

TIME-VARYING PARAMETERS PREDICTION

CARLO GRILLENZONI

*IUAV: University Institute of Architecture of Venice,
St. Croce 1957, 30135 Venezia, Italy*

(Received February 12, 1997; revised August 31, 1998)

Abstract. This paper develops a method of adaptive modeling that may be applied to forecast non-stationary time series. The starting point are time-varying coefficients models introduced in statistics, econometrics and engineering. The basic step of modeling is represented by the implementation of adaptive recursive estimators for tracking parameters. This is achieved by unifying basic algorithms—such as recursive least squares (RLS) and extended Kalman filter (EKF)—into a general scheme and next by selecting its coefficients with the minimization of the sum of squared prediction errors. This defines a non-linear estimation problem that may be analyzed in the context of the conditional least squares (CLS) theory. A numerical application on the IBM stock price series of Box-Jenkins illustrates the method and shows its good forecasting ability.

Key words and phrases: Conditional least squares, extended Kalman filter, IBM stock price series, recursive least squares, time-varying parameter models.

1. Introduction

Adaptive prediction is a forecasting technique for non-stationary time series which is based on the recursive estimation of the models of the series. Given a model with time-varying coefficients and a recursive algorithm, the adaptive predictor may be implemented by using the latest parameter estimate in the forecasting function of the model (see Zellner *et al.* (1990)). In general, no law of evolution is assumed and the approach is open to a large variety of recursive methods.

A selection principle for recursive estimators is provided by their tracking capability. The ability to approach the trajectory of parameters is determined by three factors:

- 1) the structure of the *recursive algorithm*, that is the statistical philosophy of the method;
- 2) the kind of *adaptation mechanism*, that is the form of weighting observations;
- 3) the design of the *tracking coefficients* (hyperparameters in the bayesian terminology).

In engineering literature there have been several attempts to unify recursive algorithms (see Ljung and Söderström (1983), Goodwin and Sin (1984)); however, the fundamental question of designing hyperparameters has been treated heuristically (e.g. Salgado *et al.* (1988)).

Statistical science has a solid tradition on the estimation of hyperparameters; see Pagan (1980), Kitagawa and Gersh (1985) and Grillenzoni (1993) in the context of the Kalman Filter. This strategy may be extended to general adaptive filters by defining an objective function based on squared prediction errors (Grillenzoni (1994)). Such an approach belongs to the conditional least squares (CLS) estimation for stochastic processes discussed by Klimko and Nelson (1978), Tjostheim (1986) and Tong (1990).

The paper is organized as follows: Section 2 discusses time-varying parameters with reference to ARMA models. Section 3 presents an adaptive estimator that encompasses many existing recursive algorithms. Finally, in Section 4 the adaptive framework is illustrated with an application to the IBM series of Box and Jenkins (1976).

2. Evolving ARMA models

A non-stationary process $\{z_t\}$ having an auto-regressive moving average (ARMA) representation with time-varying parameters, can be expressed as follows

$$(2.1) \quad z_t = \mu_t + \sum_{i=1}^p \phi_{it} z_{t-i} + \sum_{j=1}^q \theta_{jt} a_{t-j} + a_t, \quad a_t \sim \text{IID}(0, \sigma^2), \quad t = 0, 1, 2, \dots$$

where μ_t is a drift and a_t is the input. We suppose that the law of evolution of the parameters is unknown, in particular they may be stochastic or deterministic and can depend on a_t or t , or both; however, to allow for statistical treatment of (2.1), two fundamental assumptions are needed:

A1) The parameters $\beta_{kt} = \{\mu_t, \phi_{it}, \theta_{jt}\}$ with $k = 1, 2, \dots, (1 + p + q)$, are smooth functions. For example, they could be a mixture of random walks: $\beta_{kt} = \beta_{kt-1} + e_{kt}$, with $\text{var}(e_{kt})$ small, and polynomials of the time: $\beta_{kt} = \alpha_{0k} + \alpha_{1k}t + \alpha_{2k}t^2 + \dots$; and so on.

A2) The coefficients of the input process are constant, in particular the variance $\sigma^2 < \infty$; this means that the final purpose of a time-varying parameter modeling is to obtain stationary innovations.

This goal can be achieved by modeling the parameters in the proper way, or by using a suitable adaptive estimator for tracking their trajectory. This point will be discussed in the next Section.

Stability. Stability is a useful feature of stochastic models because it is a sufficient (although non-necessary) condition for optimal properties of parameter estimates and forecasts. As a general definition of *stochastic stability* we adopt the principle that to inputs $\{a_t\}$ bounded in probability, there must correspond outputs $\{z_t\}$ with the same feature:

$$(2.2) \quad \sup_{t>0} P(|z_t| > x) = o(1/x^\kappa) \quad \text{as } |x| \rightarrow \infty, \quad \text{with } \kappa > 0.$$

The above is stronger than the *stability in probability* defined as $\lim_{x \rightarrow \infty} \sup_{t>0} P(|z_t| > x) = 0$. This condition, sometimes called *tightness* (see Kushner (1990)), is minimal and ensures the existence of an invariant probability measure. However, the need to have $\kappa > 0$ in (2.2) follows by the fact that it is sufficient for the existence of moments of order $k < \kappa$, that is $\sup_{t>0} |E(z_t^\kappa)| < \infty$.

Now, sufficient conditions for the stochastic stability of (2.1) may be obtained by solving the model for $\{z_t; t > 0\}$ and extending those established in the constant parameter case. In particular, if S_p is the parameter space of a stationarity AR(p) process (e.g. S_2 is the well known triangular region), then (2.1) has second order moments ($\kappa = 2$) if all sample realizations of $\{\phi_{it}\}$ lie inside S_p .

PROPOSITION 1. *The non-stationary process (2.1) is second order, if the polynomials $\phi_t(B) = 1 + \phi_{1t}B + \dots + \phi_{pt}B^p$ have roots whose realizations entirely lie outside the*

unit circle, with the exception, at most, of finite sets of points. The model (2.1) is invertible (i.e. $\{a_t\}$ can be generated from $\{z_t\}$) if a similar feature holds for the polynomials $\theta_t(B) = 1 + \theta_{1t}B + \dots + \theta_{qt}B^q$.

PROOF. The result is intuitive and can be proved as in Pourahmadi (1986), Grillenzoni (1990) or Chen and Tsay (1993). First we consider the model $z_t = \phi_t z_{t-1} + a_t$; by assumption, there is a fixed sequence ϕ_t^* such that

$$|\phi_t| \leq \phi_t^* = \begin{cases} \Phi < \infty, & t \in T < \infty \\ \phi < 1, & \text{elsewhere} \end{cases} \quad \text{w.p.1}$$

with probability one (w.p.1). Thus z_t is bounded by the process $z_t^* = \phi_t^* z_{t-1}^* + a_t$ and solving for $t > 0$ with initial condition $z_0^* = a_0$ we find that

$$z_t^* = \sum_{i=0}^t \psi_{it}^* a_{t-i}, \quad \psi_{it}^* = \left(\prod_{j=1}^i \phi_{t-j+1}^* \right).$$

Since $\sup_{t>0} \psi_{it}^* = O(\Phi^T \cdot \phi^{|i-T|}) = O(\phi^{|T-i|})$, it follows that $\sup_{t>0} \sum_{i=0}^t |\psi_{it}^*|^2 < \infty$, thus z_t^* has finite second order moments, and the same occurs for the process z_t .

For the extension to AR(p) models, we consider polynomials $\phi_t^*(B) = \prod_{i=1}^p (1 - \rho_{it}^* B)$ (where B is the backward-shift operator: $B^k z_t = z_{t-k}$), whose inverse roots ρ_{it}^* are fixed and bound those of the polynomials $\phi_t(B)$ of model (2.1). As in the simpler case we have the condition

A3)

$$(2.3) \quad |\rho_{it}| \leq \rho_{it}^* = \begin{cases} P_i < \infty, & t \in T_i < \infty \\ \rho_i < 1, & \text{elsewhere.} \end{cases} \quad \text{w.p.1.}$$

The process $\{z_t\}$ is still bounded in probability by $\phi_t^*(B)z_t^* = a_t$, which has finite variance. In fact, we note that the sub-processes $z_{it}^* = (1 - \rho_{it}^* B)^{-1} z_{i-1t}^*$, with $z_{0t}^* = a_t$ and $z_{pt}^* = z_t^*$, are all second order. An algorithm for obtaining the weights $\{\psi_{it}^*\}$, in this case is given in Grillenzoni (1990).

Remark. A complete characterization of the stochastic stability for Markov processes was provided by Meyn and Tweedie (1993). It includes the definition (2.2) as a particular case. Since the model (2.1) can be expressed in Markovian form, it is possible to show that Proposition 1 allows many other stability properties, such as irreducibility, recurrence, regularity, non-evanescence and tightness. These are reviewed in Appendix 1.

Unstability. In many real situations, there may be the need to deal with non-stationary processes $\{Z_t\}$ which are also *stochastically unstable*, i.e. asymptotically unbounded. This may be achieved by including in (2.1) a linear factor whose parameter wanders outside the stability region:

$$(2.4) \quad (1 - \Phi_t B)Z_t = z_t, \quad \bar{E}|\Phi_t| = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T E|\Phi_t| \right) \geq 1.$$

The fluctuations of the unstable root Φ_t can determine in Z_t trends and cycles with complex transitory components, such as structural breaks and inverting slopes. Together

with the sets T_i in (2.3), we may conceive evolving models in which stable roots become locally unstable and vice-versa.

Prediction. The cost for the adaptability of the above framework, is in forecasting. Since the dynamic of parameters β_{kt} is unknown, the expression of the predictor $\hat{z}_t(h) = E[z_{t+h} | z_t, z_{t-1}, \dots]$, $h > 1$ cannot be derived. An approximate solution can be obtained from the formula for stationary models

$$\hat{z}_t(h) \approx \hat{\mu}_t(h-1) + \sum_{i=1}^{h-1} \hat{\phi}_{it}(h-1)\hat{z}_t(h-i) + \sum_{i=h}^p \hat{\phi}_{it}(h-1)z_{t+h-i} + \sum_{j=h}^q \hat{\theta}_{jt}(h-1)\hat{a}_{t+h-j}$$

where $\hat{\mu}_t(h-1) = E[\mu_{t+h} | z_t, z_{t-1}, \dots]$, etc... In adaptive forecasting, the parameter predictors $\hat{\beta}_{kt}(h-1)$ are usually approximated by the latest recursive estimates $\hat{\beta}_k(t)$. This solution may be improved, if on the basis of such estimates, models for $\{\beta_{kt}\}$ are built and used in forecasting.

3. Optimized adaptive estimation

In the previous section we have introduced a class of models whose parameters are time-varying. Since their dynamic is unknown, the classical solution of the Kalman Filter (KF), and its extensions (e.g. Kitagawa (1987)), cannot be used in estimation. Therefore the resort to more general algorithms described in Ljung and Söderström (1983) and Goodwin and Sin (1984) is needed. In any event, the *smoothness* assumption A1 is a necessary condition to have a good statistical performance. This fact cannot be precisely quantified because “convergence” of adaptive estimates is impossible even when parameters are constant. In practice, only bounds on the tracking errors can be achieved, but this depends on the stability of the algorithms (see e.g. Guo and Ljung (1995a, 1995b)).

Algorithms. Writing the model (2.1) as $z_t = \beta'_t x_t + a_t$, with $x'_t = [1, z_{t-1}, \dots, a_{t-q}]$ and $\beta'_t = [\mu_t, \phi_{1t}, \dots, \theta_{qt}]$, the typical recursive estimator for its parameters is

$$(3.1) \quad \hat{\beta}(t) = \hat{\beta}(t-1) + \hat{\Gamma}(t)\hat{\xi}(t)\hat{a}(t)$$

which is obtained from the iterative estimator $\hat{\beta}_T^{(i)}$ by equating number of iterations and number of processed data ($i = T$) = t . In (3.1), $\hat{\Gamma}(t)$ is the covariance matrix (it will be defined below), $\hat{\xi}(t) = [-\partial a_t(\beta_t)/\partial \beta_t]_{\beta_t = \hat{\beta}(t-1)}$ is the gradient and $\hat{a}(t) = [z_t - \hat{x}(t)'\hat{\beta}(t-1)]$ is the prediction error. From the computational viewpoint the vector $\hat{x}(t)' = [1, z_{t-1}, \dots, \hat{a}(t-q)]$ is updated with $\hat{a}(t)$, and having $\xi_t = x_t/\theta_t(B)$, the gradient can be computed on-line as $\hat{\xi}(t) = \hat{x}(t) - \sum_{j=1}^q \hat{\theta}_j(t-1)\hat{\xi}(t-j)$.

What fundamentally distinguishes the various algorithms, is the structure of the matrix $\hat{\Gamma}(t)$. The solutions related to recursive least squares (RLS, see Ljung and Söderström (1983)), extended Kalman filter (EKF, see Goodwin and Sin (1984)) and

least mean squares (LMS, see Widrow and Stearns (1985)) are given by

$$(3.2a) \quad \text{RLS} : \hat{\Gamma}(t) = \frac{1}{\lambda} \left[\hat{\Gamma}(t-1) - \frac{\hat{\Gamma}(t-1)\hat{\xi}(t)\hat{\xi}(t)'\hat{\Gamma}(t-1)}{\lambda + \hat{\xi}(t)'\hat{\Gamma}(t-1)\hat{\xi}(t)} \right], \quad 0 < \lambda < 1$$

$$(3.2b) \quad \text{EKF} : \hat{\Gamma}(t) = \hat{\Gamma}(t-1) - \frac{\hat{\Gamma}(t-1)\hat{\xi}(t)\hat{\xi}(t)'\hat{\Gamma}(t-1)}{\sigma^2 + \hat{\xi}(t)'\hat{\Gamma}(t-1)\hat{\xi}(t)} + \gamma\mathbf{I}, \quad 0 < \gamma < \infty$$

$$(3.2c) \quad \text{LMS} : \hat{\Gamma}(t) = \mu \cdot \frac{1}{[\hat{\xi}(t)'\hat{\xi}(t)]} \mathbf{I}, \quad 0 < \mu < \infty$$

where λ , γ , μ are adaptation coefficients and $\sigma^2 = E(a_t^2)$. It is well known that (3.2a) corresponds to a sequential Gauss-Newton estimator with weighted covariance matrix $\hat{\Gamma}(t) = [\sum_i^t \lambda^{t-i} \hat{\xi}(i)\hat{\xi}(i)']^{-1}$, and (3.2c) is its steepest-descent version. On the other hand, (3.2b) implies that parameters follow a random walk model with input $e_t \sim \text{NID}(\mathbf{0}, \gamma\mathbf{I})$. Thus, (3.2a) and (3.2c) are essentially non-parametric methods, suitable when the vector β_t is a non-linear process or a deterministic sequence.

As a general remark, the necessary condition which enables (3.1) to track the changes $[\beta_t - \beta_{t-1}]$ and to have $E\|\hat{\beta}(t) - \beta_t\| < \infty$, is that matrices (3.2a)–(3.2c) be positive definite and bounded: $0 < \hat{\Gamma}(t) < \infty$. As concerned *positivity*, we may note that the coefficients λ , γ , μ have the role of preventing $\hat{\Gamma}(t) \rightarrow \mathbf{0}$. Indeed, having $\Gamma_t = (\sum_i^t \lambda^{t-i} \xi_i \xi_i')^{-1}$, in the RLS the vanishing occurs if the weighting factor $\lambda = 1$, while in the LMS it happens if the step-size $\mu \rightarrow 0$ and in the EKF if the noise variance $\gamma = 0$. This motivates the constraints in (3.2a)–(3.2c). As concerned *boundedness* of $\hat{\Gamma}(t)$, which enables the stability of equation (3.1), it is fundamentally related to its invertibility. In the RLS this requires

$$(3.3) \quad \inf_{t > p} P \left[\det \left(\sum_{i=p+1}^t \lambda^{t-i} \xi_i \xi_i' + \lambda^{t-p} \gamma_0^{-1} \mathbf{I} \right) > 0 \right] = 1$$

where $\gamma_0 \mathbf{I} = \Gamma_0$. Condition (3.3) with $\lambda = 1$ is also sufficient for the exponential stability of the EKF (see Guo and Ljung (1995a), p. 1380). Now, given the relationship $\xi_t = \mathbf{x}_t / \theta_t(B)$, we may see that (3.3) is satisfied under the conditions of Proposition 1 and non-degenerate values of λ , γ_0 .

From (3.2a)–(3.2c) we may see that the various algorithms have apparent algebraic connections. This enables us to recompose them into a *general* recursive estimator, as follows

$$(3.4a) \quad \hat{\beta}(t) = \hat{\beta}(t-1) + \mu \hat{\Gamma}(t) \hat{\xi}(t) [z_t - \hat{\mathbf{x}}(t)' \hat{\beta}(t-1)], \quad \hat{\beta}(0) = \beta_0$$

$$(3.4b) \quad \hat{\Gamma}(t) = \frac{1}{\lambda} \left[\hat{\Gamma}(t-1) - \frac{\hat{\Gamma}(t-1)\hat{\xi}(t)\hat{\xi}(t)'\hat{\Gamma}(t-1)}{1 + \hat{\xi}(t)'\hat{\Gamma}(t-1)\hat{\xi}(t)} \right] + \gamma\mathbf{I}, \quad \hat{\Gamma}(0) = \gamma_0 \mathbf{I}$$

where β_0 , γ_0 are initial values. Algorithm (3.4) encompasses (3.1)–(3.2), in fact setting $\mu = 1$, $\gamma = 0$ we have the RLS, when $\mu = 1$, $\lambda = 1$ we have the EKF and for $\gamma = 0$, $1/\lambda = 0$ one obtains the LMS without the normalization $\|\hat{\xi}(t)\|$. The unification (3.4) is motivated by the aim of improving the tracking capability of (3.1), namely to minimize the distance $E\|\hat{\beta}(t) - \beta_t\|$. This approach is largely tentative, but we recall that the dynamic of parameters is unknown.

Optimization. Algorithm (3.4) involves the unknown coefficients $\delta' = [\mu, \lambda, \gamma; \gamma_0, \beta'_0]$ whose range of variation, except for $0 < \lambda < 1$, is somewhat wide. Until now, in engineering literature, only heuristic rules have been provided for their selection. Using the general bound $0 < \hat{\Gamma}(t) < \infty$, Salgado, Goodwin and Middleton (1988) have established their ranges as follows: $\lambda \in [.95, .99]$, $\mu \in [.1, .5]$, $\gamma \in [.0, .1]$. These constraints are too loose, and do not enable a good design of (3.4). Given a sample of observations $\{z_1, \dots, z_T\}$, it is appropriate to solve the problem in terms of optimal selection, by minimizing a loss function based on the prediction errors $\hat{a}(t)$

$$(3.5) \quad [\hat{\mu}, \hat{\lambda}, \hat{\gamma}; \hat{\gamma}_0, \hat{\beta}'_0]_T = \arg \min_{\delta} \left\{ \hat{Q}_T(\delta) = \sum_{t=p+1}^T [z_t - \hat{x}(t)' \hat{\beta}(t-1)]^2 \right\}.$$

This approach is similar to the optimization of smoothing coefficients of non-parametric estimators (e.g. the cross-validation selection of the bandwidths in kernel-type estimators). However, from a parametric viewpoint it belongs to the conditional least squares (CLS) method discussed in Klimko and Nelson (1978), Tjostheim (1986) and Tong ((1990), Chapter 5), where the term *conditional* concerns the set of past information, because $Q_T(\delta) = \sum_t [z_t - E_{\delta}(z_t | z_{t-1}, \dots)]^2$.

Another correspondence of (3.5) is with the maximum likelihood estimation of the parameters of state-space models, in which the likelihood function is evaluated by means of the Kalman filter (see Kitagawa and Gersh (1985)). However, major differences are the semi-parametric nature of the model (2.1), the use of filter (3.4) in place of the KF and the direct estimation of the initial values $\beta_0, \gamma_0 \mathbf{I}$. The last point, in particular, tends to avoid the drawbacks of transient behaviour and error propagation of the heuristic solutions discussed in the bayesian literature; in particular that of *diffuse priors* for $\hat{\Gamma}(0)$. On the other hand, it may rise specific problems of parametric identifiability (see Pagan (1980)). To prevent situations of this type, parsimonious parametrizations, such as $\gamma = \gamma_0/1000$, or constraints on the vector $\beta'_0 = [\mu_0, \phi_{10}, \dots, \theta_{q0}]$, can be used.

In practical terms, the optimization (3.5) is performed by iterative (off-line) estimators, such as Newton or Gauss-Newton, where the functional $Q_T(\cdot)$ is computed through the recursive filter (3.4). Given the prediction errors $\hat{a}_t(\delta) = [z_t - E_{\delta}(z_t | z_{t-1}, \dots)]$, iterative estimates are

$$(3.6a) \quad \hat{\delta}_T^{(i+1)} = \hat{\delta}_T^{(i)} + \alpha_i \left[\sum_{t=1}^T \hat{\zeta}_t^{(i)} \hat{\zeta}_t^{(i)'} \right]^{-1} \sum_{t=1}^T \hat{\zeta}_t^{(i)} \hat{a}_t^{(i)}, \quad \text{with } \zeta_t = -\frac{\partial \hat{a}_t}{\partial \delta}$$

$$(3.6b) \quad \hat{\delta}_T^{(i+1)} = \hat{\delta}_T^{(i)} + \alpha_i \left[\sum_{t=1}^T \hat{W}_t^{(i)} \right]^{-1} \sum_{t=1}^T \hat{\zeta}_t^{(i)} \hat{a}_t^{(i)}, \quad \text{with } W_t = \frac{\partial^2 \hat{a}_t}{\partial \delta \partial \delta'}$$

where α_i are stepsizes. Unlike the gradient $\{\zeta_t\}$ in (3.1), the analytical expression of the derivatives $\{\zeta_t, W_t\}$ is difficult to obtain and this complicates the statistical analysis (see Pagan (1980)).

Assuming that problems of parametric identifiability (which make the matrices $[T^{-1} \sum_t \zeta_t \zeta_t']$, $[T^{-1} \sum_t W_t]$ asymptotically singular) do not arise, the consistency of (3.6) fundamentally depends on the smoothness of $Q_T(\cdot)$.

PROPOSITION 2. *Under the conditions of Proposition 1, property (3.3) and:*
 B1) $T^{-1} \partial Q_T(\delta_0) / \partial \delta$ converges to zero with probability one (w.p.1).

B2) $M_T(\delta_0) = T^{-1} \partial^2 Q_T(\delta_0) / \partial \delta \partial \delta'$ converges w.p.1 to a positive definite matrix.

B3) $\lim_{T \rightarrow \infty} \sup_{\varepsilon \rightarrow 0} \varepsilon^{-1} |M_T(\delta^*) - M_T(\delta_0)| < \infty$ w.p.1 for $|\delta^* - \delta_0| < \varepsilon$.

the estimates (3.6) converge w.p.1 to the value δ_0 that minimizes (3.5).

PROOF. This result is a corollary of the CLS theorem reviewed in Appendix 2. In particular, Proposition 1 allows for the existence of second order moments and asymptotic independence of $\{z_t\}$. Invertibility in Proposition 1 and property (3.3) enable the algorithm (3.4) to be stable so that the functional $Q_T(\delta)$ is bounded. Conditions B1 and B2 guarantee that asymptotically Q_T has a minimum at δ_0 and B3 that the residual of the Taylor expansion of Q_T in δ_0 can be ignored.

In the application of the CLS theorem to the estimation of the drift θ_0 in the state-space model $z_t = \phi_t z_{t-1} + a_t$ with $\phi_t = \theta_0 + \theta_1 e_{t-1} + e_t$, Tjostheim (1986) required the existence of fourth order moments of $\{z_t\}$. In Appendix 3, we show that such moments are involved in the estimator (3.6), therefore they are required in Proposition 2. However, proceeding as in stationary models, one may show that $\{z_t\}$ is also a fourth order process under the conditions of Proposition 1.

Finally, on the basis of the results of Appendix 3, an extension of the RLS algorithm (3.1) and (3.2a) toward time-varying weighting factors $\{\lambda_t\}$ is proposed in Appendix 4.

4. The IBM case study

The case study concerns the IBM stock price series during the period May 17, 1961–November 2, 1962; the plot of data in levels Z_t and in differences $z_t = (Z_t - Z_{t-1})$ is given in Fig. 1 (a,b). For this series Box and Jenkins ((1976), p. 239) have confirmed the hypothesis of random walk, which is widely diffused in financial data. In fact, the ML-estimation of the identified IMA(1,1) model $z_t = \theta a_{t-1} + a_t$ provided a non-significant MA-parameter. The same data-set has been republished by Tong ((1990), p. 512), together with its calendar. This enables us to obtain the weekly average of the IBM series, which contains $T = 77$ observations and has a richer dynamic structure.

As is known from Weiss (1984), given a process $Z_t \sim \text{ARIMA}(p, d, q)$ (where d is the order of integration), the aggregated series $\bar{Z}_\tau = (Z_t + Z_{t-1} + \dots + Z_{t-k})$ behaves like an $\text{ARIMA}(p, d, r)$ where $r = \text{int}[(p+d+1) + (q-p-d-1)/k]$ and $\text{int}(\cdot)$ takes the integer part. Weiss (1984) also showed that as $k \rightarrow \infty$ the model becomes IMA(d, d); therefore if Z_t is a random walk and k is not small, then \bar{Z}_τ should be an IMA(1,1).

Empirical investigation of the IBM series has provided results more complex than those expected from the previous discussion. Indeed, the models identified for the weekly average in differences and in levels, were ARI(3,1) and ARMA(2,1):

$$(4.1a) \quad \bar{z}_\tau = \underset{(3.5)}{.394} \bar{z}_{\tau-1} - \underset{(1.1)}{.127} \bar{z}_{\tau-2} + \underset{(2.5)}{.292} \bar{z}_{\tau-3} + \bar{a}_\tau, \quad \hat{Q}_{77} = 9,905$$

$$(4.1b) \quad \bar{Z}_\tau = \underset{(12.8)}{1.378} \bar{Z}_{\tau-1} - \underset{(3.7)}{.380} \bar{Z}_{\tau-2} + \underset{(2.1)}{.264} \bar{a}_{\tau-3} + \bar{a}_\tau, \quad \hat{Q}_{77} = 10,005.$$

Unlike the daily series, these models are significantly better, in terms of the statistic Q_T , than the corresponding random walk $\bar{Z}_\tau = .9972 \bar{Z}_{\tau-1} + \bar{a}_\tau$, $\hat{Q}_{77} = 13,332$. As shown by Rao (1961), the OLS estimator applied to models with unstable roots substantially improves its consistency property, although estimates are not asymptotically normal. Hence, there are not counterindications to the implementation of the model (4.1b). The

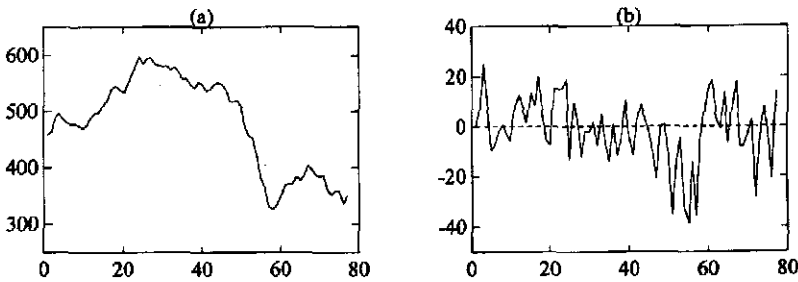


Fig. 1. Plot of the IBM series: (a) in levels; (b) in differences.

Table 1. CLS estimates of the coefficients of algorithm (3.4) applied to models (4.1).

Model	γ_0	λ	μ	ϕ_{10}	ϕ_{20}	ϕ_{30}/θ_{30}	Q_T
(4.1a)	.938494 (140)	.782296 (125)	-1.05758 (31)	.546323 (141)	-.451328 (63)	.068321 (55)	5410
(4.1a)	.354499 (35)	.920214 (68)	-.578551 (18)	.461552 (99)	-.360024 (27)	.098062 (29)	6047
(4.1b)	1.034802 (10.3)	.956084 (18.3)	-.159910 (8.5)	1.356285 (23.5)	-.357421 (5.7)	.151143 (2.5)	8270
AR(1)	.038408 (1.2)	.625269 (10.5)	-.349404 (22.2)	1.012237 (123)			10016

constant parameter models (4.1a) and (4.1b) will be the benchmarks for evaluating the performance of the adaptive techniques discussed in Section 3.

We now apply the adaptive framework (3.4)–(3.5) to the models of the weekly IBM series. The best algorithm structure was tentatively identified without the Kalman filter component, because the estimates of the coefficient γ were non significant. CLS estimations were carried out with the MAXLIK routine of the Gauss package; results are reported in Table 1. The t -statistics (in parentheses) must be considered with caution because the asymptotic normality may not hold for highly nonlinear estimations (Tjostheim (1986)).

We now describe in detail the content of Table 1: Row 1 contains the estimates of the coefficients of algorithm (3.4) with the constraint $\gamma \equiv 0$ and applied to model (4.1a). Row 2 concerns the same experiment but performed with the constraint $\gamma = \gamma_0/100$, which was suggested by the unconstrained estimation. As we may see, the statistical performance slightly worsens. Row 3 deals with the model (4.1b), of the IBM series in levels, and the constraint $\gamma = \gamma_0/1000$. Despite the non-stability of the series $\{\bar{Z}_t\}$, the iterative estimation converged; however, the statistic Q_T is not as good as before. Row 4 concerns the same estimation as in row 3, but done on an AR(1) model. With this experiment we are mainly interested to track the time-path of an unstable root.

Important remarks can be made on Table 1: In all cases we may see that in-sample forecasting performance, expressed by the statistic \hat{Q}_T , significantly improves over the constant parameter models (4.1a) and (4.1b). In contrast with the expectation of the heuristic design of hyperparameters (see e.g. Ljung and Söderström (1983)), the value of the LMS coefficient μ is negative. This is probably associated to non-linear characteristics of the IBM series (see Tong (1990)). The factor λ belongs to the admissible range (0,1), but estimated values do not fall in the set [.95,.99] which is usually recommended. Moreover, as for γ , the coefficient γ_0 does not seem significant.

In summary, the reduction of the statistic Q_T allowed by the method (3.4)–(3.5) over the constant parameter models (4.1a) and (4.1b), ranges from -45% of row 1 to -18%

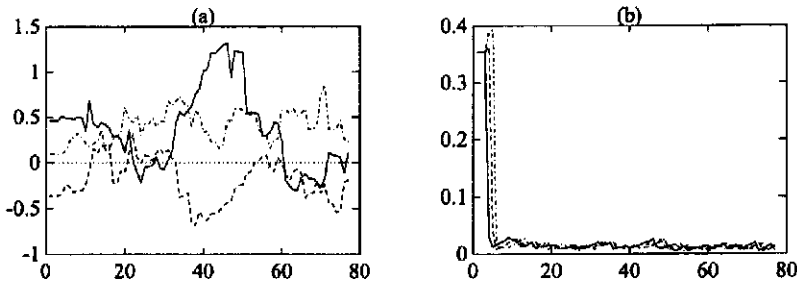


Fig. 2. Recursive estimates generated with the coefficients in row 2 of Table 1: (a) Parameters $\hat{\phi}_1(t)$ —, $\hat{\phi}_2(t)$ ---, $\hat{\phi}_3(t)$ - · -; (b) Variances $\hat{\Gamma}_{ii}(t)$, $i = 1, \dots, 3$.

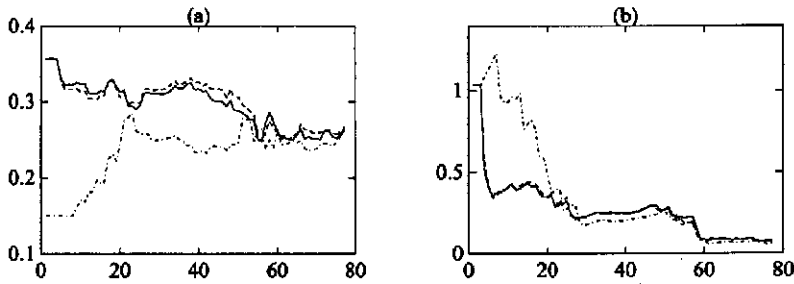


Fig. 3. Recursive estimates generated with the coefficients in row 3 of Table 1: (a) Parameters $[\hat{\phi}_1(t) - 1]$ —, $-\hat{\phi}_2(t)$ ---, $\hat{\phi}_3(t)$ - · -; (b) Variances $\hat{\Gamma}_{ii}(t)$, $i = 1, \dots, 3$.

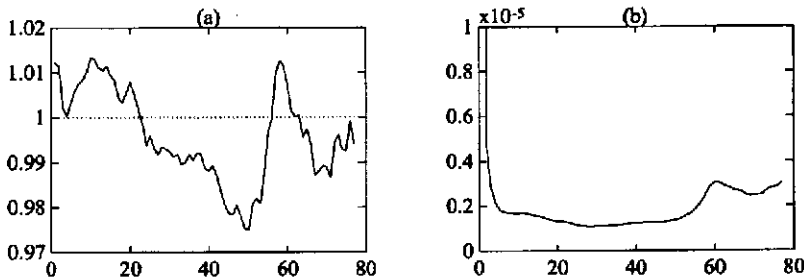


Fig. 4. Estimates $\hat{\phi}_1(t)$ and $\hat{\Gamma}_{11}(t)$ generated by the coefficients in row 4 of Table 1.

of row 3. These are good results which confirm the non-constancy of the parameters of the IBM models. Figures 2–4 plot the recursive estimates $\hat{\beta}_i(t)$ and $\hat{\Gamma}_{ii}(t)$ generated by the algorithm (3.4) with the coefficients in Table 1. Particularly interesting is the path on the unit circle of the “root” $\hat{\phi}_1(t)$ obtained from row 4.

Forecasting. In the previous experiments the orders (p, q) of the models were selected from the off-line identification and estimation (4.1). In the context of time-varying parameters, the validity of this approach is based on the assumption that the average values $\bar{\beta}_{iT} = \bar{E}_T(\beta_{it}) = T^{-1} \sum_{t=1}^T E(\beta_{it})$ were different from zero. Since this condition may not be satisfied, the sole effective way for improving adaptive modeling is to increase (p, q) tentatively. With reference to the series $\{\bar{z}_t\}$ and using F-tests, we have checked that the most significant reduction of Q_T is achieved by an AR(5) model. Estimates of the coefficients of the filter (3.4) with the preceding constraint are given in Table 2. It can be seen that the improvement of Q_T over the model (4.1a) is now -63% .

Table 2. CLS estimates of the coefficients of algorithm (3.4) applied to an AR(5) model.

Series	γ_0	λ	μ	ϕ_{10}	ϕ_{20}	ϕ_{30}	ϕ_{40}	ϕ_{50}	Q_T
\bar{z}_t	.06018 (40)	.79848 (151)	-1.1303 (14)	.79005 (62)	-.70908 (15)	.27563 (15)	.04705 (3.5)	-.06957 (4.4)	3677

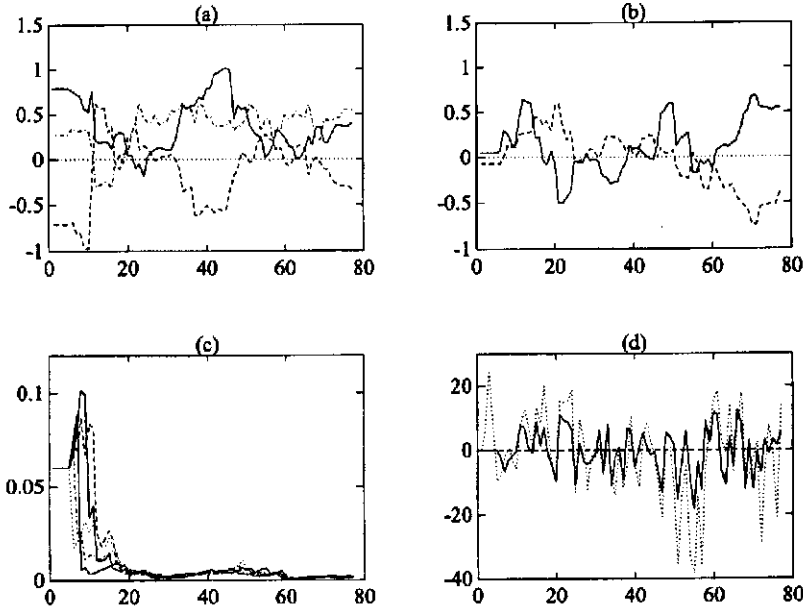


Fig. 5. Recursive estimates generated by the coefficients in Table 2: (a) $\hat{\phi}_{1t}$ —, $\hat{\phi}_{2t}$ ---, $\hat{\phi}_{3t}$ - · - ·; (b) $\hat{\phi}_{4t}$ —, $\hat{\phi}_{5t}$ ---; (c) $\hat{\Gamma}_{iit}$, $i = 1, \dots, 5$; (d) \hat{a}_t —, \bar{z}_t · · ·.

Since AR models are linear in the parameters, we may simplify the notation used for their recursive estimates as $\hat{\beta}_t$, $\hat{\Gamma}_t$. Now, Fig. 5 (a,b) show the series $\hat{\phi}_{it}$, $i = 1, \dots, 5$ generated by algorithm (3.4) with the coefficients in Table 2. Figure 5 (c,d) show the paths of the corresponding “variances” $\hat{\Gamma}_{iit}$ and innovations \hat{a}_t . We may see that the mean values of $\hat{\phi}_{4t}$, $\hat{\phi}_{5t}$ approach zero and the series $\{\hat{a}_t\}$ is nearly stationary. This is one the final goals of the adaptive modeling.

Prediction experiments were carried out on the series $\{\bar{z}_t\}$ using the adaptive AR(3) and AR(5) models implied by Tables 1 and 2. Prediction algorithms used a multilayer structure, in the sense that ARMA models were constructed for the recursive estimates; for those in Fig. 2 we had

$$\begin{aligned} \hat{\phi}_{1t} &= .031 + .918 \hat{\phi}_{1t-1} + e_{1t}, & R^2 &= 84\% \\ & \quad (1.2) \quad (19.8) \\ \hat{\phi}_{2t} &= -.028 + .863 \hat{\phi}_{2t-1} + e_{2t}, & R^2 &= 75\% \\ & \quad (1.4) \quad (14.9) \\ \hat{\phi}_{3t} &= .095 + .796 \hat{\phi}_{3t-1} + e_{3t}, & R^2 &= 64\%. \\ & \quad (2.9) \quad (11.6) \end{aligned}$$

We may see that, despite the random walk structure of the equation (3.4a), the above models do not contain unit roots. From Section 2, the approximate forecasting function is given by

$$\bar{z}_t(h) = \sum_{i=1}^3 \hat{\phi}_{it}(h-1)\bar{z}_t(h-i), \quad \hat{\phi}_{it}(h) = \alpha_i + \beta_i \hat{\phi}_{it}(h-1), \quad i = 1, 2, 3.$$

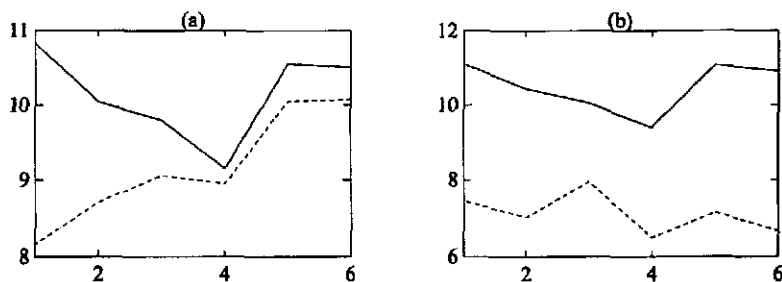


Fig. 6. Plot of statistics (4.2) for stationary (—) and adaptive (---) models of the series $\{\bar{z}_t\}$: (a) AR(3) models; (b) AR(5) models.

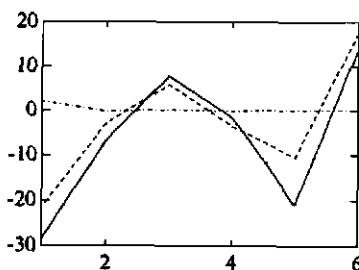


Fig. 7. Plot of data $\{\bar{z}_{72}, \dots, \bar{z}_{77}\}$ (—) and their forecasts obtained from stationary (---) and adaptive (···) AR(5) models.

To compare the forecasting ability of stationary and adaptive AR models we have used mean absolute forecast error (MAFE) statistics, defined as

$$(4.2) \quad \text{MAFE}_n(h) = \frac{1}{n} \sum_{s=1}^n |\bar{z}_{t+s+h} - \bar{z}_{t+s}(h)|, \quad h = 1, 2, \dots$$

where n is the sample size of the mean and s shifts the forecast origin. In the IBM application we have taken $t = 61$, $n = 10$ and $h = 1, 2, \dots, 6$; specifically, forecast origin was changed 10 times, and each time 6 forecasts were computed. As it can be seen in Fig. 1b, the forecast period is complete because contains positive and negative trends and a turning point. Plot of statistics (4.2) for the various models are given in Fig. 6.

In comparing Fig. 6(a) and 6(b), we may see that the AR(5) specification yields a set of MAFE statistics which are uniformly (in h) and significantly lower than those of the stationary model. The gain of forecasting capability of the adaptive AR(5) model can be quantified, on average, as 35%; although it is smaller than the in-sample performance, expressed by Q_T , this result legitimates the complexity of techniques proposed in the paper. Furthermore, from Fig. 7, which plots the last set of forecasts, we may see that adaptive predictions may locally (in t) be very accurate.

Acknowledgements

The author would like to thank the referees and the editors for the constructive review process.

Technical Appendices

Appendix 1 *Definitions of stochastic stability*

A general characterization of the stochastic stability for Markovian processes $\{z_t\}$, was provided by Meyn and Tweedie (1993). Letting $\tau_A = \inf_{t>0}\{t : z_t \in A\}$ the *hitting time* (i.e. the first time the process returns in the set A) and $P^t(x, A) = P(z_t \in A \mid z_0 = x)$ the t -th step transition probability from the initial state x , the process $\{z_t\}$ is defined:

- 1) *irreducible*, if $P(\tau_A < \infty \mid z_0 = x) > 0$ for every x, A bounded.

This means that any bounded state can be reached in a finite time from any initial condition.

- 2) *recurrent*, if $P(\tau_A < \infty \mid z_0 = x) = 1$, and further $E(\tau_A \mid z_0 = x) < \infty$.

This concept clearly strengthens that of irreducibility introduced above.

- 3) *non-evanescent*, if $P(z_t \rightarrow \infty \mid z_0 = x) = 0$, which means asymptotic boundedness.

This is the condition of stability which is more close to that introduced in the paper.

- 4) *tight*, if $\lim_{t \rightarrow \infty} \inf P^t(x, C_\varepsilon) \geq (1 - \varepsilon)$, for any $\varepsilon > 0$ and C_ε compact.

This concept concerns the transition probabilities, rather than the process itself.

- 5) *ergodic*, if $P^t(x, C) \xrightarrow{D} \pi(C)$, where $\pi(\cdot)$ is an invariant probability measure.

This property is the most strong and defines the convergence in probability of the process.

Given the non-stationarity of model (2.1), ergodicity cannot be guaranteed in our case. However, under Proposition 1 the process is non-evanescent and therefore is *bounded in probability on average*, in the sense that $\bar{P}_T(x, \cdot) = T^{-1} \sum_{t=1}^T P^t(x, \cdot)$ is tight (see Meyn and Tweedie (1993), Chapter 12). As a consequence we have that $\bar{P}_T(x, C)$ converges weakly to $\bar{\pi}(C)$ which is a sub-invariant measure for the average probability of any compact. This measure is related the average moments, defined as $\bar{\mu}_k = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(z_t^k)$, which exist for $k \leq 4$ under the conditions of Proposition 1.

Appendix 2 *The conditional least squares theorem*

The theorem provides the existence of a consistent estimator for a vector of nonlinear parameters δ in a stochastic process. Letting $\mathfrak{S}_{t-1} = \{z_{t-1}, z_{t-2}, \dots\}$ the set of past information, it requires the distribution of the prediction errors $\hat{a}_t(\delta) = [z_t - E_\delta(z_t \mid \mathfrak{S}_{t-1})] = (z_t - \mathbf{x}'_t \hat{\beta}_{t|t-1})$, where $\hat{\beta}_{t|t-1} = E(\beta_t \mid \mathfrak{S}_{t-1})$. In conditional form, this is given by $(\hat{a}_t \mid \mathfrak{S}_{t-1}) \sim \text{IID}[0, v_t^2 = (\mathbf{x}'_t \Gamma_t \mathbf{x}_t + \sigma^2)]$ where v_t^2 and $\Gamma_t = E[(\beta_t - \hat{\beta}_{t|t-1})(\beta_t - \hat{\beta}_{t|t-1})' \mid \mathfrak{S}_{t-1}]$ are conditional dispersions. Finally, let us denote partial derivatives of \hat{a}_t as follows: $\zeta_t(\delta) = -\partial \hat{a}_t(\delta) / \partial \delta$ and $W_t(\delta) = \partial^2 \hat{a}_t(\delta) / \partial \delta \partial \delta'$.

THEOREM. *Assume that $\{z_t\}$ is a second order process, and that $\{\hat{a}_t\}$ is almost surely twice continuously differentiable in an open set D_δ containing δ_0 , the unknown parameters. Moreover, assume that there are positive constants C_1, C_2, C_3 such that, for $t > p + 1$, the following conditions hold:*

- (i) $E[\zeta'_t(\delta_0) v_t^2(\delta_0) \zeta_t(\delta_0)] \leq C_1, \quad C_1 < \infty$
- (ii) $E[W'_t(\delta_0) v_t^2(\delta_0) W_t(\delta_0)] \leq C_2 \mathbf{I}, \quad C_2 < \infty$
- (iii) $\lim_{T \rightarrow \infty} \inf_T \left\{ R_T(\delta_0) = \frac{1}{T} \sum_{t=p+1}^T [\zeta_t(\delta_0) \zeta'_t(\delta_0)] \right\} \stackrel{\text{a.s.}}{\geq} C_3 \mathbf{I}, \quad C_3 > 0$
- (iv) $\lim_{T \rightarrow \infty} \sup_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left\| R_T(\delta) - R_T(\delta_0) + \frac{1}{T} \sum_{t=p+1}^T [\hat{a}_t(\delta) W_t(\delta) - \hat{a}_t(\delta_0) W_t(\delta_0)] \right\| \stackrel{\text{a.s.}}{<} \infty$ on the set $D_\varepsilon = \{\delta : \|\delta - \delta_0\| < \varepsilon\} \subset D_\delta$, where $\|\cdot\|$ is the euclidean norm.

Then, there exists a sequence of estimates $\{\hat{\delta}_T\}$ strongly consistent for δ_0 and a T_0 such that $Q_T(\delta)$ attains a relative minimum at $\hat{\delta}_T$ for $T > T_0$.

As in non-linear least squares (NLS) theory, the result is proved by using a Taylor expansion of $Q_T(\delta)$ around δ_0 (see Tjostheim (1986) or Tong (1990), Chapter 5). Looking at the assumptions, we see that (i) and (ii) define second order properties of the processes $\{\zeta_t, W_t\}$. By stochastic calculus, these are satisfied if $\{z_t\}$ is a fourth order process, that is if the assumptions of Proposition 1 hold. (iii) is a condition of parametric identifiability, that is of feasibility of the Gauss-Newton step in the algorithms (3.6). Finally, (iv) is a condition of smoothness of $\{\hat{a}_t, \zeta_t, W_t\}$ around δ_0 .

Appendix 3 Fourth order moments in CLS estimates

Focusing on an AR(1) model and the RLS algorithm with coefficient λ , the prediction error is defined as $\hat{a}_t(\lambda) = z_t - \hat{\phi}_{t-1} z_{t-1}$. Since $\hat{\phi}_t$ can be expressed as a weighted OLS estimator, we have

$$\begin{aligned}
 \zeta_t &= -\frac{\partial \hat{a}_t}{\partial \lambda} = \frac{\partial z_{t-1} \hat{\phi}_{t-1}}{\partial \lambda} \\
 &= \frac{z_{t-1} \partial (\sum_{i=2}^{t-1} \lambda^{t-i-1} z_{i-1}^2)^{-1} (\sum_{i=2}^{t-1} \lambda^{t-i-1} z_{i-1} z_i)}{\partial \lambda} \\
 \text{(A.1a)} \quad &= z_{t-1} \left(\sum_{i=2}^{t-1} \lambda^{t-i-1} z_{i-1}^2 \right)^{-2} \\
 &\quad \cdot \left\{ \left[\sum_{i=2}^{t-1} (t-i-1) \lambda^{t-i-2} z_{i-1} z_i \right] \left[\sum_{i=2}^{t-1} \lambda^{t-i-1} z_{i-1}^2 \right] \right. \\
 &\quad \left. - \left[\sum_{i=2}^{t-1} (t-i-1) \lambda^{t-i-2} z_{i-1}^2 \right] \left[\sum_{i=2}^{t-1} \lambda^{t-i-1} z_{i-1} z_i \right] \right\} \\
 \text{(A.1b)} \quad &\stackrel{\text{def}}{=} z_{t-1} R_{t-1}^{-2} \{ [(t-1)A_{t-1} - B_{t-1}]R_{t-1} - [(t-1)C_{t-1} - D_{t-1}]S_{t-1} \} \\
 &= z_{t-1} \left(\sum_{i=2}^{t-1} \lambda^{t-i-1} z_{i-1}^2 \right)^{-2} \sum_{i=2}^{t-1} \sum_{j=2}^{t-1} (t-i-1) \lambda^{2t-i-j-3} \\
 &\quad \cdot z_{i-1} z_{j-1} (z_i z_{j-1} - z_j z_{i-1})
 \end{aligned}$$

thus, we may see that 4th moments of $\{z_t\}$ are involved in the estimators (3.6). However, since $\{k \cdot \lambda^k\} \rightarrow 0$ as $k \rightarrow \infty$, it is clear that under Proposition 1 (which also allows for existence of such moments), conditions (i) and (ii) of the CLS theorem are satisfied.

Appendix 4 An extension of the RLS algorithm

Expression (A.1) for the gradient may be directly extended to AR(p) models by substituting the scalar z_{t-1} with the vector $\mathbf{x}'_i = [z_{t-1}, \dots, z_{t-p}]$. It is relatively complex, but can still be computed recursively; in particular, replacing z_{i-1}^2 with $\mathbf{x}_i \mathbf{x}'_i$ in the second term of (A.1a), we have

$$\begin{aligned}
 \text{(A.2)} \quad &\sum_{i=2}^{t-1} (t-1-i) \lambda^{t-i-2} \mathbf{x}_i \mathbf{x}'_i = (t-1) \sum_{i=2}^{t-1} \lambda^{t-i-1} \lambda^{-1} \mathbf{x}_i \mathbf{x}'_i \\
 &\quad - \sum_{i=2}^{t-1} \lambda^{t-i-1} \lambda^{-1} (i) \mathbf{x}_i \mathbf{x}'_i \stackrel{\text{def}}{=} (t-1) \mathbf{C}_{t-1} - \mathbf{D}_{t-1}
 \end{aligned}$$

$$= (t - 1)[\lambda \mathbf{C}_{t-2} + \lambda^{-1} \mathbf{x}_{t-1} \mathbf{x}'_{t-1}] - [\lambda \mathbf{D}_{t-2} + \lambda^{-1} (t - 1) \mathbf{x}_{t-1} \mathbf{x}'_{t-1}].$$

This suggests that a recursive algorithm for estimating the *hyperparameter* λ may be implemented. Indeed, using the sequential loss function $Q_t(\lambda) = \hat{a}_t^2(\lambda)$, from the first order condition $\zeta_t \hat{a}_t = 0$, expressions (A.1b) and (A.2), we may obtain the LMS algorithm

$$(A.3a) \quad \hat{\lambda}_t = \hat{\lambda}_{t-1} + \mu^* \hat{\zeta}_t (z_t - \mathbf{x}'_t \hat{\phi}_{t-1})$$

$$(A.3b) \quad \hat{\zeta}_t = \mathbf{x}'_t \hat{\mathbf{R}}_{t-1}^{-2} \{ \hat{\mathbf{R}}_{t-1} [(t - 1) \hat{\mathbf{a}}_{t-1} - \hat{\mathbf{b}}_{t-1}] - [(t - 1) \hat{\mathbf{C}}_{t-1} - \hat{\mathbf{D}}_{t-1}] \hat{\mathbf{s}}_{t-1} \}$$

$$\begin{aligned} \hat{\mathbf{R}}_{t-1} &= \hat{\lambda}_{t-1} \hat{\mathbf{R}}_{t-2} + \mathbf{x}_{t-1} \mathbf{x}'_{t-1} & \hat{\mathbf{C}}_{t-1} &= \hat{\lambda}_{t-1} \hat{\mathbf{C}}_{t-2} + \hat{\lambda}_{t-1}^{-1} \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \\ \hat{\mathbf{a}}_{t-1} &= \hat{\lambda}_{t-1} \hat{\mathbf{a}}_{t-2} + \hat{\lambda}_{t-1}^{-1} \mathbf{x}_{t-1} z_{t-1} & \hat{\mathbf{D}}_{t-1} &= \hat{\lambda}_{t-1} \hat{\mathbf{D}}_{t-2} + \hat{\lambda}_{t-1}^{-1} (t - 1) \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \\ \hat{\mathbf{b}}_{t-1} &= \hat{\lambda}_{t-1} \hat{\mathbf{b}}_{t-2} + \hat{\lambda}_{t-1}^{-1} (t - 1) \mathbf{x}_{t-1} z_{t-1} & \hat{\mathbf{s}}_{t-1} &= \hat{\lambda}_{t-1} \hat{\mathbf{s}}_{t-2} + \mathbf{x}_{t-1} \hat{z}_{t-1} \end{aligned}$$

where μ^* a coefficient that enables to track the possible variability of the factor λ . Algorithm (A.3) must run in parallel with the RLS estimator $\hat{\phi}'_t = [\hat{\phi}_{1t}, \dots, \hat{\phi}_{pt}]$; in fact, the two filters share the prediction errors \hat{a}_t . Finally, the RLS version of (A.3) can be obtained as in (3.2a) by substituting μ^* with $\hat{\gamma}_t = (\lambda^* \hat{\gamma}_{t-1}^{-1} + \hat{\zeta}_t^2)^{-1}$, where $0 < \lambda^* < 1$.

Extension of the methods (A.3a) and (A.3b) to general (ARMA) models is difficult because the parameters require nonlinear estimators which cannot be expressed in the compact form $\hat{\beta}_t = (\sum_i \lambda^{t-i} \mathbf{x}_t \mathbf{x}'_t)^{-1} \sum_i \lambda^{t-i} \mathbf{x}_t z_t$. Moreover, the vector of regressors \mathbf{x}_t involves lagged errors $\hat{a}(t - j)$, $1 \leq j \leq q$ which depend on $\beta(t - j - 1)$ and consequently on λ . All of these facts make computation of derivative (A.1) extremely complex. The same situation occurs for other kinds of hyperparameters, such as those of LMS and EKF methods. For example, in an AR model, algorithms (3.1)–(3.2c) can be expressed as $\hat{\phi}_t = \sum_{i=1}^t \prod_{j=i+1}^t (\mathbf{I} - \mu \mathbf{x}_j \mathbf{x}'_j) \mu \mathbf{x}_i z_i$ but $\partial \hat{a}_t(\mu) / \partial \mu$ has not a compact form.

REFERENCES

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

Chen, R. and Tsay, R. S. (1993). Functional coefficient autoregressive models, *J. Amer. Statist. Assoc.*, **88**, 298–311.

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

Grillenzoni, C. (1990). Modeling time-varying dynamical systems, *J. Amer. Statist. Assoc.*, **85**, 499–507.

Grillenzoni, C. (1993). ARIMA processes with ARIMA parameters, *J. Bus. Econom. Statist.*, **11**, 272–289.

Grillenzoni, C. (1994). Optimal recursive estimation of dynamic models, *J. Amer. Statist. Assoc.*, **89**, 777–787.

Guo, L. and Ljung, L. (1995a). Exponential stability of general tracking algorithms, *IEEE Trans. Automat. Control*, **AC-40**, 1376–1387.

Guo, L. and Ljung, L. (1995b). Performance analysis of general tracking algorithms, *IEEE Trans. Automat. Control*, **AC-40**, 1388–1402.

Kitagawa, G. (1987). Non-gaussian state-space modeling of Nonstationary time series, *J. Amer. Statist. Assoc.*, **82**, 1032–1041.

Kitagawa, G. and Gersh, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary time series, *IEEE Trans. Automat. Control*, **AC-30**, 48–56.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes, *Ann. Statist.*, **6**, 629–642.

- Kushner, H. (1990). *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhauser, Berlin.
- Ljung, L. and Söderström, T. (1983). *Theory and Practice of Recursive Identification*, The MIT Press, Cambridge (MA).
- Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability*, Springer, New York.
- Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficients, *J. Econometrics*, **13**, 341–363.
- Pourahmadi, M. (1986). On stationarity of the solution of a doubly stochastic model, *J. Time Ser. Anal.*, **7**, 123–131.
- Rao, M. M. (1961). Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations, *Ann. Math. Statist.*, **32**, 195–218.
- Salgado, M. E., Goodwin, C. G. and Middleton, R. H. (1988). Modified least squares algorithms incorporating exponential resetting and forgetting, *Internat. J. Control*, **49**, 477–491.
- Tjøstheim, D. (1986). Estimation of nonlinear time series models, *Stochastic Process. Appl.*, **21**, 225–273.
- Tong, H. (1990). *Nonlinear Time Series Analysis: A Dynamical System Approach*, Clarendon Press, Oxford (UK).
- Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models, *J. Econometrics*, **26**, 271–281.
- Widrow, B. and Stearns, S. D. (1985). *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs (NJ).
- Zellner, A., Hong, C. and Min, C. (1990). Forecasting turning points in international output growth rates, *J. Econometrics*, **49**, 275–304.