very constructive review process.

Journal of the American Statistical Association
September 1994, Vol. 89, No. 427, Applications & Case Studies

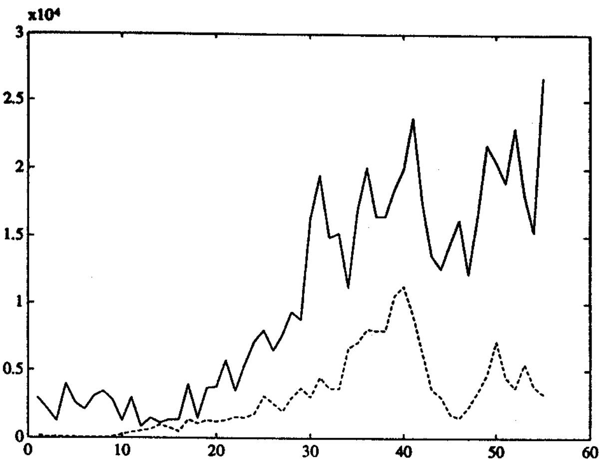

Figure 1. Plot of Medical Time Series: Y_t(——), X_t (- - -).

main causal action, in terms of cross-correlation, is that of x_{t-k} on y_t. This is a natural consequence of the central role played by white blood cells in the immune system. Applying the Box-Jenkins analysis to rows 1 and 3 of Table 1 leads us to the ARX model,

$$y_t = \underset{(3.2)}{-0.378}\, y_{t-2} + \underset{(5.6)}{1.647}\, x_{t-2} - \underset{(3.8)}{1.141}\, x_{t-3} + \hat{a}_t,$$

$$R^2 = .49, \quad Q = 273 \times 10^6, \quad (1)$$

where the values in parentheses are t ratios and Q is the sum of squared residuals (in-sample prediction errors). The weakness of the feedback $y_t \Rightarrow x_t$ is confirmed by the fact that the R^2 coefficient of the inverse model

$$x_t = \underset{(2.0)}{.264}\, x_{t-1} + \underset{(2.2)}{.114}\, y_{t-1} - \underset{(2.9)}{.164}\, y_{t-2} + \hat{e}_t$$

was .27.

Under normal circumstances, biological systems may present nonlinear dynamics (see Tong 1990) but are substantially stationary. But when diseases occur and the effective treatments are supplied they become, by definition, evolutive. In particular, this is the case for radical interventions such as transplants, because they involve the adaptation of a donor's hematopoietic system in the patient's body. To check the stationarity of model (1), we now introduce recursive estimation methods that allow the regression coefficients to vary over time.

Writing the ARX system in polynomial form with the Box-Jenkins notation, we obtain $(1 - \phi_1 B^2) y_t = (\omega_0 + \omega_1 B) x_{t-2} + a_t$, where B is the lag operator. This model may be written in regression form as $y_t = \boldsymbol{\beta}' \mathbf{z}_t + a_t$ where $\mathbf{z}_t' = [y_{t-2}, x_{t-2}, x_{t-3}]$ is the vector of "regressors" and $\boldsymbol{\beta}' = [\phi_1, \omega_0, \omega_1]$ are the parameters. Now, replacing the constant vector $\boldsymbol{\beta}$ by a time-varying vector $\boldsymbol{\beta}_t$, a recursive estimator for $\boldsymbol{\beta}_t$ that combines the main adaptive algorithms is given by

$$\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t-1} + \alpha \mathbf{P}_t \mathbf{z}_t (y_t - \mathbf{z}_t' \hat{\boldsymbol{\beta}}_{t-1}), \qquad \hat{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0 \quad (2a)$$

$$\mathbf{P}_t = \frac{1}{\lambda}\left(\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{z}_t\mathbf{z}_t'\mathbf{P}_{t-1}}{\lambda + \mathbf{z}_t'\mathbf{P}_{t-1}\mathbf{z}_t}\right) + \gamma_1 \mathbf{I}, \qquad \mathbf{P}_0 = \gamma_0 \mathbf{I}, \quad (2b)$$

where $(\alpha, \lambda, \gamma_1)$ are tracking coefficients and $(\boldsymbol{\beta}_0, \gamma_0)$ are parametric initial values. $\mathbf{P}_t$ is the "gain" matrix, which may approach the dispersion matrix of $\hat{\boldsymbol{\beta}}_t$. Some alternative formulations leading to (2) are discussed in the Appendix.

Algorithm (2) encompasses three basic adaptive estimators:

1. The RLS with exponentially weighted observations (EW-RLS), when $(\alpha = 1, \gamma_1 = 0)$. In this case (2a) is asymptotically equivalent to the discounted ordinary least squares (OLS) estimator $\hat{\boldsymbol{\beta}}_t = (\sum_{i=1}^t \lambda^{t-i} \mathbf{z}_i \mathbf{z}_i')^{-1} (\sum_{i=1}^t \lambda^{t-i} \mathbf{z}_i y_i)$, where $0 < \lambda < 1$. This version involves direct matrix inversions but does not require initial conditions.
2. The simplified KF, when $(\alpha = 1, \lambda = 1)$. A probabilistic interpretation of the filter is that it provides the optimal mean squared error (MSE) estimator when the parameters follow the Gaussian random-walk dynamics $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{e}_t$, $\mathbf{e}_t \sim \mathrm{IN}(\mathbf{0}, \gamma_1 \mathbf{I})$ with initial condition $\boldsymbol{\beta}_{t_0} \sim \mathrm{N}(\boldsymbol{\beta}_0, \gamma_0 \mathbf{I})$. In practice, $\hat{\boldsymbol{\beta}}_t$ minimizes the distance $E\|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t\|^2$ for each t.
3. The least mean squares (LMS), when $(\gamma_1 = 1, 1/\lambda = 0)$. This algorithm is also known as the stochastic gradient, because it implies that $\mathbf{P}_t = \mathbf{I}$ for all t and α is the learning rate.

These algorithms are discussed in detail in the books of Ljung and Söderström (1983), Goodwin and Sin (1984), and Widrow and Stearns (1985). The need to unify them into a general scheme arises from the facts that no method is the best one and that the real dynamics of parameters are unknown. In particular, they might be nonlinear or deterministic.

In general, algorithms 1 and 3 do not assume an explicit model for the parameter evolution. But given the algebraic relationships between the various schemes, the EW-RLS may be viewed as a KF in which the variance component $\gamma_1 \mathbf{I}$ is replaced by

$$\boldsymbol{\Sigma}_t = \left(\frac{1}{\lambda} - 1\right)\left[\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{z}_t\mathbf{z}_t'\mathbf{P}_{t-1}}{\lambda + \mathbf{z}_t'\mathbf{P}_{t-1}\mathbf{z}_t}\right]$$

$$= \left(\sum_{i=1}^{t} \lambda^{t-i} \mathbf{z}_i \mathbf{z}_i'\right)^{-1} - o_p(1)$$

Table 1. Sample Correlation Functions of Medical Series

Lag	0	1	2	3	4	5	6	7	8	9	10
$r(y_t y_{t-k})$	1	−.160	−.302	.069	−.047	.023	.090	−.069	−.062	.007	.209
$r(x_t x_{t-k})$	1	.165	−.069	.146	−.184	−.057	−.102	−.369	−.118	.049	.112
$r(y_t x_{t-k})$	.106	.034	.438	−.336	−.303	.169	−.001	−.005	−.094	−.063	.113
$r(x_t y_{t-k})$	.106	.317	−.293	−.082	.318	−.064	−.184	−.166	.072	.206	.008

(see the Appendix). Thus the "implicit" parameter dynamics of algorithm 1 is a random walk $\beta_t = \beta_{t-1} + e_t$ in which the input has the conditional distribution $(e_t|z_t, z_{t-1} \ldots) \sim \text{IN}(0, \Sigma_t)$. A similar interpretation holds for the LMS (see Ljung and Gunnarsson 1990); however, it is quite difficult to understand the meaning of the matrix Σ_t. Because the filter (2) may be further extended, we prefer to look at the foregoing unification in terms of a nonparametric regression, in which $\beta_t = f(z_t, z_{t-1} \ldots)$ is a function of lagged values of input and output $z'_{t-k} = [y_{t-k-1}, x_{t-k-1}]$ and (2) is a one-sided smoother.

It is easy to check that both coefficients (λ, γ_1) have the role of preventing the matrix P_t from tending to 0, which is the essential condition for tracking the parameter changes $(\beta_t - \beta_{t-1})$. As in the nonparametric regression, their design should provide a suitable trade-off between tracking of the regression function and accuracy (smoothness) of its trajectory. An optimal selection procedure for the coefficients of (2) arises from CLS estimation (see Hall and Heyde 1980, p. 172; Klimko and Nelson 1978; and Tjostheim 1986). This amounts to minimizing a penalty function given by the sum of squared innovations (one-step-ahead prediction errors),

$$(\hat{\alpha}, \hat{\lambda}, \hat{\gamma}_1; \hat{\beta}_0, \hat{\gamma}_0)_T = \arg\min\left\{Q_T = \sum_{t=1}^{T} [y_t - E(y_t|z_t, z_{t-1} \ldots)]^2\right\}, \quad (3)$$

where T is the number of observations. In Gaussian state-space systems, the criterion $-Q_T$ approximates the log-likelihood function when the KF approaches the steady state; that is, when $P_t \to \bar{P}$ constant (see Harvey 1989, p. 185 and appendix).

Calculation of Q_T is realized by (2) with the prediction errors $\tilde{a}_t = (y_t - z'_t\hat{\beta}_{t-1})$ of Equation (2a). These must not be confused with the recursive residuals $\hat{a}_t = (y_t - \hat{\beta}'_t z_t)$, which tend to 0 as P_t increases, by letting $\lambda \to 0$ or $\gamma_1 \to \infty$. Usually Q_T has a nonzero lower bound and may be minimized even with respect to the starting values (β_0, γ_0), which have an important role in the tracking. To simplify the nonlinear optimization (3), and to avoid identification problems, it may be reasonable to constrain the value of some coefficients such as $(\gamma_0 = \gamma_1) = \gamma$.

We now apply the adaptive framework (2)–(3) to the ARX model of the bone marrow transplant data. Optimization (3) was carried out with the MAXLIK routine of the GAUSS package using the designs adopted in system identification as initial values for the coefficients—namely, $\alpha = .5$, $\lambda = .97$, $\gamma_1 = .001$, and $\gamma_0 = 1$—and letting $\beta_0 = \hat{\beta}_T$, the OLS estimates in (1). Convergence required many iterations; results are reported in Table 2 in constrained and unconstrained form. The t statistics (in parentheses) are quite significant, but asymptotic normality may not hold for highly nonlinear estimation problems such as (3). Constrained estimation was concerned with fixing $\alpha = -1$, $\gamma_1 = 0$ as suggested by the first row of Table 2; it provides the best result and points out the importance of LMS-RLS adaptation mechanisms. The final reduction of statistic Q_T with respect to the constant parameter model (1) is approximately 50%.

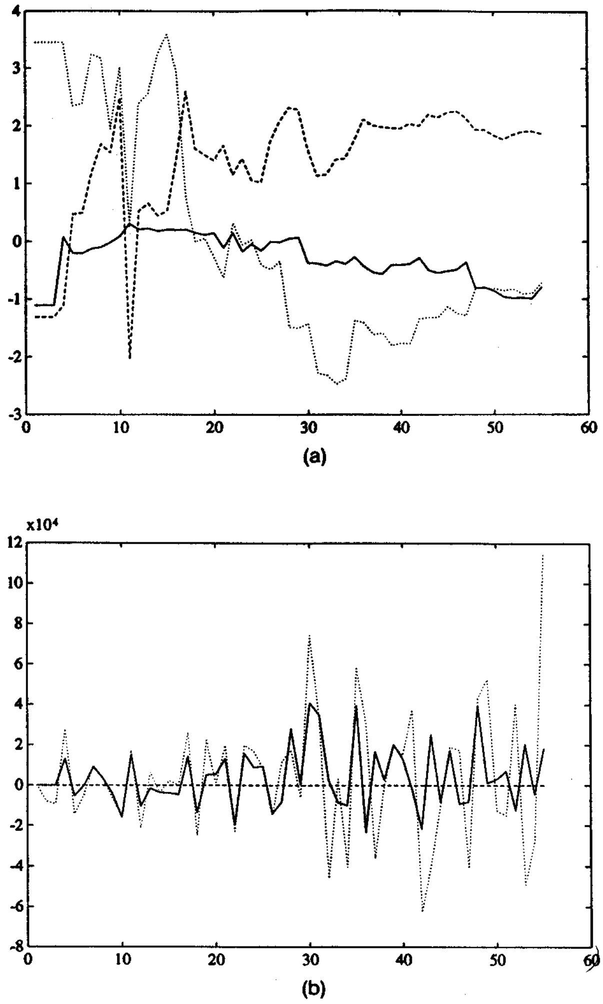

Figure 2. Plots of (a) Parameters ($\hat{\phi}_t$ ——, $\hat{\omega}_{0t}$ – – –, $\hat{\omega}_{1t}$ · · · ·) and (b) Innovations ($\tilde{a}_t$ ——, y_t · · · · ·).

Figure 2a shows the trajectories of the recursive estimates $\hat{\beta}_t$; Figure 2b plots the sequence of innovations $\tilde{a}_t$. Both are generated with the coefficients in row 1 of Table 2. It can be seen that mean values of $\hat{\phi}_{1t}$, $\hat{\omega}_{0t}$, and $\hat{\omega}_{1t}$ approach the OLS estimates in (1) and that the sequence $\tilde{a}_t$ has a nonincreasing variance. This is an important result, as the ultimate purpose

Table 2. CLS Estimates (and t Ratios) of the Coefficients of Algorithm (2)

Estimates	$\hat{\gamma}_0$	$\hat{\phi}_{10}$	$\hat{\omega}_{00}$	$\hat{\omega}_{10}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\gamma}_1$	Q_T
Normal	.014539 (3.1)	−1.108 (23.5)	−1.313 (10.9)	3.445 (12.8)	−.5638 (23.1)	.8380 (48.2)	.648E − 7 (1.5)	156.3E + 6
Constrained	.000352	−1.5318	−1.658	5.491	−1*	.8013	0*	136.7E + 6

* Designates constrained (fixed) values.

of a time-varying parameter modeling is to obtain stationary innovations.

3. AN INDUSTRIAL APPLICATION

In this section we apply and extend the adaptive framework (2)–(3) to the transfer function (TF) models of Box and Jenkins (1976), focusing on the well-known gas furnace data. This case study has already been treated with recursive estimators by Young (1984, 1985), who changed the order specification of Box–Jenkins. This will not be done in our application, because our aim is to compare constant and variable parameter methods.

Dynamic models with rational transfer functions and Gaussian noise $\{a_t\}$ connect an output process $\{y_t\}$ to a control input $\{x_t\}$ in the following way:

$$y_t = \frac{(\omega_0 + \omega_1 B + \cdots + \omega_s B^s)}{(1 - \delta_1 B - \cdots - \delta_r B^r)} x_{t-b} + \frac{(1 + \theta_1 B + \cdots + \theta_q B^q)}{(1 - \phi_1 B - \cdots - \phi_p B^p)} a_t, \quad a_t \sim \mathrm{IN}(0, \sigma^2), \quad (4)$$

where $(\omega_i, \delta_j; \theta_i, \phi_j)$ are parameters. Stability conditions require that polynomials $\delta(B)$, $\phi(B)$, and $\theta(B)$ have stable roots and that the coefficients $\omega_0, \omega_1 \ldots \omega_s$ are bounded. Model (4) is nonlinear in the parameters δ_i and θ_j but linear in the variables. To appreciate this, we define the auxiliary systems $m_t = [\omega(B)/\delta(B)]x_{t-b}$, $n_t = [\theta(B)/\phi(B)]a_t$ and derive the pseudolinear representation

$$y_t = \left(\sum_{i=1}^{r} \delta_i m_{t-i} + \sum_{i=0}^{s} \omega_i x_{t-b-i}\right) + \left(\sum_{j=1}^{p} \phi_j n_{t-j} + \sum_{j=1}^{q} \theta_j a_{t-j}\right) + a_t. \quad (5)$$

This equation may be rewritten as $y_t = \beta' \mathbf{z}_t + a_t$, where $\beta' = [\delta_1 \ldots, \omega_0 \ldots, \phi_1 \ldots, \theta_1 \ldots]$ is the vector of parameters and $\mathbf{z}_t' = [m_{t-1} \ldots, x_{t-b} \ldots, n_{t-1} \ldots, a_{t-1} \ldots]$ is the vector of "regressors." Note that the unobservable terms $\{m_{t-i}, n_{t-j}, a_{t-j}\}$ may be generated sequentially starting from $\{x_t, y_t\}$, as discussed by Box and Jenkins (1976, p. 390).

The gas furnace data set of Box and Jenkins (1976) concerns a sample of 296 observations from a gas furnace in which air and methane are combined to form a mixture of gases containing carbon dioxide (CO_2). The air feed was kept constant, but the methane feed rate (input) was varied to form the desired CO_2 concentration (output); the sampling interval was 9 seconds. Series $\{y_t, x_t\}$ with mean 0 are plotted in Figure 3a, together with the residuals (in-sample innovations) $\{\hat{a}_t\}$ of the fitted model

$$y_t = \begin{pmatrix} -.531 & -\ .378B & -\ .518B^2 \\ (-7.1) & (-3.6) & (-4.8) \end{pmatrix} \begin{pmatrix} 1 - .550B \\ (-15.4) \end{pmatrix}^{-1} x_{t-3} + \begin{pmatrix} 1 - 1.533B + .634B^2 \\ (-32.1) \quad (12.5) \end{pmatrix}^{-1} \hat{a}_t, \quad (6)$$

where $Q = \sum_{t=4}^{296} \hat{a}_t^2 = 16.67$. This model was identified by Box–Jenkins by means of sample correlation functions; some doubts about the adequacy of the specification are raised by the residuals in Figure 3b. We see in fact that the variance of $\hat{a}_t$ significantly increases over the last 100 observations, probably because of changes in the physical/chemical characteristics of the gas furnace and of its inputs. These changes are usually associated with variations of temperature of the plants and nonhomogeneity of the quality of raw materials. In both cases, the effects on the variability of system parameters are remarkable.

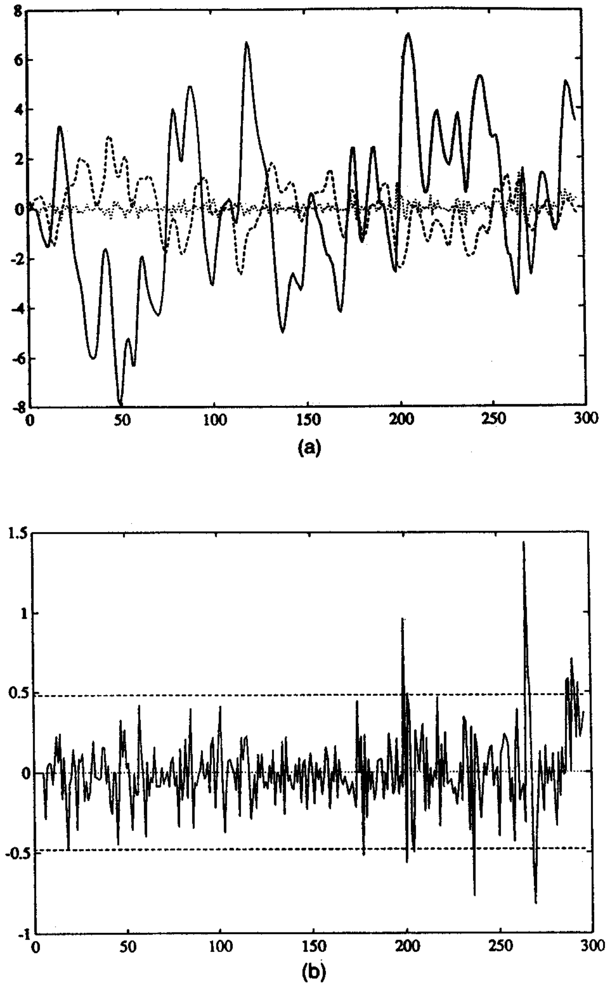

Figure 3. Plots of (a) Series (y_t ——, x_t - - -, $\hat{a}_t$ · · · · ·) and (b) Residuals ($\hat{a}_t$ ——, $\pm 2\hat{\sigma}$ - - -).

Nonlinearity in the parameters of (4) has practical consequences on the estimation, because iterative algorithms must be used: $\hat{\beta}_T^{(k)} = \hat{\beta}_T^{(k-1)} + \hat{\delta}_T^{(k)}$, where $\hat{\delta}_T^{(k)}$ is the adjustment in the kth iteration. Simple algebraic transformations, based on equating the number of iterations and the number of observations, $(k = T) = t$, yield the recursive version of such algorithms. An adaptive scheme that reconciles many of the existing algorithms for TF models—such as recursive maximum likelihood (RML; Ljung and Söderström 1983), extended Kalman filter (EKF; Goodwin and Sin 1984), and refined instrumental variables (RIV; Young 1984)—can be obtained from the approach suggested by Salgado et al. (1988) and by Bittanti et al. (1989):

$$\hat{\beta}(t) = \hat{\beta}(t-1) + (1 + \alpha_1 B + \alpha_2 B^2)\mathbf{P}(t)\hat{\zeta}(t)\tilde{a}(t), \qquad \hat{\beta}(0) = \beta_0 \quad (7a)$$

and

$$\mathbf{P}(t) = \frac{1}{\lambda}\mathbf{P}(t-1) - \mu\frac{\mathbf{P}(t-1)\hat{\zeta}(t)\hat{\zeta}(t)'\mathbf{P}(t-1)}{1+\hat{\zeta}(t)'\mathbf{P}(t-1)\hat{\zeta}(t)} + \gamma_1\mathbf{I} - \delta\mathbf{P}^{\kappa}(t-1), \qquad \mathbf{P}(0) = \gamma_0\mathbf{I}, \tag{7b}$$

where $\hat{\zeta}(t) = -[\partial a_t/\partial\beta]_{\beta=\hat{\beta}(t-1)}$ is the gradient vector and $\tilde{a}(t) = [y_t - \hat{\mathbf{z}}(t)'\hat{\beta}(t-1)]$ is the prediction error. All of the coefficients (λ, μ, γ_1, δ, κ) have the role of keeping the gain matrix $\mathbf{P}(t)$ positive definite and uniformly bounded in probability: $P_r[0 < \mathbf{P}(t) < \infty] = 1$, which is the essential condition for tracking the parameter changes.

Apart from the nonlinearity (incorporated by the gradient), Algorithm (7) differs from (2) for three structural reasons:

1. The stable polynomial $\alpha(B) = (1 + \alpha_1 B + \alpha_2 B^2)$. This makes (7) a multistep algorithm in the sense that past adjustments $\mathbf{P}(t-j)\hat{\zeta}(t-j)\tilde{a}(t-j), j > 0$ intervene in updating the current estimate. This approach was discussed by Benveniste (1987).
2. The linearity of Equation (7b) with respect to ($1/\lambda$, μ, γ_1, δ). Such an implementation introduces flexibility and may simplify the optimization of criterion (3).
3. The presence of the factor $-\delta\mathbf{P}^{\kappa}(t-1)$. This has the role of balancing the excess of discounting activity induced by the coefficients λ and γ_1, which may cause $\mathbf{P}(t)$ to blow up. Indeed, Salgado et al. (1988) have shown that for $\delta > 0$, $\kappa = 2$, the factor has the same effect as does the addition of a positive definite matrix to $\mathbf{P}^{-1}(t)$.

At the computational level, what fundamentally distinguishes the various estimators is the way the vector $\hat{\zeta}(t)$ is calculated. Standard differentiation shows that the analytical expression of the gradient of model (4) has the representation

$$\xi_t = \left[-\frac{\partial a_t}{\partial\beta}\right] = \mathbf{G}(B)\mathbf{z}_t,$$

$$\mathbf{G}(B) = \operatorname{diag}\left[\frac{\phi(B)}{\theta(B)\delta(B)}\mathbf{I}_{(r+s+1)}; \frac{1}{\theta(B)}\mathbf{I}_{(p+q)}\right],$$

which may be adapted to online calculation. In particular, note that for ARMAX models, where $\delta(B) \equiv \phi(B)$, we have $\hat{\xi}(t) = [\hat{\mathbf{z}}(t) - \sum_{i=1}^{q}\hat{\theta}_i(t)\hat{\xi}(t-i)]$. Thus for the choice of $\{\zeta_t\}$, there are two possibilities: setting $\zeta_t = \xi_t$ we have nonlinear (NL) estimators, whereas setting $\zeta_t = \mathbf{z}_t$ we have pseudolinear (PL) algorithms. In this context, the recursive residuals $\hat{a}(t) = [y_t - \hat{\beta}(t)'\hat{\mathbf{z}}(t)]$ have an important role, as they may be used for updating the vector of "regressors" $\hat{\mathbf{z}}(t+1)' = [\hat{m}(t) \ldots\ldots \hat{a}(t-q+1)]$.

We now apply the foregoing algorithm to the gas furnace model, trying to estimate its coefficients with the CLS approach (3). Leaving the parameter vector β_0 unconstrained, the parsimonious parameterization $\gamma_0\mathbf{I}$ of the covariance matrix may be justified by the fact that (ω_0, ω_1, ω_2; δ_1, ϕ_1, ϕ_2) have similar sizes. Because the coefficient κ is highly nonlinear, and its estimation may cause numerical problems, its value should be determined with a search procedure. In practice, given a small grid, $\kappa = .5, 2, 3$, CLS estimations of the coefficients λ, μ, $\ldots$, γ_0 are carried out conditional on each value of κ. Next, the best solution is selected on the basis of the lower value of Q_T.

Table 3 reports the estimates obtained with both the gradients $\mathbf{z}_t$ and ξ_t and conditional on the choice $\kappa = \frac{1}{2}$. Some of the coefficients are not significant (e.g., α_2 and γ_1), which means that a more parsimonious algorithm may be sought. The reduction of statistic Q_T over model (6) is greater than 25%; owing to the significance of the factor λ, this result may be largely ascribed to the RLS component and to the pseudolinear implementation $\zeta_t = \mathbf{z}_t$. The path of the parameter estimates $\hat{\beta}(t)$ generated with the coefficients in the first row of Table 3 can be seen in Figure 4. Moreover, Figure 5a shows the innovations $\tilde{a}(t)$ with the adaptive confidence bands $\pm 2\tilde{\sigma}(t)$, where $\tilde{\sigma}^2(t) = \hat{\lambda}\tilde{\sigma}^2(t-1) + (1-\hat{\lambda})\tilde{a}^2(t)$.

Given the presence of many significant coefficients, there is a need to check whether the reduction of the statistic Q_T is significant. Because the stationary model (6) is encompassed by the adaptive system in Table 3, and offline and online innovations are independent Gaussian, standard F tests may be applied. Doing so, we obtain $\hat{F} = 13.64$, which is 1% significant because $F_{1\%}(7,280) = 2.71$. Unlike in the medical application, here we have a system that becomes increasingly nonstationary and for which adaptive estimation is unable to provide stationary innovations. But apart from a couple of outliers, the conclusion of stationarity may be accepted for the recursive residuals $\hat{a}(t)$ in Figure 5b.

Concerning the nature of the change in the parameters of Figure 4, some hypotheses have been already introduced by Young (1984, 1985), who fitted a simplified version of the gas furnace model with a KF based on instrumental variables. A number of possibilities exist, such as (a) bad data quality, in particular systematic measurement errors over the last 50 samples; (b) a progressive change in the physical or chemical characteristics of the inputs; and (c) nonlinearities in the dynamics of the system (e.g., of bilinear type), which may arise as a consequence of a change in the temperature of the plant. In particular, if temperature rises, then the volume of the input gases increases, and it is difficult to control the desired CO_2 concentration.

Table 3. CLS Estimates of the Coefficients of Algorithm (7) with $\kappa = 1/2$

$\hat{\zeta}_t$	$-\hat{\omega}_0(0)$	$-\hat{\omega}_1(0)$	$-\hat{\omega}_2(0)$	$\hat{\delta}_1(0)$	$\hat{\phi}_1(0)$	$-\hat{\phi}_2(0)$	$\hat{\gamma}_0$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\gamma}_1$	$\hat{\delta}$	$-\hat{\alpha}_1$	$\hat{\alpha}_2$	Q_T
$\mathbf{z}_t$	.853 (4.5)	.270 (1.1)	.332 (1.2)	.509 (3.0)	1.08 (5.6)	.130 (.6)	.0165 (1.8)	.974 (125)	.167 (2.9)	.49E − 4 (.7)	.00061 (2.1)	.664 (1.9)	.232 (.9)	12.43
ξ_t	.790 (3.5)	.271 (1.0)	.309 (1.1)	.548 (3.8)	1.17 (3.1)	.272 (.9)	.0003 (.1)	.956 (60.1)	.441 (1.8)	.45E − 4 (.7)	.00151 (.7)	.572 (2.1)	.221 (.8)	12.94

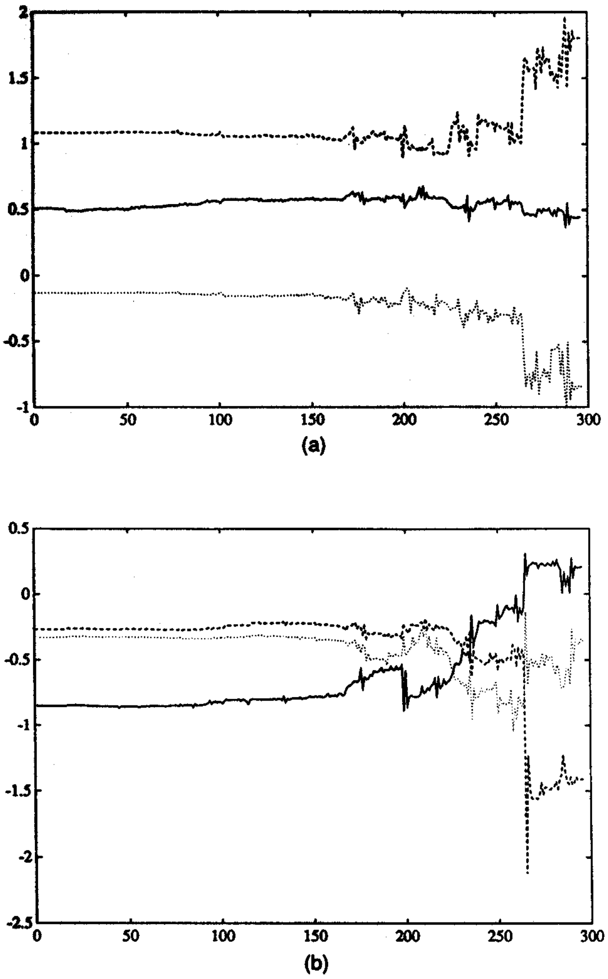

Figure 4. Recursive Estimates Generated by the Coefficients in Row 1 of Table 3: (a) (δ_1 ——, ϕ_1 - - -, ϕ_2 · · · · ·), (b) (ω_0 ——, ω_1 - - -, ω_2 · · · · ·).

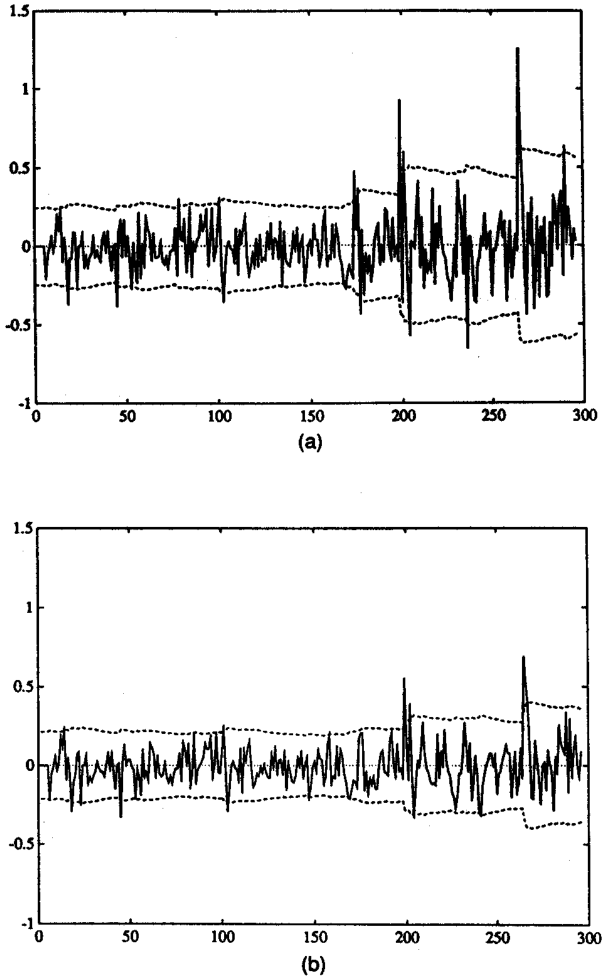

Figure 5. Graphs of (a) Prediction Errors $\tilde{a}(t)$ and (b) Recursive Residuals $\hat{a}(t)$.

4. AN ECONOMIC APPLICATION

The third case study concerns the relationships between short-term (Y) and long-term (X) interest rates. According to Keynes's economic theory, the variable X should determine Y, based on the fact that it has lower variability. To check this hypothesis, we have considered two time series: X = "interest rate on 3-month loans in the money market" and Y = "yields on bonds outstanding for total fixed interest securities" for West Germany (Frankfurt main market), for the period January 1960–December 1987. Monthly data, for a total $T = 336$, were provided by the Deutsche Bundesbank (see Lütkephol 1990, p. 505).

As in the medical application, here stationarity in mean of X_t and Y_t was achieved by differencing. The resulting series x_t and y_t are displayed in Figure 6, showing a clear situation of nonstationarity in covariance. The big outlier y_{255} was replaced by $\bar{y}_{255} = \frac{1}{2}(y_{254} + y_{256})$.

Sample correlation functions are given in Table 4 and lead to the TF model

$$y_t = \begin{pmatrix} .806 \\ (7.8) \end{pmatrix} \begin{pmatrix} 1 - .356B \\ (4.3) \end{pmatrix}^{-1} x_t + \begin{pmatrix} 1 + .301B \\ (5.1) \end{pmatrix} \begin{pmatrix} 1 - .405B^{12} \\ (5.9) \end{pmatrix}^{-1} \hat{a}_t,$$

$$Q = \sum_{13}^{336} \hat{a}_t^2 = 54.8. \quad (8)$$

Table 4. Sample Correlation Functions of Economic Series

Lag	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
$r(y_t y_{t-k})$	1	.32	.18	.03	.12	.09	.03	−.03	−.02	−.16	−.03	−.05	.29
$r(x_t x_{t-k})$	1	.48	.08	−.01	.00	.03	.00	−.05	.01	.01	.06	.09	.08
$r(y_t x_{t-k})$	.47	.32	.14	.19	.15	.09	.01	−.08	−.07	−.11	−.10	−.01	.12
$r(x_t y_{t-k})$	.47	.20	.11	.02	−.01	.05	.02	.03	.05	.08	.11	.15	.12

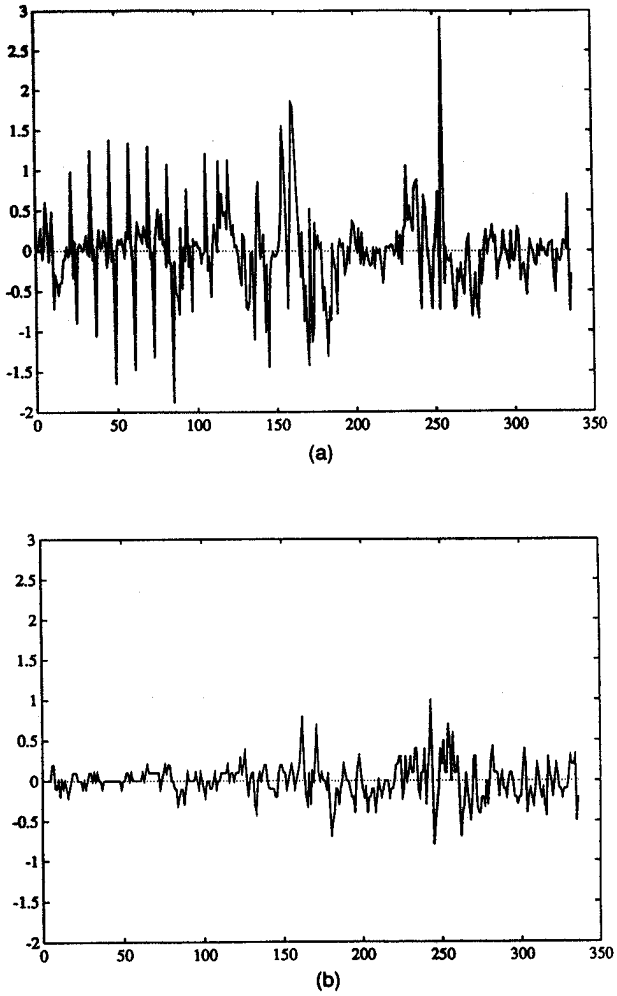

Figure 6. Differenced Series: (a) Short-Term Interest Rate; (b) Long-Term Interest Rate.

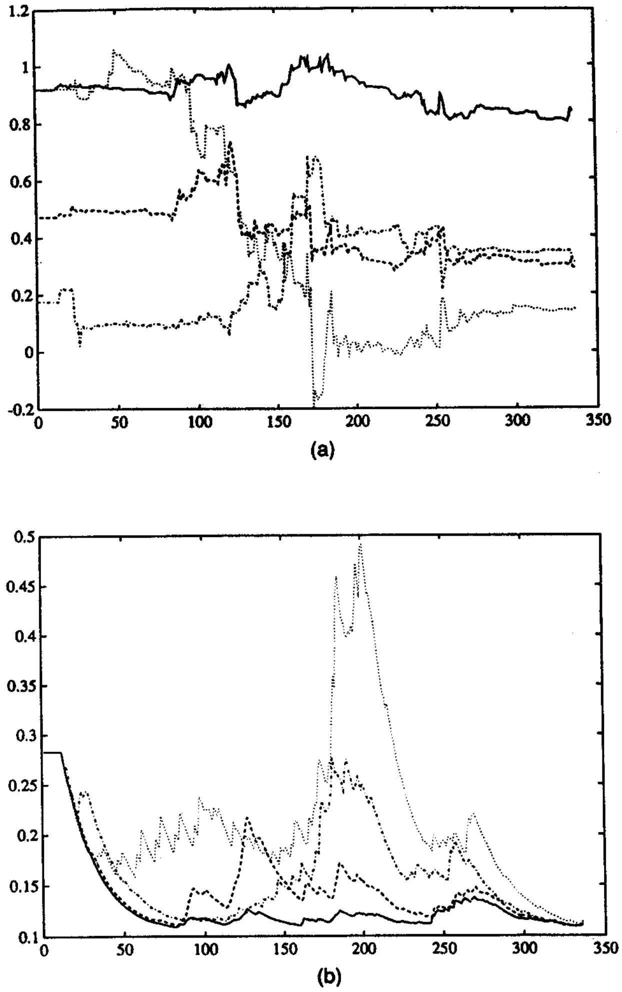

Figure 7. Paths of (a) Parameter Estimates $\hat{\beta}(t)$ and (b) Diagonal Elements of **P***(t), Generated by the Coefficients in Row 2 of Table 5: ($\hat{\omega}_0$ ——; $\hat{\delta}_1$ – – –; $\hat{\phi}_{12}$ · · · · ·; $\hat{\theta}_1$ – · – · –).*

With this, the Keynes's hypothesis of causality, $x_t \Longrightarrow y_t$, is confirmed.

Whereas in the estimation of (8), the outlier at $t = 255$ was dropped for reasons of robustness, in the recursive context this adjustment was necessary to assure minimal tracking properties. Indeed, the CLS estimation of the coefficients of (7) provided a value of $\hat{\lambda} > 1$ together with $\hat{\gamma}_1 = 0$. The "new" observation $\bar{y}_{255}$ has partially solved the problem for γ_1; results for a simplified version of (7) (without α_2, δ) are presented in Table 5. A general feature of the coefficients in Table 5 is that $\hat{\lambda} > 1$ is balanced by $\hat{\mu} < 0$. But this may not be sufficient to keep $\mathbf{P}(t) > \mathbf{O}$, and only for $\zeta_t = \xi_t$ do we have $\hat{\gamma}_1 > 0$ significantly. In any event, the implementation $\zeta_t = \mathbf{z}_t$ still provides the greatest reduction of $Q_T(-17\%)$.

The recursive estimates $\hat{\beta}(t)$ generated by coefficients in row 1 of Table 5 converge (toward constant values), whereas the solution in row 2 shows a sufficient tracking ability. The paths of $\hat{\beta}(t)$ and of the diagonal elements of $\mathbf{P}(t)$ are given in Figure 7; they suggest some remarks:

1. Consistently with Figure 6a, system (8) is stationary at the beginning ($t < 80$) and at the end ($t > 260$) of the sample, although with different patterns.

2. The greatest change is in the noise component, where the AR filter $(1 - \phi_{12}B^{12})$ is gradually replaced by the MA one $(1 + \theta_1 B)$. This may be interpreted as the tendency of short-term variables (e.g., speculative bubbles) to prevail in financial markets.

Table 5. CLS Estimates of the Coefficient of Algorithm (7) Applied to (8)

ζ_t	$\hat{\omega}_0(0)$	$\hat{\delta}_1(0)$	$\hat{\phi}_{12}(0)$	$\hat{\theta}_1(0)$	$\hat{\gamma}_0$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\gamma}_1$	$\hat{\alpha}_1$	Q_T
$\mathbf{z}_t$	.819 (6.1)	.269 (2.7)	1.204 (4.1)	.105 (.7)	.519 (2.9)	1.035 (66.2)	−.557 (.9)	.0002 (1.4)	.651 (1.3)	46.25
ξ_t	.919 (2.1)	.473 (2.8)	.921 (3.1)	.176 (.6)	.283 (1.4)	1.054 (21.3)	−.680 (1.1)	.0052 (2.9)	−.223 (.8)	47.38

The previous outlier adjustment creates a need for robust recursive estimates. Under the assumption of Gaussian innovations, a robustification of algorithm (7) that adopts the two-sigma rule consists of replacing $\tilde{a}(t)$ with $\tilde{a}^*(t) = [\tilde{\mu}(t)\tilde{a}(t)]$, where

$$\tilde{\mu}(t) = 1 \qquad \text{if } |\tilde{a}(t)| < 2\tilde{\sigma}(t-1)$$
$$= 2\tilde{\sigma}(t-1)|\tilde{a}(t)|^{-1} \qquad \text{if } |\tilde{a}(t)| \geq 2\tilde{\sigma}(t-1), \quad (9a)$$

$$\tilde{\sigma}^2(t) = \lambda\tilde{\sigma}^2(t-1) + (1-\lambda)[\tilde{\mu}(t)\tilde{a}(t)]^2, \qquad \tilde{\sigma}^2(0) = \sigma_0^2, \quad (9b)$$

and $\tilde{\sigma}^2(t)$ is a robust adaptive estimator of the innovation variance. Filter (9) must be placed before (7) and requires a suitable initial value of σ_0^2. Robustness aside, it tends to smooth recursive estimates and thus is useful in tracking slowly varying parameters.

5. TWO SIMULATION STUDIES

In this section we check, with small-sample Monte Carlo experiments, the effectiveness of the previous methods in tracking the parameters of evolving ARX models. Simulation consists of 30 independent realizations of length $T = 200$ from a system

$$y_t = \phi_t y_{t-1} + \omega_t x_{t-1} + a_t, \qquad x_t = .7x_{t-1} + e_t, \qquad a_t, e_t \sim \text{IN}(0, 1), \quad (10)$$

in which ϕ_t, ω_t are functions of time that change smoothly and suddenly.

The first experiment considers relatively smooth parameter functions, such as

$$\phi_t = -7 \cdot \left[\frac{\sin(.15(t-100+35))}{(t-100+35)} - \frac{\sin(.15(t-100-35))}{(t-100-35)}\right],$$
$$\omega_t = \frac{-100{,}000(t-100)}{2{,}000{,}000 + (t-100)^4}. \quad (11)$$

The plots of these functions are given in Figure 8, c and d; Figure 8a provides a typical realization of the implied process y_t. The marked nonstationary pattern is mainly attributable to the excursions of the AR parameter ϕ_t outside the stability region $(-1, +1)$. The algorithm used for fitting, with criterion

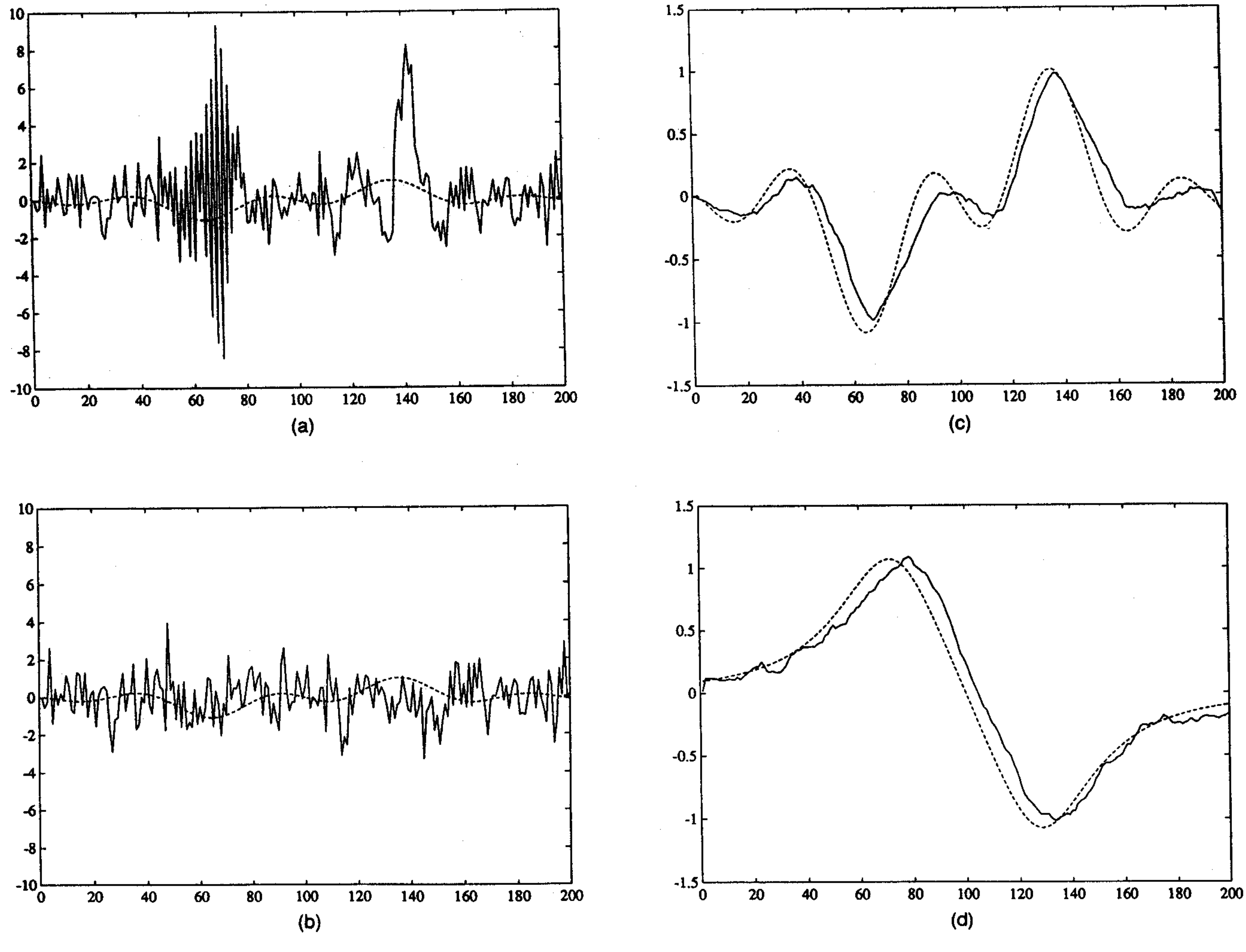

Figure 8. Plots of (a) a Realization of the Process (10)–(11), (b) Its Recursive Innovations, and (c, d) Estimates $\tilde{\phi}_t$, $\tilde{\omega}_t$ (——) and Functions ϕ_t, ω_t (- - -).

Table 6. Mean Values (and Standard Errors) of the Coefficients of Algorithm (2), Estimated with Criterion (3), in 30 Independent Replications of System (10)

Model	$\bar{\gamma}$	$\bar{\phi}_0$	$\bar{\omega}_0$	$\bar{\lambda}$	$\bar{\alpha}$	$\bar{Q}_T$	$\bar{Q}_{OLS}$	S_ϕ	S_ω
(11)	.0096 (.0001)	.0168 (.0127)	.1103 (.0232)	.9651 (.0037)	.3123 (.0552)	274	828	7.8	5.7
(12)	.0017 (.0001)	.5754 (.0185)	−.6310 (.0326)	.9866 (.0025)	.4204 (.3281)	233	452	3.4	1.9

(3), the various replications was (2)–(9) with the constraints $(\gamma_0 = \gamma_1) = \gamma$ and $\alpha = (1 + \alpha B)$, which introduces multistep.

Estimation was carried out with the FMINS routine of the MATLAB package, which is based on the simplex method and is somewhat time-consuming. Table 6 reports mean values and standard errors of the coefficients in 30 replications and the mean value of Q_T in OLS regressions. It also provides the statistics S_ω and $S_\phi = \sum_{t=1}^{200} (\bar{\phi}_t - \phi_t)^2$, where $\bar{\phi}_t = 30^{-1} \sum_{i=1}^{30} \hat{\phi}_{it}$ is the mean value of the recursive estimates. Figures 8c and 8d show the paths of $\bar{\phi}_t$ and $\bar{\omega}_t$ (which are close to the parameter functions), and Figure 8b shows the recursive innovations (which are nearly stationary) of the series y_t in Figure 8a. Because the OLS estimates of the parameters of model (10) approach 0 (for ϕ_t and ω_t moving around 0), OLS innovations do coincide with the series y_t.

The second experiment deals with suddenly changing parameters, such as

$$\begin{aligned} \phi_t &= .6, \quad t = 1\ldots50,\ 151\ldots200 \\ &= .9, \quad t = 51\ldots100 \\ &= .3, \quad t = 101\ldots150, \\ \omega_t &= -.60, \quad t = 1\ldots25,\ 176\ldots200 \\ &= -.95, \quad t = 51\ldots75 \\ &= -.25, \quad t = 151\ldots175. \end{aligned} \tag{12}$$

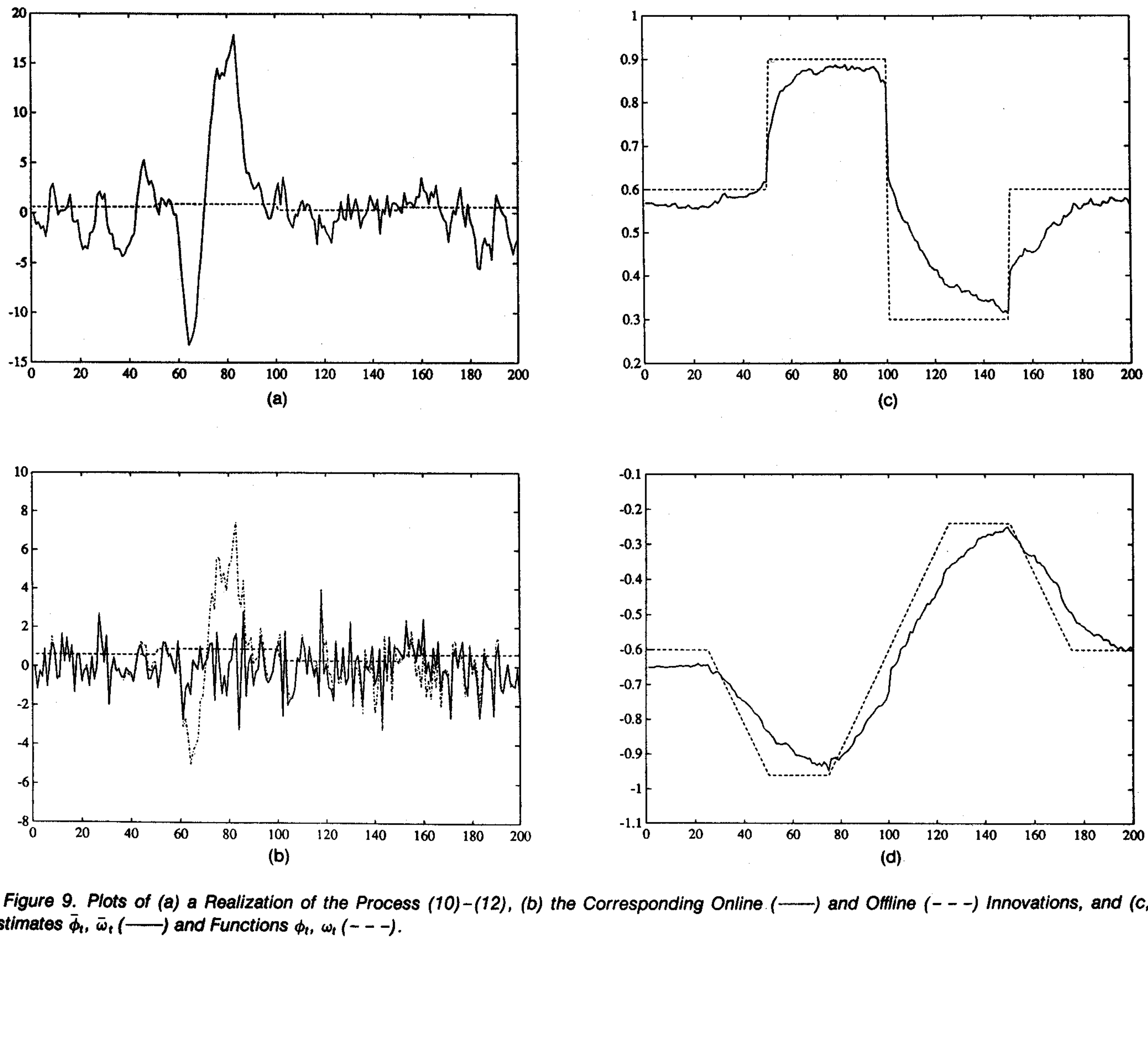

Figure 9. Plots of (a) a Realization of the Process (10)–(12), (b) the Corresponding Online (——) and Offline (- - -) Innovations, and (c, d) Estimates $\bar{\phi}_t$, $\bar{\omega}_t$ (——) and Functions ϕ_t, ω_t (- - -).

These functions are displayed in Figure 9,c and d; Figure 9a shows a typical realization of the implied process (10). The algorithm used for fitting, with criterion (3), the various replications was (2) with the constraint $\gamma_0 = \gamma_1$ and with variable tracking coefficients. A rule to make adaptive the discounting factor of RLS algorithms has been discussed by Fortesque, Kershenbaum, and Ydstie (1981) and Bittanti et al. (1989); it consists of creating an inverse relationship between factor and innovations. Applying this principle to (2), we may define

$$\lambda_t = [1 - (1 - \lambda)\tilde{a}_t^2], \qquad \gamma_t = \gamma \cdot \tilde{a}_t^2 \tag{13}$$

which have a similar effect on the gain matrix. In practice, as $|\tilde{a}_t|$ worsens, $\mathbf{P}_t$ increases so as to improve the tracking capability of the algorithm.

Estimation results are presented in row 2 of Table 6, and Figure 9,c and d, plot the mean values of recursive estimates. The general indication is that the algorithm (2)–(13) is quite good in tracking jumps of medium size, such as ±.3. Finally, Figure 9b compares the innovations of the series in Figure 9a obtained with the adaptive algorithm with those of an OLS regression on the model (10). As in the previous experiment, we may check that recursive innovations are nearly stationary and have a variance significantly lower than that of the OLS ones. This property is systematically confirmed by statistics $\bar{Q}$ in Table 6; the different sizes of Q and S in the two experiments is mainly due to the different range of variation of the parameter functions (11)–(12).

6. CONCLUSIONS

In this article we have considered an adaptive regression methodology based on the unification and optimization of existing recursive algorithms. Using several numerical examples, we have demonstrated its ability to track the parameters of evolving dynamic models and to improve predictions. General indications of the various applications can be summarized as follows:

1. In all the CLS fits, the most significant tracking coefficient was the RLS component λ, rather than the LMS component α or the KF component γ. This can be explained by the fact that it updates the gain matrix $\mathbf{P}_t$ in a more flexible and adaptive way.
2. In the application of Algorithm (7) to TF models, the pseudolinear implementation (approximate gradient) is more effective than that of Gauss–Newton in minimizing the statistic Q_T. The reason is that computation of the exact gradient $\xi_t = \mathbf{G}(B)\mathbf{z}_t$ slows down the adaptation speed of the algorithm.
3. Optimal values of α, λ, and γ and β_0 and γ_0 can seldom be determined on the basis of a priori information. Heuristic design of these coefficients is dangerous, because the value of Q_T may turn out to be greater than that of constant parameter models. On the other hand, statistical properties of CLS estimates have not been entirely investigated (see Tjostheim 1986).

The general conclusion at this point is that adaptive methods represent important tools in the analysis of nonlinear and nonstationary time series. Nonetheless, further developments are still possible, leaving this field of study substantially open.

APPENDIX: BACKGROUND

We review known results that lead to Algorithm (2) and Criterion (3).

KF

Given a regression model $y_t = \beta_t'\mathbf{x}_t + a_t$ with time-varying parameters, the typical assumption for parameter dynamics is the random walk. Such an assumption has the advantage of simplicity and flexibility and leads to a dynamical system that has a state-space interpretation:

$$\beta_t = \beta_{t-1} + \mathbf{e}_t, \qquad \mathbf{e}_t \sim \text{IN}(\mathbf{0}, \mathbf{\Sigma}) \tag{A.1a}$$

$$y_t = \beta_t'\mathbf{x}_t + a_t, \qquad a_t \sim \text{IN}(0, \sigma^2). \tag{A.1b}$$

Given values for the coefficients σ^2, $\mathbf{\Sigma}$ (a positive definite matrix), the optimal MSE estimator of the unobservable sequence $\{\beta_t\}$ is the simplified KF (see Goodwin and Sin 1984)

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \mathbf{P}_t\mathbf{x}_t(y_t - \mathbf{x}_t'\hat{\beta}_{t-1}) \tag{A.2a}$$

$$\mathbf{P}_t = \mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{x}_t\mathbf{x}_t'\mathbf{P}_{t-1}}{\sigma^2 + \mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t} + \mathbf{\Sigma}, \tag{A.2b}$$

where $\mathbf{P}_t$ is the dispersion matrix of $\hat{\beta}_t$. Assuming that $\beta_{t_0} \sim \text{N}(\beta_0, \mathbf{P}_0)$, suitable starting values for the filter (A.2) are β_0 and $\mathbf{P}_0$. To reduce the number of coefficients to be specified a priori or to be estimated, the covariance matrices are usually designed as $\mathbf{\Sigma} = \gamma_1\mathbf{I}$ and $\mathbf{P}_0 = \gamma_0\mathbf{I}$. This may be admissible when the regressors x_{it} have the same scale.

RLS

The previous framework becomes problematic when the parameters are deterministic or their dynamics are nonlinear and generally unknown. In this case, it is necessary to use the approach of nonparametric regression, in which the stress is on the design of suitable smoothers and local estimators. Local regression obtained by discounting observations with exponential weights $\hat{\beta}_t = (\sum_1^t \lambda^{t-i}\mathbf{x}_i\mathbf{x}_i')^{-1}(\sum_1^t \lambda^{t-i}\mathbf{x}_i y_i)$, $0 < \lambda < 1$ is preferable to the method of sliding windows (rolling regression). The reason is that the exponential profile gives more weight to recent observations and, therefore, it may track sudden changes and nonlinear oscillations. Moreover, it is easy to manage recursively (see Ljung and Söderström 1983):

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \mathbf{P}_t\mathbf{x}_t(y_t - \mathbf{x}_t'\hat{\beta}_{t-1}) \tag{A.3a}$$

$$\mathbf{P}_t = \frac{1}{\lambda}\left[\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{x}_t\mathbf{x}_t'\mathbf{P}_{t-1}}{\lambda + \mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t}\right], \tag{A.3b}$$

starting from suitable initial values $\hat{\beta}_0$ and $\mathbf{P}_0$. The resemblance of Algorithms (A.2) and (A.3) is apparent; they differ only in the way the matrix $\mathbf{P}_t$ is prevented from tending to 0. The method used by the latter is more adaptive and depends only on a scalar λ.

As stated before, the approach of local regression is similar to nonparametric estimation; however, in the case of exponential weights there exists a functional interpretation. Because Algorithm (A.3) may be obtained from (A.2b) by setting $\sigma^2 = \lambda$ and

$$\mathbf{\Sigma}_t = \left(\frac{1}{\lambda} - 1\right)\left[\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{x}_t\mathbf{x}_t'\mathbf{P}_{t-1}}{\lambda + \mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t}\right] = \left(\sum_{i=1}^{t} \lambda^{t-i}\mathbf{x}_i\mathbf{x}_i'\right)^{-1} - \left(\frac{1}{\lambda}\sum_{i=1}^{t} \mathbf{x}_i\mathbf{x}_i'\right)^{-1},$$

the "implicit" assumption of parameter evolution underlying EW-RLS is the random walk (A.1a) with nonstationary input $\mathbf{e}_t \sim \text{IN}(\mathbf{0},$

Σ_t) and $a_t \sim \mathrm{IN}(0, \lambda)$. In any event, the meaning of the implied state-space system is difficult to understand, and we prefer to regard local regression as a method consistent with the nonparametric assumption $\beta_t = f(y_{t-k}, \mathbf{x}_{t-k}; k > 0)$.

CLS

In econometrics the coefficients of system (A.1) are estimated efficiently by the method of maximum likelihood (see Pagan 1980). The log-likelihood function has the form

$$-\log L_T(\sigma^2, \Sigma) \propto \left[\sum_{t=1}^{T} \frac{(y_t - \mathbf{x}_t'\hat{\beta}_{t-1})^2}{\sigma^2 + \mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t} + \sum_{t=1}^{T} \log(\sigma^2 + \mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t) \right], \quad (\text{A.4})$$

which is the sum of two incompatible cost functions: the sum of squared standardized innovations, and the sum of their variances in logarithm. The filter (A.2) provides the necessary tool for computing the elements of (A.4), but introduces further coefficients to estimate: β_0 and $\mathbf{P}_0$.

Now, if the KF approaches the steady-state (i.e., $\mathbf{P}_t \to \bar{\mathbf{P}}$, constant) and the regressors $\mathbf{x}_t$ are fixed, then the variance of innovations may be approximated by a constant. Consequently, (A.4) becomes proportional to the sum of squared prediction errors (see Grillenzoni 1993). This functional is defined autonomously in the CLS estimation (see Tjostheim 1986) and may be used for selecting the coefficients of (A.3), namely

$$Q_T(\lambda, \beta_0, \gamma_0) = \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\hat{\beta}_{t-1})^2, \quad (\text{A.5})$$

in which the algorithm of computation is (A.3) itself with $\hat{\beta}_0 = \beta_0$ and $\mathbf{P}_0 = \gamma_0\mathbf{I}$.

[Received February 1989. Revised December 1993.]

REFERENCES

Benveniste, A. (1987), "Design of Adaptive Algorithms for the Tracking of Time-Varying Systems," in *Advances in Statistical Signal Processing* (Vol. 1), ed. V. H. Poor, London: JAI Press, pp. 15–52.

Bittanti, S., Bolzern, P., and Campi, M. (1989), "Adaptive Identification via Prediction-Error Directional-Forgetting Factor," *International Journal of Control,* 50, 2407–2421.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.

Fortesque, T. R., Kershenbaum, L. S., and Ydstie, B. E. (1981), "Implementation of Self-Tuning Regulators with Variable Forgetting Factors," *Automatica,* 17, 831–835.

Goodwin, C. G., and Sin, K. S. (1984), *Adaptive Filtering, Prediction and Control,* Englewood Cliffs, NJ: Prentice-Hall.

Grillenzoni, C. (1993), "ARIMA Processes with ARIMA Parameters," *Journal of Business and Economic Statistics,* 11, 235–250.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and its Applications,* New York: Academic Press.

Härdle, W., Hall, P., and Marron, J. S. (1988), "How Far are Automatically Chosen Regression Smoothing Parameters From Their Optimum?," *Journal of the American Statistical Association,* 83, 86–94.

Harvey, A. C. (1989), *Forecasting, Structural Time-Series Models and the Kalman Filter,* Cambridge, UK: Cambridge University Press.

Klimko, L. A., and Nelson, P. I. (1978), "On Conditional Least Squares Estimation for Stochastic Processes," *The Annals of Statistics,* 6, 629–642.

Ljung, L., and Gunnarsson, S. (1990), "Adaptation and Tracking in System Identification—A Survey," *Automatica,* 26, 7–21.

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

Lütkephol, H. (1990), *Introduction to Multiple Time Series Analysis,* Berlin: Springer-Verlag.

Pagan, A. (1980), "Some Identification and Estimation Results for Regression Models With Stochastically Varying Parameters," *Journal of Econometrics,* 13, 341–363.

Salgado, M. E., Goodwin, C. G., and Middleton, R. H. (1988), "Modified Least Squares Algorithm Incorporating Exponential Resetting and Forgetting," *International Journal of Control,* 49, 477–491.

Tjostheim, D. (1986), "Estimation of Non-Linear Time Series Models," *Stochastic Processes and Their Applications,* 21, 225–273.

Tong, H. (1990), *Non-Linear Time Series Analysis: a Dynamical System Approach.* Oxford, UK: Clarendon Press.

Widrow, B., and Stearns, S. (1985), *Adaptive Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall.

Young, P. C. (1984), *Recursive Estimation and Time Series Analysis,* Berlin: Springer-Verlag.

——— (1985), "Recursive Identification, Estimation and Control," in *Handbook of Statistics* (Vol. 5), eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, Amsterdam: North-Holland, pp. 213–255.