International Journal of Forecasting 14 (1998) 469–482

Forecasting unstable and nonstationary time series

Carlo Grillenzoni*

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

Abstract

Many time series are asymptotically unstable and intrinsically nonstationary, i.e. satisfy difference equations with roots greater than one (in modulus) and with time-varying parameters. Models developed by Box–Jenkins solve these problems by imposing on data two transformations: differencing (unit-roots) and exponential (Box–Cox). Owing to the Jensen inequality, these techniques are not optimal for forecasting and sometimes may be arbitrary. This paper develops a method for modeling time series with unstable roots and changing parameters. In particular, the effectiveness of recursive estimators in tracking time-varying unstable parameters is shown with applications to data-sets of Box–Jenkins. The method is useful for forecasting time series with trends and cycles whose pattern changes over time.

Keywords: Adaptive forecasting; Airline series; IBM stock prices; Recursive estimators; Time-varying parameters; Unstable roots

1. Introduction

Parzen (1982) introduced the idea of modeling nonstationary time series by estimating the models on the series in levels, i.e. by avoiding differencing as in Box and Jenkins (1976). This approach is suitable for modeling macroeconomic series with trends because the assumption of *unit roots* may sometimes hold only for financial processes. Moreover the probability of having processes with roots exactly on the unit circle is zero.

Although Parzen (1982) did not provide theoretical motivations, his approach can be supported by several results in mathematical statistics, in particular, those concerned with the properties of least squares (LS) estimators of the parameters of *explosive* autoregressive models (see Rubin, 1950; White, 1958, 1959; Anderson, 1959; Rao, 1961) and those of their LS predictors (see Fuller and Hasza, 1980, 1981).

The approach of directly modeling original data may be further extended to include the case of avoiding power transformations and allow ''explosive roots'' to change over time. The rationale is that roots wandering on the unit circle may generate stochastic trends with non-homogeneous components, such as structural breaks and inverting slopes. For multiple time series an original approach to forecasting unstable processes may be obtained from cointegration analysis (see Engle et al., 1989).

By following the approach of optimized recursive estimation (see Grillenzoni, 1994), this paper provides a framework for modeling *time-varying unstable roots*. This will be done in Section 3, after an integration to the work of Parzen (1982).

2. Unstable time series

By definition, *long-memory* or *unstable* time series

*Corresponding author. E-mail: carlog@iuav.unive.it

0169-2070/98/$ – see front matter
PII: S0169-2070(98)00039-9

$\{Z_t\}$ are realizations of stochastic processes which fail to satisfy standard requirements of asymptotic independence and boundedness in probability. The problem of modeling such processes may be tackled as in Parzen (1982): Given a short-memory series $\{z_t\}$ a dynamic model is a mathematical device $\phi(\cdot)$ that transforms it into white noise $\{a_t\}$; by extending this definition to the relationships between long-memory and short-memory time series one may obtain the sequence

$$\{Z_t\}\rightarrow\begin{bmatrix}\text{Stabilizing}\\ \text{Filter } \Phi(\cdot)\end{bmatrix}\rightarrow\{z_t\}\rightarrow\begin{bmatrix}\text{Whitening}\\ \text{Filter } \phi(\cdot)\end{bmatrix}\rightarrow\{a_t\}$$

Using linear representations and allowing for the existence of periodicities of size $s>1$ in the dynamics of the processes, we may define the system

$$Z_t - \Phi_1 Z_{t-s} - \cdots - \Phi_d Z_{t-ds} = z_t,$$
$$\Phi_d(B^s)Z_t = z_t \tag{1a}$$

$$z_t - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p} = a_t,$$
$$\phi_p(B)z_t = a_t \tag{1b}$$

where (B) is the lag operator: $B^s Z_t = Z_{t-s}$. Thus, the previous definition of memory-length may be checked as stability properties of linear difference equations, i.e. size of the roots P_i, ρ_j of the filters $\Phi(B^{-1})$, $\phi(B^{-1})$:

$$\Phi(B^s) = \prod_{i=1}^{d}(1 - P_i B^s),\ |P_i| \geq 1,\ i = 1,2\ldots d \tag{2a}$$

$$\phi(B) = \prod_{j=1}^{p}(1 - \rho_j B),\ |\rho_i| < 1,\ j = 1,2\ldots p \tag{2b}$$

Some ambiguities remain in the case of unit-root processes, because they have first moments but not second moments in the limit. Other special features arise in statistical inference where LS estimators do not have standard distributions (see White, 1959). In any event, the probability to meet with a process having roots exactly on the unit circle is zero. Hence, models of practical interest may simply be classified as stable or unstable.

From simulation experiments in Grillenzoni (1993), other important remarks are as follows: (*i*) trends generated by random walks plus drift and explosive models are qualitatively different: the first ones are linear, whereas the subsequent are exponential; (*ii*) models with two or more unstable roots generate series which are unsuitable for representing real data because they are smooth like deterministic functions. As a consequence, the number of unstable roots in the model (2a) may be restricted as $d=1$, independently of the width of the periodicity s.

Combining equations (2a,b) we obtain the *multiplicative* model $\phi_p(B)\Phi_d(B^s)Z_t = a_t$ with $a_t \sim \text{IID}(0, \sigma^2 < \infty)$. Apart from the usefulness in terms of interpretation, the advantage of multiplicative filters over the additive ones is essentially a question of parsimony. This is apparent in the presence of periodicities because the two parameter model $(1-\phi B)\cdot(1-\Phi B^s)Z_t = a_t$ generates an AR(3) process. Problems arise in the procedures of identification and estimation, because nonlinear algorithms are required.

Sequential procedures may be used in the forecasting phase without loss of optimality. Indeed, the LS-predictor at origin t for the lead time h must satisfy the equation

$$\phi_p(B)\Phi_d(B^s)\hat{Z}_t(h) = 0,\ \hat{Z}_t(h) = \text{E}[Z_{t+h}|Z_t, Z_{t-1}, \ldots]$$

hence, the sequential solution becomes

$$\hat{z}_t(h) = \sum_{j=1}^{h-1}\phi_j\hat{z}_t(h-j) + \sum_{j=h}^{p}\phi_j z_{t-j} \tag{3a}$$

$$\hat{Z}_t(h) = \sum_{i=1}^{h-1}\Phi_i\hat{Z}_t(h-is) + \sum_{i=h}^{d}\Phi_i Z_{t-is} + \hat{z}_t(h) \tag{3b}$$

The idea of merging forecasts obtained from short- and long-term models has also been developed in the framework of cointegration by Engle et al. (1989).

Statistical properties of the above predictor have been investigated by Fuller and Hasza (1980), (1981). In particular, if $(d=p=s)=1$ the mean squared error (MSE) is

$$\begin{aligned}\hat{\Sigma}_t(h) &= \text{E}\{[Z_t - \hat{Z}_t(h)]^2\}\\ &= \hat{\sigma}_t(h)[h^2\Phi^{2(h-1)}(\Phi^2 - 1)\\ &\quad + (\Phi^{2h} - 1)/(\Phi^2 - 1)]\end{aligned}$$

$$\hat{\sigma}_t(h) = \text{E}\{[z_t - \hat{z}_t(h)]^2\} = \sigma_a^2(1 - \phi^{2h})/(1 - \phi^2).$$

This shows that forecasting unstable processes by means of explosive roots may have significant consequences in terms of forecast variability. On the other hand, improper differencings make predictors statistically biased.

For any data-transformation $Y_t = g(Z_t, Z_{t-1} \ldots)$, the optimal (minimum MSE) predictor of the original series does not coincide with the inverse transformation of the predictor of the transformed series, i.e. $\hat{Z}_t(h) \neq g^{-1}[\hat{Y}_t(h)]$, $h > 0$. In fact, by extending the Jensen inequality to the conditional expectation (see White, 1984 p. 54) one has

$$\hat{Y}_t(h) = \mathrm{E}[g(Z_{t+h})|Z_t, Z_{t-1}, \ldots] \neq g[\mathrm{E}(Z_{t+h}|Z_t, Z_{t-1}, \ldots)] = g[\hat{Z}_t(h)]$$

where the sign $>$, $<$ of the inequality depends on the convexity or concavity of $g(\cdot)$. In particular, for the log-transformation one has $\hat{Y}_t(h) \leq \log[\hat{Z}_t(h)]$, so that the sub-optimal predictor $\tilde{Z}_t(h) = \exp[\hat{Y}_t(h)] \geq \hat{Z}_t(h)$ tends to be biased upward. With reference to the airline application of Box and Jenkins (Box and Jenkins, 1976, p. 308), we now show that the predictor (3) performs better than that which works on the log-transformed series.

2.1. Application 1

The airline series, displayed in Fig. 1, concerns monthly observations of passenger totals in U.S. international air travels from 1949 to 1960; the sample size is $N = 144$ and the periodicity is $s = 12$. The model identified by Box–Jenkins was

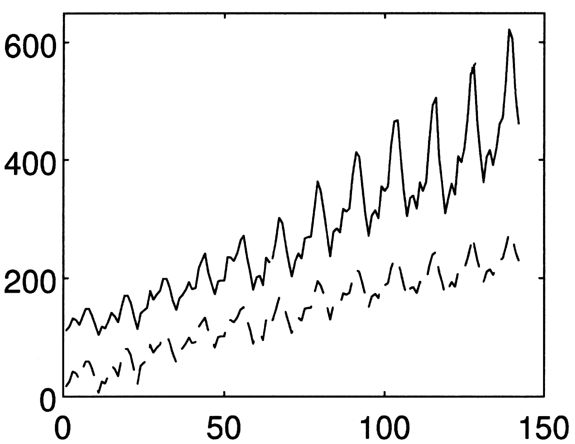

Fig. 1. Plot of airline data-set: Z_t ———, log (Z_t) - - -.

$$(1-B)(1-B^{12})\log(Z_t) = (1 - \underset{(23.4)}{0.377}B)(1 - \underset{(42.3)}{0.587}B^{12})\hat{a}_t, \qquad Q_N = \sum_{t=14}^{144} \hat{a}_t^2 = 0.182 \tag{4}$$

where the statistics in parentheses are t-statistics and Q_N is the sum of squared residuals (in-sample prediction errors). In (4) the log-transformation not only serves to stabilize the variance of $\{Z_t\}$ and to improve the efficiency of estimates, but it is essential to the existence of MA components. In fact, the model fitted on original series is

$$(1-B)(1-B^{12})Z_t = (1 - \underset{(56.3)}{0.310}B)(1 - \underset{(20.7)}{0.113}B^{12})\hat{a}_t, \qquad Q_N = 17\,752$$

Issues raised in the previous discussion concern: (i) consistency of the unit-root assumption with data; (ii) feasibility and efficacy of unstable root modelings; (iii) relationships between trend and seasonal components; (iv) effect of data transformations on forecasts. The first step toward answering these questions is to estimate the model identified by Box–Jenkins without the unit-root constraint:

$$(1 - \underset{[-7.7]}{0.948}B)(1 - \underset{[31.7]}{1.021}B^{12})\log(Z_t) = (1 - \underset{(20.3)}{0.361}B)(1 - \underset{(41.1)}{0.582}B^{12})a_t, \qquad Q_N = 0.178 \tag{5a}$$

$$(1 - \underset{[-26.8]}{0.915}B)(1 - \underset{[166.1]}{1.118}B^{12})Z_t = (1 - \underset{(59.2)}{0.399}B)(1 - \underset{(71.8)}{0.489}B^{12})a_t, \qquad Q_N = 13\,982 \tag{5b}$$

where in brackets are reported the τ-statistics $(\hat{\phi}_N - 1)/\mathrm{SE}(\hat{\phi}_N)$, whose distributions have been tabulated by Fuller (1996) for simpler cases. We may note that the unit root assumption is rejected in all cases; however, only in the model without log-transformation is the coefficient Φ far from unity and does it substantially improve the fit over the previous estimate. The introduction of a deterministic drift in the model (5b) did not improve the statistic Q_N; the reason is that unstable roots alone may generate *local trends*.

Another estimation experiment concerns the sequential treatment of multiplicative models. The results confirm the crucial importance of unstable periodic roots

$$(1-\underset{[531.2]}{1.114}B^{12})Z_t = z_t, \qquad Q_z = 35\,920 \tag{6a}$$

$$(1-\underset{[-29.8]}{0.913}B)z_t = (1-\underset{(64.2)}{0.393}B)(1-\underset{(80.9)}{0.474}B^{12})a_t,$$
$$Q_N = 14\,003 \tag{6b}$$

In this case the introduction of a drift in the first equation has significant effects on the size of the root Φ, but the global fitting performance Q_N slightly worsens.

At this point the following remarks can be made: (*i*) With or without log-transformation, the presence of roots significantly greater (Φ) and lower (ϕ) than unity is detected in the airline data-set. Without logarithms the relaxation of the unit-root hypothesis improves the fitting by about 20%. (*ii*) Periodic unstable models of the type $Z_t = \Phi Z_{t-s} + z_t$ with $|\Phi| > 1$ are capable of simultaneously capturing seasonal and trend components. This is concretely shown in Fig. 2 which compares the differenced series $(Z_t - Z_{t-12})$ with $(Z_t - 1.114\ Z_{t-12})$, where the latter is nearly stationary. The conclusion is that seasonal and trend components are neither independent nor separable. (*iii*) The high *t*-statistics in the previous models are due to the peculiar asymptotic properties of the LS estimator in the presence of unstable roots. As shown in Appendix A, standard limit theorems do not hold and *t*-ratios may be biased indicators.

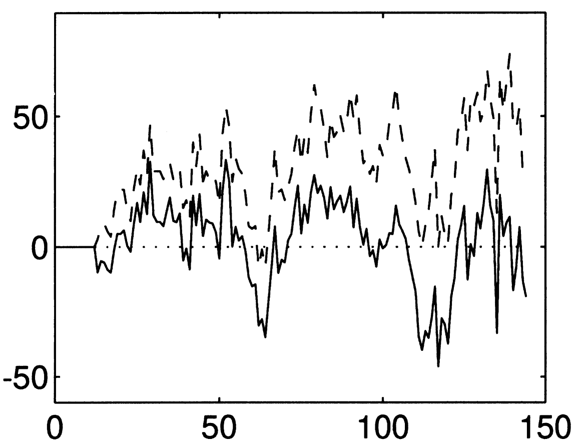

Fig. 2. Plot of filtered series: $(Z_t - 1.114\ Z_{t-12})$ ———, $(Z_t - Z_{t-12})$ - - -.

From previous estimations on log-transformed data it is not possible to discriminate precisely between models with unit or unstable roots; moreover, the values of the statistic Q_N cannot be compared with those of the models estimated on original data. Thus, the sole objective way to compare the various models is that of making forecasts of real data and computing prediction errors. Given the origin t and the steps ahead $1 < h < 12$, the predictor of multiplicative models as (5a), is given by

$$\begin{aligned}\tilde{Z}_t(h) = \exp[&\phi \log(\tilde{Z}_t(h-1)) + \Phi \log(Z_{t+h-12}) \\ &- \phi\Phi \log(Z_{t+h-13}) - \Theta a_{t+h-12} \\ &+ \theta\Theta a_{t+h-13}]\end{aligned}$$

One of the hypotheses to be checked is whether this predictor is better than that which assumes $\phi = \Phi = 1$, but worse than $\hat{Z}_t(h)$ which works on original data.

In the literature on forecasting, absolute percentage errors (APE) provide basic statistics for model comparison. In order to mitigate the dependence of individual errors from the particular forecast origin, it is necessary to average over *t*, obtaining the mean (M) statistics

$$\text{MAPE}_n(h) = \frac{1}{n}\sum_{\tau=1}^{n}\left|\frac{Z_{t+\tau+h} - \hat{Z}_{t+\tau}(h)}{Z_{t+\tau+h}}\right| \tag{7}$$

where n is the sample size of the mean and τ shifts the forecast origin. In the airline application we have taken $t = 1958.12$, $n = 12$ and $h, \tau = 1, 2, \ldots 12$; more specifically, forecast origin was changed 12 times, and each time 12 forecasts were computed.

A plot of MAPE statistics generated by models (4) and (5) is given in Fig. 3; the best one is (5b) which does not transform data, whereas the worst one is (4) proposed by Box–Jenkins. Intermediate performance is provided by model (5a), with unstable roots but log-transformed data. The performance of (5b) could be further improved by re-identifying the ARMA model of the series $\{z_t\}$ generated by (6a). However, the task of the above exercise was the evaluation of the effects of data transformations for a *given* model structure. As we may see in Fig. 3, these effects are

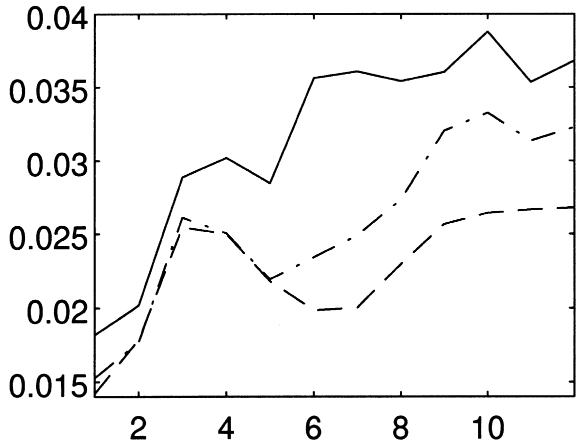

Fig. 3. Plot of statistics (7) for models (4) ———; (5a) –·–; (5b) - - -.

present in the long run because statistics (7) differ, on average, by about 50%. In any event, the performance of the various models is good because all MAPE are less than 4%.

2.2. Application 2

The second application focuses on another data-set of Box–Jenkins: the daily IBM stock prices from May 17, 1961 to Nov 2, 1962. Many financial series are difficult to forecast because they behave like random walks. For the IBM data, this hypothesis is confirmed by the estimate $Z_t = 0.9995\, Z_{t-1} + a_t$ and by the nonsignificant MA parameter of the IMA(1,1) model identified by Box and Jenkins (Box and Jenkins, 1976, p. 239). The same data-set has been republished by Tong (Tong, 1990, p. 512) together with its calendar. This enables us to obtain the weekly average Z_t (in Fig. 4) which is not a random walk:

$$Z_t = \underset{(12.8)}{1.378}\, Z_{t-1} - \underset{(3.7)}{0.380}\, Z_{t-2} + \underset{(2.1)}{0.264}\, a_{t-3} + a_t, \quad Q_N = 10\,005 \qquad (8a)$$

$$z_t = \underset{(3.5)}{0.394}\, z_{t-1} - \underset{(1.3)}{0.127}\, z_{t-2} + \underset{(2.5)}{0.292}\, z_{t-3} + a_t, \quad Q_N = 9905 \qquad (8b)$$

where $N = 77$ and $z_t = (Z_t - Z_{t-1})$ are first differences.

In practice the parameters of model (8a) are on the border of the stationarity region of an AR(2) process, and therefore imply nonstability. Unlike Application 1 the fitting performance of the unstable model is slightly worse than that of (8b). However, computation of MAPE statistics over the last 16 weeks with $t = 61$, $n = 10$, $h = 1 \ldots 6$ has shown a better performance of model (8a). The results are displayed in Fig. 5.

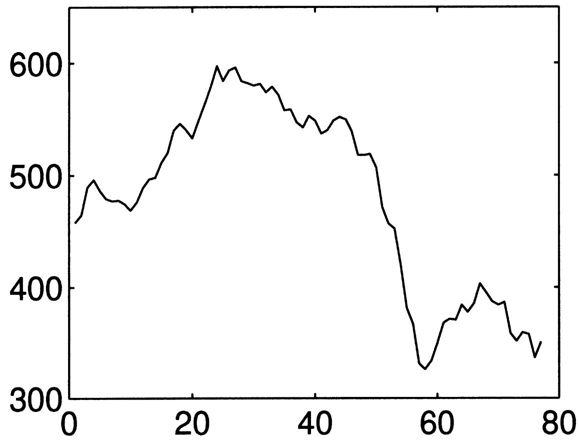

Fig. 4. Plot of the weekly average of the IBM series.

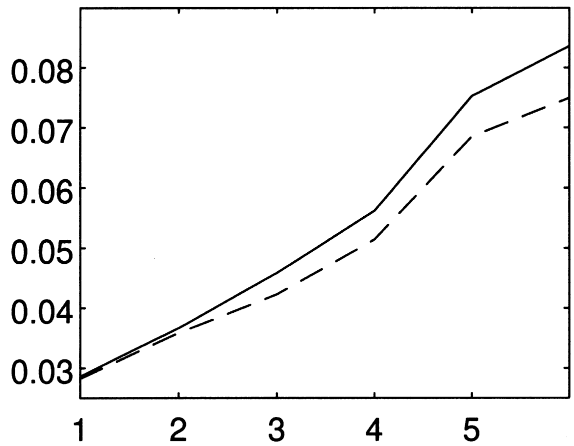

Fig. 5. Plot of MAPE statistics of models (8b) ——— and (8a) - - -.

3. Nonstationary time series

This section deals with the problem of time-varying parameters. The question is important for unstable models because we cannot expect real time series to grow indefinitely. More specifically, unstable roots wandering *on* the unit circle (i.e. inside and outside the stability region) can invert the slope of the trends and, therefore, may have a stabilizing effect on the series. This behaviour may be easily checked by simulations.

A time-varying parameter extension of the framework (1), with the constraint $d=1$ motivated in Section 1 and with $\{z_t\}$ having an ARMA representation, is given by

$$Z_t = \Phi_t Z_{t-s} + z_t, \qquad \bar{\mathrm{E}}(\Phi_t) = (\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \Phi_t) \geq 1 \tag{9a}$$

$$z_t = \sum\nolimits_{i=1}^{p} \phi_{it} z_{t-i} + \sum_{j=1}^{q} \theta_{jt} a_{t-j} + a_t,$$
$$a_t \sim \mathrm{IID}(0, \sigma_t^2) \tag{9b}$$

where $\{\Phi_t, \phi_{it}, \theta_{jt}, \sigma_t\}$ are deterministic sequences. It should be noted that deterministic is not synonymous with smooth and in (9a) the time-average of the evolving root Φ_t lies on or outside the stability region. The above may be viewed either as a two-stage system or as a model having a multiplicative structure: $\phi_t(B)\Phi_t(B)Z_t = \theta_t(B)a_t$.

Stability conditions for model (9b) require that MA parameters be uniformly bounded and the others fluctuate inside the parameter space S_p of a stationary AR(p) model. For example, if $p=2$ then S_2 is the well known triangular region and we must have $\phi_{it} \in S_p$, $i=1, 2$ for every $t \geq 0$ with the exception, at most, of finite sets of points. The presence of these sets and of Eq. (9a) may create trends and cycles in Z_t with complex transitory components, such as structural breaks and inverting slopes. It may also capture situations in which stable roots become unstable and changes in the periodicity $s>0$. Finally, the variance σ_t^2 in (9b) might be assumed as constant because the *ultimate* purpose of a time-varying modeling is to obtain stationary innovations.

If parameters are non-stochastic, the optimal forecasting function for (9) is a generalization of formula (3)

$$\hat{z}_t(h) = \sum_{i=1}^{h-1} \phi_{it+h} \hat{z}_t(h-i) + \sum_{i=h}^{p} \phi_{it+h} z_{t+h-i} + \sum_{j=h}^{q} \theta_{jt+h} \hat{a}_{t+h-j} \tag{10a}$$

$$\hat{Z}_t(h) = \Phi_{t+h} \hat{Z}_t(h-s) + \hat{z}_t(h) \tag{10b}$$

Obviously, a fundamental question is the estimation of parameters $\beta_{kt} = \phi_{it}, \theta_{jt}, \Phi_t$ where $k=1 \ldots p+q+1$. A natural tool is provided by recursive estimation methods for stochastic systems presented in Ljung and Söderström (1983). Once these methods have provided the estimates $\hat{\beta}_{kt}$, the approach of *adaptive forecasting* uses them in (10) in place of β_{kt+h}. Alternatively, one may build ARMA models for $\hat{\beta}_{kt}$ and obtain the forecasts $\hat{\beta}_{kt}(h)$.

3.1. Estimation

The derivation of recursive estimators for models which have a multiplicative structure is difficult because they involve constraints on the vector of parameters $\boldsymbol{\beta}$ which cannot be managed on-line. Moreover, the gradient $\partial a_t(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ has a nonlinear expression even in the autoregressive case. For the system (9) we must thus resort to approximate solutions which treat the model in 1-stage or 2-stage forms.

1-Stage. The first approach consists of merging Eq. (9a,b) into an ''additive'' model of order $p^* \geq (s+p)$ and then applying known algorithms. However, the attempt to recover the estimates $\hat{\Phi}_t$, which have interesting interpretation in terms of signal extraction, is subject to identification problems.

2-Stage. The second solution proceeds sequentially by applying known algorithms first to Eq. (9a) and next to Eq. (9b). In the first stage the series $\{z_t\}$ is generated as recursive innovations together with recursive parameters $\hat{\Phi}_t$. The fundamental statistical problem of these estimates is their inefficiency.

To simplify the presentation of recursive algorithms we focus on AR models; these can be written in regression form as $Z_t = \boldsymbol{\beta}_t' \boldsymbol{x}_t + a_t$ where $\boldsymbol{\beta}_t' = [\phi_{1t} \ldots \phi_{pt}]$ and $\boldsymbol{x}_t' = [Z_{t-1} \ldots Z_{t-p}]$ is the vector of regressors. Now, an algorithm that unifies the main adaptive estimators has been obtained in Grillenzoni (1994) and has the structure

$$\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t-1} + \alpha \hat{\boldsymbol{\Gamma}}_t \boldsymbol{x}_t [Z_t - \boldsymbol{x}_t' \hat{\boldsymbol{\beta}}_{t-1}], \qquad \hat{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0 \tag{11a}$$

$$\hat{\boldsymbol{\Gamma}}_t = \frac{1}{\lambda}\,\hat{\boldsymbol{\Gamma}}_{t-1} - \mu\left[\frac{\hat{\boldsymbol{\Gamma}}_{t-1}\boldsymbol{x}_t\boldsymbol{x}_t'\hat{\boldsymbol{\Gamma}}_{t-1}}{1+\boldsymbol{x}_t'\hat{\boldsymbol{\Gamma}}_{t-1}\boldsymbol{x}_t}\right] + \gamma_1\boldsymbol{I}_p,$$

$$\hat{\boldsymbol{\Gamma}}_0 = \gamma_0\boldsymbol{I}_p \qquad (11b)$$

where $\hat{\boldsymbol{\Gamma}}_t$ is the dispersion matrix of the estimator. In the above scheme $0<(\lambda,\ \mu)<1$ and $0<(\alpha,\ \gamma_1)<\infty$ are adaptation coefficients and $(\boldsymbol{\beta}_0,\ \gamma_0)$ are initial values. Notice that the series $[Z_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{t-1}]$ provides the prediction errors and when $\boldsymbol{x}_t = Z_{t-1}$, $\hat{\boldsymbol{\beta}}_t = \hat{\Phi}_t$ it may be used for generating the unobservable process $\{z_t\}$ in the system (9).

Filter (11) encompasses several algorithms. In particular, when $(\mu = 1/\lambda,\ \gamma_1 = 0)$ we have the recursive least squares (RLS) estimator with exponentially weighted observations; whereas for $(\lambda = 1,\ \mu = 1)$ we have the simplified Kalman filter (SKF) for the model $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{e}_t$ with $\boldsymbol{e}_t \sim \mathrm{N}(\boldsymbol{0},\ \gamma_1\boldsymbol{I}_p)$ and $\beta_{t_0} \sim \mathrm{N}(\boldsymbol{\beta}_0,\ \gamma_0\boldsymbol{I}_p)$. In general, we prefer to interpret (11) as a nonparametric smoother which is consistent with the assumption of parameters as unknown functions of time.

Algorithm (11) involves $p+5$ unknown coefficients, whose range of variation is somewhat wide. Until now, only heuristic rules have been provided for their design (see Salgado et al., 1988). Given a sample $\{Z_1, Z_2, \ldots, Z_N\}$, it seems appropriate to solve the problem in terms of parametric estimation, by optimizing a loss function based on prediction errors (see Grillenzoni, 1994)

$$[\hat{\alpha},\ \hat{\lambda},\ \hat{\mu},\ \hat{\gamma}_1;\ \hat{\gamma}_0,\ \hat{\boldsymbol{\beta}}_0]_N = \arg\min\left\{Q_N = \sum_{t=p+1}^{N}[Z_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{t-1}]^2\right\} \qquad (12)$$

This belongs to the conditional least squares (CLS) method discussed in Klimko and Nelson (1978); Tjostheim (1986), where ‘‘conditional’’ refers to the set of past information of prediction errors. It is worth noting that in constant parameter models the statistic Q_N has the same meaning as the residual sum of squares (RSS); therefore, we have a basis for making comparisons with the results of the previous section. Since estimation (12) might involve problems of parametric identifiability, parsimonious parametrizations, such as $\mu = \lambda$, $\gamma_0 = \gamma_1$ may be recommended.

3.2. IBM series

We start with the IBM application because models (8a,b) have not a multiplicative structure. The sole adjustment that we need when applying the method (11)–(12) to (8b) is to substitute series Z_t with $z_t = (Z_t - Z_{t-1})$. In all cases, best performance was provided by the algorithm (11) with the constraints $\mu = 1/\lambda$ and $\gamma_1 = 0$, which yield the covariance matrix of the RLS algorithm. Estimates of the adaptation coefficients with criterion (12) are given in Table 1; as we may see, the reduction of the statistic Q_N with respect to the values in (8) is very significant. Moreover, even better results were obtained by modeling the differenced series z_t as an AR(5) scheme (see last row).

To further comment on Table 1 we note that, unlike results in (8), the solution *with* the unit-root constraint is significantly better, in terms of statistic Q_N, than the other (in row 1). This is due to optimization (12) which is nonlinear and performs better on stabilized (differenced) series. The negative value of the ‘‘stepsize’’ coefficient (α) may be attributed to nonlinearity (in the variables) of the IBM series (see Tong, 1990). Fig. 6 shows the path of the estimates $\{\hat{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\Gamma}}_t, \hat{a}_t\}$ generated by filter (11) with the coefficients in row 3 of Table 1. It can be seen that the final goal of adaptive modeling, i.e. to obtain stationary innovations $\hat{a}_t = z_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{t-1}$, was nearly achieved.

As was previously discussed, adaptive forecasting may be improved by building ARMA models for the

Table 1
CLS Estimates of the coefficients of filter (11) applied to the models (8a,b)

Model	γ_0	$\lambda = 1/\mu$	α	ϕ_{10}	ϕ_{20}	ϕ_{30}/θ_{30}	ϕ_{40}	ϕ_{50}	Q_N
(8a)	1.0348	0.9561	−0.1599	1.3563	−0.3574	0.1511	–	–	8270
(8b)	0.9385	0.7823	−1.0576	0.5463	−0.4513	0.0683	–	–	5410
z_t~AR(5)	0.0602 (40.1)	0.7985 (151.3)	−1.1302 (14.5)	0.7901 (61.8)	−0.7091 (15.2)	0.2756 (14.9)	0.0471 (3.5)	−0.0696 (4.4)	3677

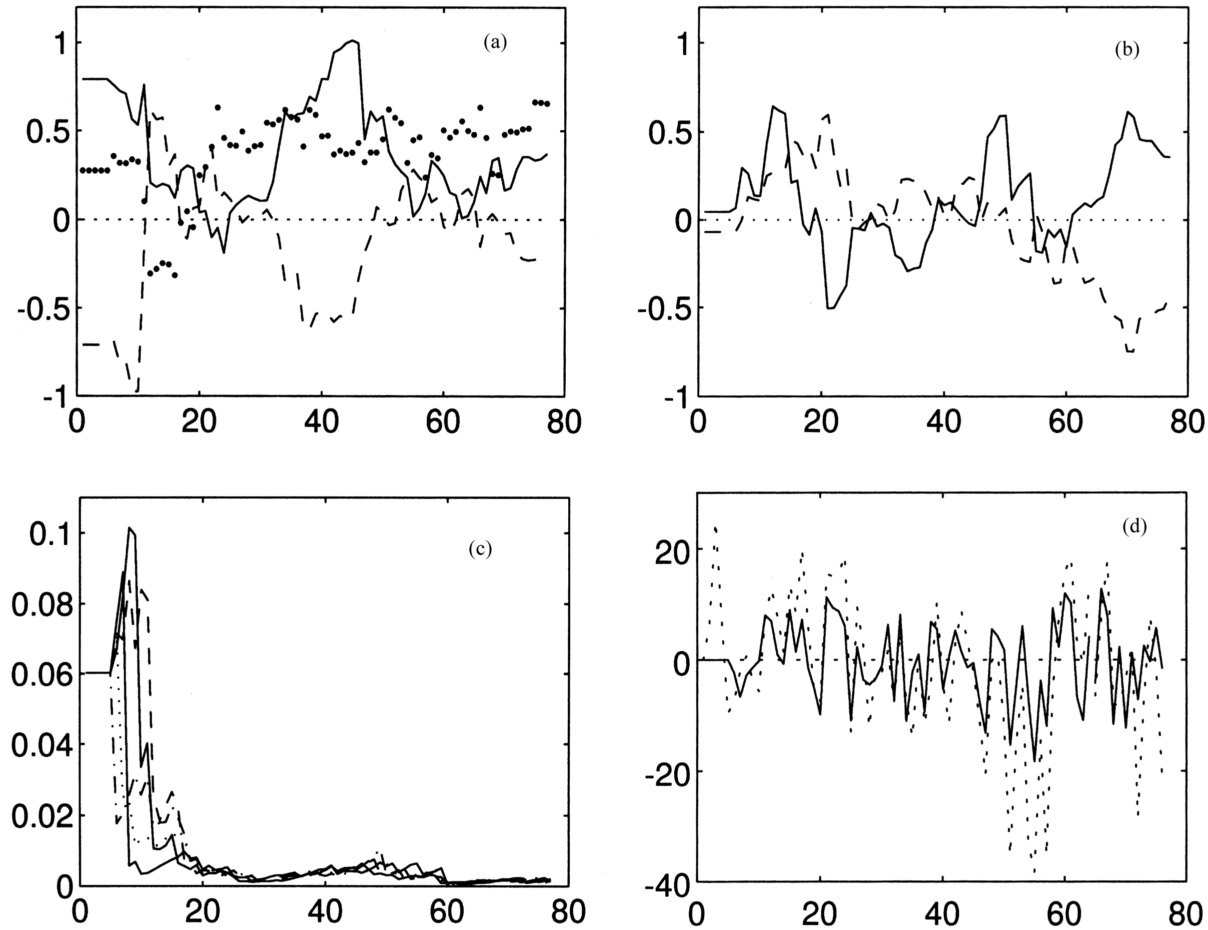

Fig. 6. Path of recursive estimates of the AR(5) model of the differenced IBM series: (a) $\hat{\phi}_{1t}$ ———, $\hat{\phi}_{2t}$ - - -, $\hat{\phi}_{3t}$ · · · ·; (b) $\hat{\phi}_{4t}$ ———, $\hat{\phi}_{5t}$ - - -; (c) $\hat{\Gamma}_{iit}$, $i=1\ldots5$; (d) $\hat{a}_t$ ———, z_t . . . . .

parameter estimates $\hat{\boldsymbol{\beta}}_t$. Analysis of series $\hat{\phi}_{it}$ in Fig. 6(a,b) has led to AR(1) schemes; thus, for the model with unit root the predictor becomes

$$\hat{\phi}_{it}(h) = \alpha_i + \beta_i \hat{\phi}_{it}(h-1) \quad (13a)$$

$$\hat{z}_t(h) = \sum_{i=1}^{5} \hat{\phi}_{it}(h-1)\hat{z}_t(h-i) \quad (13b)$$

$$\hat{Z}_t(h) = \hat{Z}_t(h-1) + \hat{z}_t(h) \quad (13c)$$

with initial value $\hat{Z}_t(0)=Z_t$. Comparisons with constant parameter models (8) were done with MAPE statistics computed as in Fig. 5; specifically, forecast origin was changed 10 times starting from $t=61$, and each time 6 forecasts were computed. Plots of these statistics are given in Fig. 7 for model (8a), and in Fig. 8 for model (8b) augmented as an AR(5). We may see that the adaptive solutions are significantly better.

To comment on Fig. 7, we note that time-varying models perform better than their constant parameter versions over the whole forecasting horizon. In comparing dashed lines, we also note that the solution *without* the unit-root constraint has lower MAPE for all $h>2$. This substantially changes the conclusion obtained from the analysis of the fitting statistics Q_N in Table 1 and confirms the results of previous section.

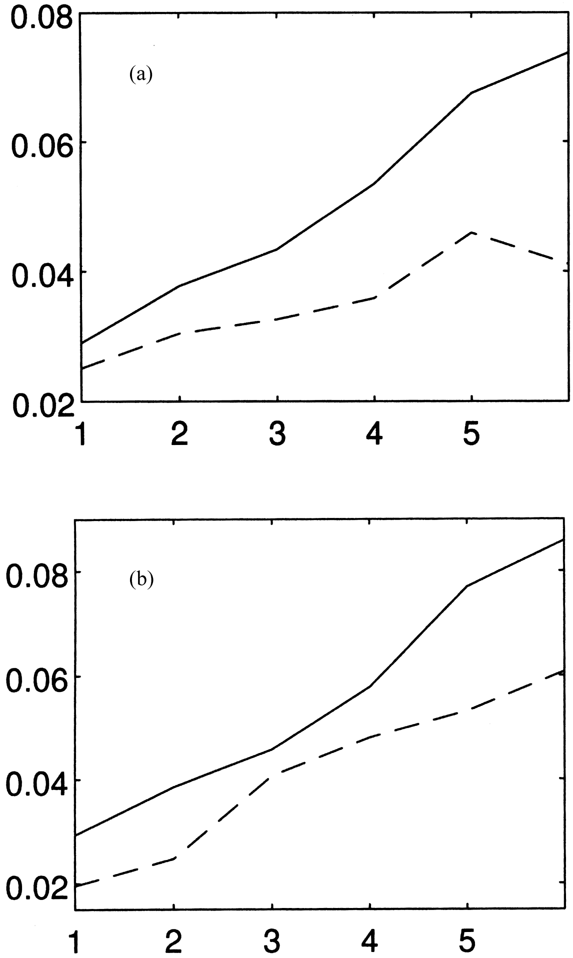

Fig. 7. MAPE statistics of constant (———) and adaptive (- - -) models of IBM series: (a) model (8a): Z_t~ARMA (2, 1); (b) model (8b) augmented: Z_t~ARI (5, 1).

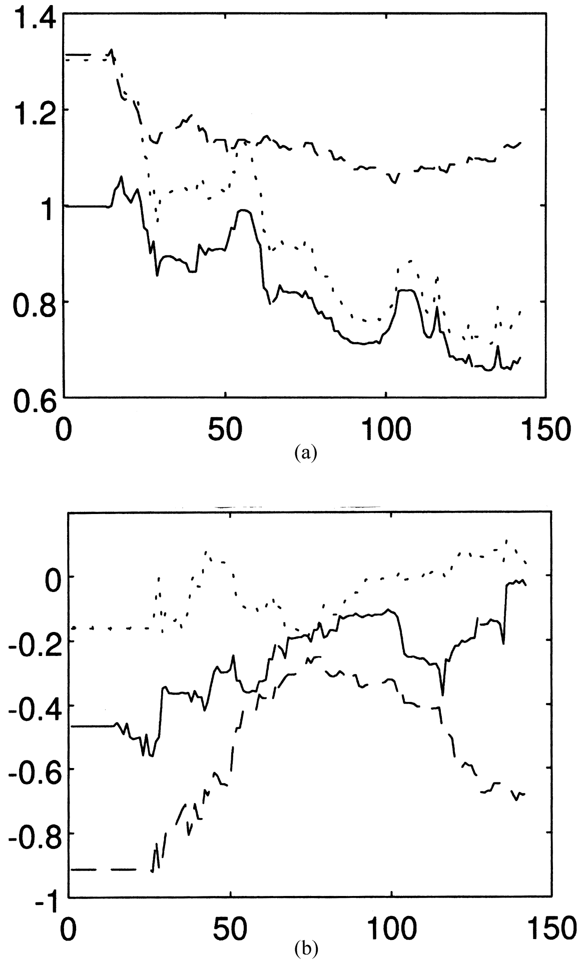

Fig. 8. Recursive estimates of the parameter of model (14) of the Airline series: (a) $\hat{\phi}_{1t}$ ———, $\hat{\phi}_{12t}$ - - -, $-\hat{\phi}_{13t}$ ····; (b) $\hat{\theta}_{1t}$ ———2, $\hat{\theta}_{12t}$ - - -, $-\hat{\theta}_{13t}$ ····.

3.3. Algorithm

It is useful to summarize the procedure we have developed so far:

Step 1. Identify the order of the constant parameter models in the usual ways.

Step 2. Estimate the parameters of these models by means of the recursive algorithm (11) providing suitable values for the coefficents α, λ, μ, γ_1 and $\boldsymbol{\beta}_0$, γ_0.

Step 3. Optimize the designs selected at Steps 1 and 2 by means of the criterion (12).

Step 4. Generate the recursive estimates $\hat{\boldsymbol{\beta}}_t$ by means of the coefficients obtained at Step 3 and algorithm (11).

Step 5. Identify constant ARMA models for these estimates and predict the original series z_t, Z_t by means of algorithms as (13).

The crucial phase of the method is represented by Step 3. Estimation (12) is highly nonlinear and its convergence may require suitable initial values and identification constraints. However, simulation results in Grillenzoni (1994) encourage its use.

3.4. Airline (1-stage)

As was previously discussed, recursive estimators for multiplicative models are difficult to obtain and approximate solutions must be developed. The 1-stage approach requires that such models be expressed in their equivalent additive form; for the Airline scheme (5b), reestimation of parameters provided

$$Z_t = \underset{(10.8)}{0.935}\, Z_{t-1} + \underset{(32.6)}{1.126}\, Z_{t-12} - \underset{(9.9)}{1.053}\, Z_{t-13} - \underset{(3.5)}{0.431}\, a_{t-1} - \underset{(6.2)}{0.626}\, a_{t-12} + \underset{(3.1)}{0.378}\, a_{t-13} + a_t \quad (14)$$

with $Q_N = 13\,662$. Now, the application of method (11)–(12) to model (14) may be simply obtained by defining $\boldsymbol{\beta}' = [\phi_1 \ldots \theta_q]$ and $\mathbf{x}_t' = [Z_{t-1} \ldots a_{t-q}]$ (see Grillenzoni, 1991).

As for the IBM case, the best performance was provided by algorithm (11) with the constraints $\mu = 1/\lambda$, $\gamma_1 = 0$; estimates of the adaptation coefficients with criterion (12) are reported in Table 2. We point out that, despite the strong unstability of the airline series, the minimization (12) converged and the value of statistic Q_N is significantly lower than that of (14). Finally, recursive estimates of the parameters of model (14) generated with algorithm (11) and the coefficients in Table 2 are reported in Fig. 8.

As in the IBM application, adaptive forecasting was performed by building AR(1) models for the series in Fig. 8. MAPE statistics were computed as in Fig. 3, namely forecast origin was shifted $n = 12$ times, starting from $t = 121$, and each time forecasts were computed for $h = 1 \ldots 12$; results are reported in Fig. 9. We may see that the adaptive modeling

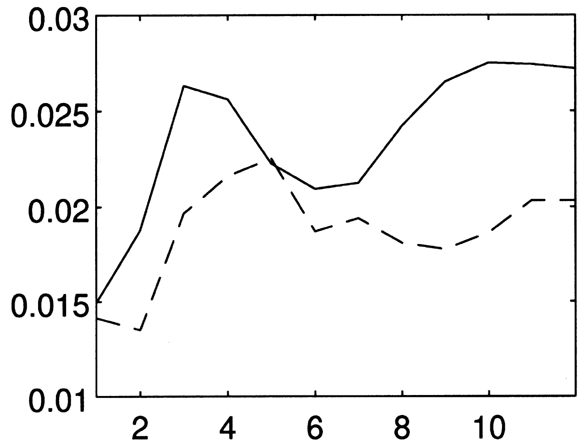

Fig. 9. MAPE statistics of model (14) (———) and its adaptive version (- - -).

performs better than its constant parameter version (14), but the improvement is not as good as in the IBM application. This is due to the nature of the Airline series which is more explosive, but less volatile, than the IBM; as a consequence, optimized adaptive techniques do not work satisfactorily.

3.5. Airline (2-stage)

In this subsection we present results of the two stage modeling of the Airline series. In the first step we applied method (11)–(12) to the model $Z_t = \Phi_t Z_{t-12} + z_t$; estimates of the coefficients are given in Table 3. It is worth noting that the value of Q_N decreases by about 50% over the corresponding statistic of model (6a).

Fig. 10 shows the path of recursive estimates generated with the algorithm (11) implemented with the coefficients in Table 3. Looking at the series $\hat{\Phi}_t$ in Fig. 10(a) one may see that its average approximates the LS estimate in (6a). Further, its trajectory is uniformly outside the unit circle and only approaches the boundary 1 at two points. It is interesting to note that in correspondence to these points, the

Table 2
CLS Estimates of the coefficients of filter (11) applied to the model (14)

γ_0	$\lambda = 1/\mu$	α	$\phi_{1,0}$	$\phi_{12,0}$	$\phi_{13,0}$	$\theta_{1,0}$	$\theta_{12,0}$	$\theta_{13,0}$	Q_N
0.0100 (1.6)	0.9575 (60.1)	− 0.4047 (5.3)	0.9974 (14.1)	1.314 (17.3)	− 1.302 (12.5)	− 0.4664 (4.1)	− 0.9137 (5.6)	0.1624 (3.2)	10 730

Table 3
CLS estimates of the coefficents of algorithm (11) applied to model (9a)

Coefficient	Φ_0	γ_0	λ	α	Q_N
Estimate	1.0973 (24.6)	0.00334 (4.2)	0.26921 (37.8)	− 0.22856 (33.0)	18 640

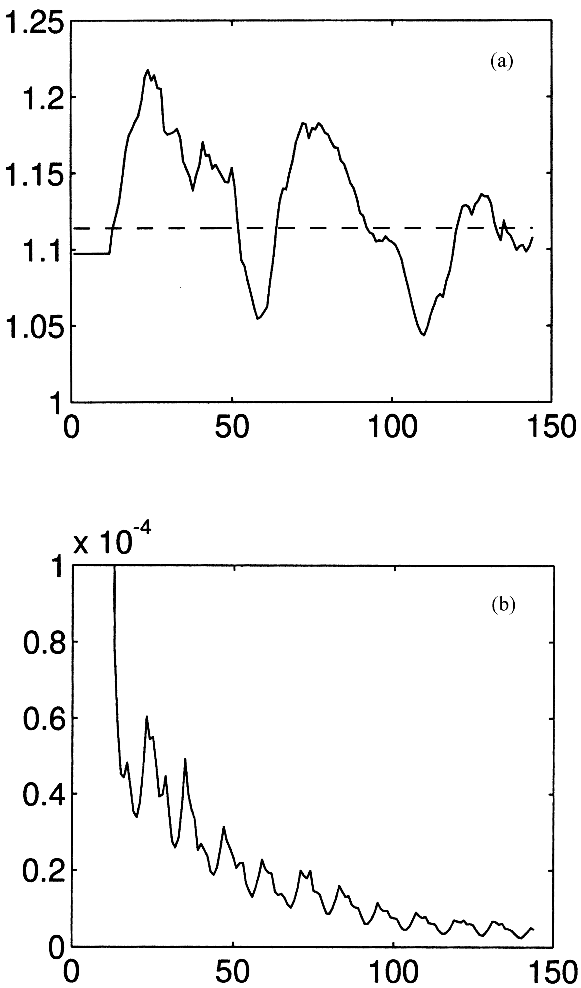

Fig. 10. Recursive estimates of model (9a) obtained with filter (11) and the coefficients in Table 3: (a) $\hat{\Phi}_t$; (b) $\hat{\Gamma}_t$.

slope of the trend of the series Z_t changes slightly (see Fig. 2). All of these features confirm the validity of our method in tracking unstable nonstationary models.

In the second stage we modeled the stabilized series in (9b), that was estimated as recursive innovations $\hat{z}_t = (Z_t - \hat{\Phi}_{t-1} Z_{t-s})$ in the first stage. It is reasonable to focus on this series, rather than on the recursive residuals $\hat{z}_t = (Z_t - \hat{\Phi}_t\ Z_{t-s})$, because the latter may be reduced to zero by simply taking $\tilde{\Phi}_t = Z_t / Z_{t-s}$. In Grillenzoni (1993) it is shown that $\{\hat{z}_t\}$ is nearly stationary and may be represented by a constant parameter model. To cope with the cross-correlation between $\hat{z}_t$, $\hat{\Phi}_t$ a *bivariate* AR model was constructed. The statistic Q_N associated with the equation of $\hat{z}_t$ turned out to be slightly lower than that of model (6b). Joint modeling of series $\hat{z}_t$, $\hat{\Phi}_t$ is also useful in forecasting because the adaptive predictor $\hat{Z}_t(h) = \hat{\Phi}_t(h)\hat{Z}_t(h-12) + \hat{z}_t(h)$ needs the joint forecasts $\hat{z}_t(h)$, $\hat{\Phi}_t(h)$.

MAPE statistics produced by this method were greater than those displayed in Fig. 9; this means that, for the Airline series, the 1-stage approach is preferable in forecasting. In general, the choice between the two methods cannot be defined a priori; it must be made on the basis of the empirical evidence. However, there are structural factors that must be considered when modeling nonstandard time series.

(*i*) The first is the inverse relationship between unstability and nonstationarity. By comparing previous applications one realizes that the greater an unstable root, the smaller its tendency to fluctuate over time. In other words, roots lying well outside the stability region tend to remain in that region just because evolutionary components (such as trends, cycles and seasonalities) are persistent.

(*ii*) The second is the inverse relationship between model complexity and algorithm complexity. In the airline application we have seen that the problem arises from the fact that optimization (12) is sensitive to the number of coefficients $\boldsymbol{\delta}' = [\boldsymbol{\alpha}, \boldsymbol{\beta}_0]$ (where $\boldsymbol{\alpha}' = [\lambda, \ldots, \gamma_0]$) to be estimated. Hence, as we fix dim $(\boldsymbol{\delta})$, if the order of the model dim$(\boldsymbol{\beta})$ increases, then recursive algorithms must be simplified. The parsimony principle discussed in Box and Jenkins (1976) assumes crucial importance here.

(*iii*) Finally, the goal of the research (namely, forecasting or signal extraction) should be precisely defined. If the latter is concerned, then one may adopt the two-stage approach which allows using complex algorithms in the first stage. It is worth noting that the procedure discussed above provides a time-varying parameter extension of the 2-stage method of Parzen (1982).

All of these factors must be evaluated, both theoretically and empirically, when applying adaptive methods. On the other hand, other solutions can be generated by combining the strategies and the algorithms described in this section.

3.6. Simulations

In this subsection we present results of simulation experiments that were performed to check the validity of the methods we have developed in this paper. The first experiment deals with constant parameter models; we simulated the process

$$Z_t = 1.03Z_{t-1} + a_t,$$
$$a_t \sim \mathrm{IN}(0, 1),\ t = 1,2 \ldots 100 \qquad (15)$$

and we compared the performance of predictors based on original and differenced series. The latter were fitted with an AR(1) scheme and out-of-sample forecasts were computed for $h=1,\ 2\ldots10$ steps ahead, starting from $t=91$. Mean values of MAPE statistics over $N=1000$ replications are shown in Fig. 11(a); as we may see, forecasts produced by differencing series are worse. This gap increases by increasing the size of the root in (15).

The second experiment focuses on varying parameter models. We considered sinusoidal parameters fluctuating on the unit circle: $\Phi_t=1-0.05\sin(t/16)$, $t=1,\ 2\ldots100$ and the same design conditions as in (15). We compared adaptive forecasts $\hat{Z}_t(h)=\hat{\Phi}_t(h)\hat{Z}_t(h-1)$, $h=1,\ 2\ldots10$, $t=91$ – where estimates $\hat{\Phi}_t$ were fitted with an AR(1) model – with those of a standard ARI(1, 1) scheme. Recusive estimates were generated with an RLS algorithm implemented with $\gamma_0=1$, $\lambda=0.85$; however, mean performance of the adaptive predictor was relatively insensitive to these designs. Mean values of MAPE statistics over $N=1000$ replications are displayed in Fig. 11(b). As before we may see that the adaptive predictor is better than that based on unit roots; moreover, this performance improves if Φ_t lye entirely outside the unit circle.

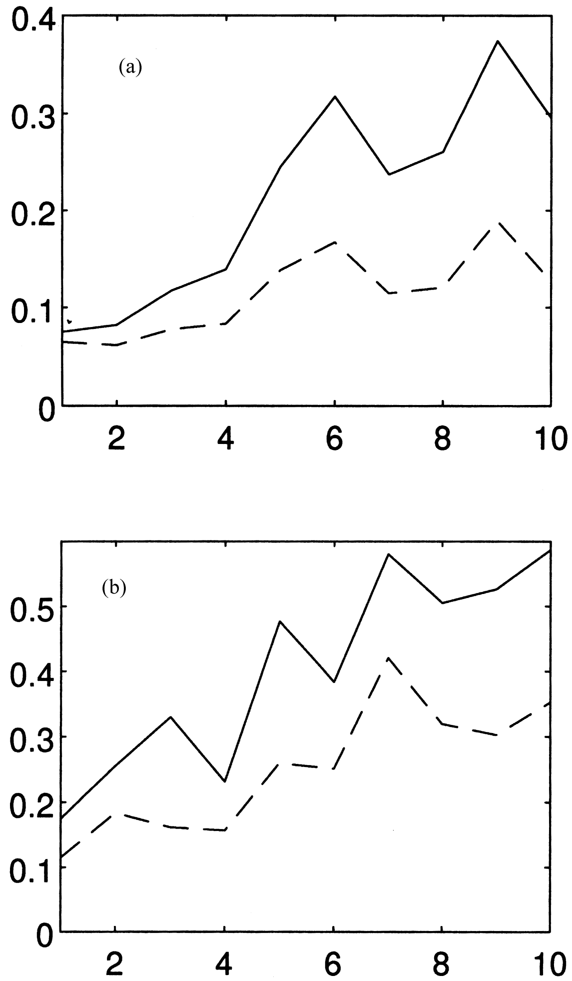

Fig. 11. Results of simulation experiments. Mean MAPE statistics of models with (———), without (- - -) unit roots: (a) constant parameters, (b) varying parameters.

4. Conclusions

In this paper we have developed a method of adaptive forecasting which is based on the optimization of recursive estimators. Through applications to well known data-sets we have demonstrated its validity in several implementations and for different model structures. An important feature is that such a method may run even on time series that contain unstable components such as trends, cycles and periodicities.

The main conclusion is that unstable models are better than those with unit-root constraints either in fitting or forecasting; moreover, this statement may be partly extended to time-varying parameter modelings. In general, these modelings improve the corresponding constant parameter solutions (with and without unit roots).

The results achieved are uniformly and significantly better than those of standard time series models. However, we do not claim that our method provides the "best" and "final" solution. Rather, it may be further extended by means of adaptation mechanisms used in signal processing, such as variable tracking coefficients (see Benveniste, 1987). Development of empirical applications remains the sole way for testing the method.

Appendix A

Review of super-consistency

Original statistical analysis of the estimators of unstable AR-processes, focused on first order models such as $Z_t = \Phi Z_{t-s} + a_t$, $a_t \sim \mathrm{IID}(0, \sigma^2 < \infty)$, $t = s, s+1\ldots$ where $1 \leq |\Phi| < \infty$ and $0 \leq |Z_0| < \infty$. If Z_0 is fixed or stochastically defined as $Z_0 = a_0$, the distribution of the process $\{Z_t\}$ is entirely determined by that of the sequence $\{a_t\}$; in the gaussian case, LS and maximum likelihood methods provide the same estimator

$$\hat{\Phi}_N = \left(\sum_{t=s}^{N} Z_t Z_{t-s} \right) \left(\sum_{t=s}^{N} Z_{t-s}^2 \right)^{-1},$$

$$\hat{\sigma}_N^2 = \frac{1}{N-s} \left(\sum_{t=s}^{N} Z_t - \hat{\Phi}_N Z_{t-s} \right)^2$$

Under stationarity it is well known that $\sqrt{N}(\hat{\phi}_N - \phi) \rightarrow \mathrm{N}[0, 1(1-\phi^2)^{-1}]$; but this does not apply to $\hat{\Phi}_N$ because its numerator and denominator increase at a rate faster than N.

The convergence of the above estimator may still be obtained by defining a standardizing function $g(N) \gg \sqrt{N}$; see Rubin (1950); White (1958), (1959); Anderson (1959); Rao (1961). These works have been surveyed in Grillenzoni (1993), and the fundamental findings are summarized in Table A. The exponential nature of the function $g(N)$ may be explained by the fact that for $Z_0 = 0$ an unstable AR(1) process may be solved as $Z_t = \Phi^{t} \Sigma_{i=0}^{t-1} \Phi^{i-t} a_{t-i}$; hence, the sum of squares of Z_t grows at the rate Φ^{2N}. In general, with Table A we may see that the convergence of the LS estimator applied to unstable processes is faster than in the case of stationary models. The cost of this *super-consistency* property is the difficulty of making inferences, for the asymptotic distributions are non-standard and their dispersions are unknown.

However, in practical terms the natural approach to follow for testing the hypothesis H_0: $\Phi = \Phi_0$ is to use the Studentized statistics $\hat{\tau}_N = (\hat{\Phi}_N - \Phi_0)[\hat{V}(\hat{\Phi}_N)]^{-1/2}$ where $\hat{V}(\hat{\Phi}_N) = \hat{\sigma}_N^2 / \left(\sum_{i=s}^{N} Z_{t-s}^2 \right)$. From Table A it is clear that for the conditions $Z_0 = a_0$ and $a_t \sim \mathrm{IN}(0, \sigma^2)$ the limiting distribution of the statistic $\hat{\tau}_N$ under the null $\Phi_0 > 1$ is N(0, 1)

An empirical way for checking the super-consistency property is to make simulations using the least squares estimator in recursive form (RLS). On-line implementation processes data one observation at time and provides corresponding parameter estimates; in this way it is possible to have a clear evidence of the speed of convergence. In Grillenzoni (1993) several simulation experiments were performed, and their results support the conclusions of Table A.

Table A. Asymptotic distribution of $g(N)(\hat{\Phi}_N - \Phi)$ when $|\Phi| > 1$.

Case	$\lvert Z_0 \rvert$	$\{a_t\}$	$g(N)$	$f(\hat{\Phi}_N)$
1)	$=0$	$\mathrm{IN}(0, \sigma^2)$	$\lvert\Phi\rvert^N(\Phi^2-1)^{-1}$	Cauchy
2)	$\neq 0$	$\mathrm{IN}(0, \sigma^2)$	$\lvert\Phi\rvert^N(\Phi^2-1)^{-1}$	Ad-Hoc
3)	$=0$	$\mathrm{IID}(0, \sigma^2$	$\lvert\Phi\rvert^N(\Phi^2-1)^{-1}$	Unknown
4)	$<\infty$	$\mathrm{IN}(0, \sigma^2)$	$[\Sigma_{t=s}^{N} Z_{t-s} / \hat{\sigma}_N^2$	N(0,1)

References

Anderson, T. W., (1959). On the asymptotic distribution of estimates of parameters of stochastic difference equation. *Annals of Mathematical Statistics*, *30*, 676–687.

Benveniste, A., (1987). Design of one-step and multistep adaptive algorithms for the tracking of time varying systems. *International Journal of Adaptive Control and Signals Processing 1*, 3–29.

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

Engle, R. F., Granger, C. W. J., & Hallman, J. J., (1989). Merging short- and long-run forecasts. *Journal of Econometrics*, *40*, 45–62.

Fuller, W. A., (1996). *Introduction to Statistical Time Series*, 2nd ed. Wiley, New York.

Fuller, W. A., & Hasza, D. P., (1980). Predictors for the first-order autoregressive process. *Journal of Econometrics*, *13*, 139–157.

Fuller, W. A., & Hasza, D. P., (1981). Properties of predictors for autoregressive time series. *Journal of the American Statistical Association*, *76*, 155–161.

Grillenzoni, C., (1991). Iterative and recursive estimation of transfer functions. *Journal of Time Series Analysis*, *12*, 105–127.

Grillenzoni, C., (1993). Forecasting unstable and nonstationary time series. *Working Papers in Economics*, n.93/1, European University Institute, Florence, Italy.

Grillenzoni, C., (1994). Optimal recursive estimation of dynamic models. *Journal of the American Statistical Association*, *89*, 777–787.

Klimko, L. A., & Nelson, P. I., (1978). On conditional least squares estimation for stochastic processes. *Annals of Statistics*, *6*, 629–642.

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

Parzen, E., (1982). AR–ARMA models for time series analysis and forecasting. *Journal of Forecasting*, *1*, 67–82.

Rao, M. M., (1961). Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Annals of Mathematical Statistics*, *32*, 195–218.

Rubin, H., (1950). Consistency of maximum likelihood estimates in the explosive case. In: Koopmans, T. C. (Ed.), *Statistical Inference in Dynamic Economic Models*. Wiley, New York, pp. 356–364.

Salgado, M. E., Goodwin, G. C., & Middleton, R. H., (1988). Modified least sqaures algorithm incorporating exponential resetting and forgetting. *International Journal of Control*, *47*, 477–491.

Tjostheim, D., (1986). Estimation of nonlinear time series models. *Stochastic Processes and Applications*, *21*, 225–273.

Tong, H., (1990). *Nonlinear Time Series Analysis: A Dynamical System Approach*. Claredon Press, Oxford.

White, J. S., (1958). The limiting distribution of of the serial correlation coefficient in the explosive case I. *Annals of Mathematical Statistics*, *29*, 1188–1197.

White, J. S., (1959). The limiting distribution of of the serial correlation coefficient in the explosive case II. *Annals of Mathematical Statistics*, *30*, 831–834.

White, H., (1984). *Asymptotic Theory for Econometricians*. Academic Press, Orlando.

Biography: Carlo GRILLENZONI is Associate Professor of Statistics at the IUAV (University Institute of Architecture of Venice) in the PTUA (Territorial, Urban and Environmental Planning) division. He obtained a BA in Economics from the University of Modena and a PhD in Statistics from the University of Padova. His research interests are in nonlinear and nonstationary time series, adaptive prediction, and nonlinear optimization. He has published papers in: Statistica, Metron, Journal of Forecasting, Journal of Time Series Analysis, Journal of Business and Economic Statistics, Journal of the American Statistical Association, Journal of Econometrics and Technometrics.