Statistica Applicata Vol. 13, n. 3, 2001

NONLINEAR PREDICTIONS OF FINANCIAL TIME SERIES

Carlo Grillenzoni

Dipartimento di Pianificazione, IUAV Venezia, Italy

Summary

This paper develops statistical techniques for building multilinear (or polynomial) ARMA models for nonlinear time series. In particular, tests for linearity use sample multi-correlation functions; stability properties are investigated by means of simulations; structure identification is based on subset regression; parameter estimation follows a pseudolinear regression approach; forecasting algorithms adopt deterministic extrapolating functions. Throughout, an extended numerical application on the IBM data-set of Box and Jenkins (1976) illustrates and checks the various solutions.

Key Words: IBM time series, Multicorrelation Functions, Nonlinear Models, Pseudolinear Regression, Recursive Algorithms.

1. INTRODUCTION

In Grillenzoni (1993) a general class of *multilinear* (M) ARMA models for nonlinear time series was introduced. It extendes the bilinear representation of Granger and Andersen (1978) by including general monomials of lagged input (the disturbances) and output (the series) in the regressors. The resulting framework unifies and encompasses other nonlinear schemes proposed in the literature, such as the polynomial AR models of Mittnik (1990) and the quadratic MA models of Hinich and Patterson (1985).

The advantage of the multilinear approach, with respect to other nonlinear models discussed in Tong (1990) and Granger and Terasvirta (1991), is in retaining a genuine regression structure, i.e. it is linear (or at least *pseudolinear*) in the parameters. This feature enables one to apply typical recursive algorithms of standard time series analysis in the procedures of identification, estimation and forecasting.

The central purpose of this article is to provide further evidence of the validity of the MARMA representation, with special reference to its forecasting ability. This will be done in Section 6 with an extended numerical application to the IBM data-

set published in Box and Jenkins (1976), and already investigated by Granger and Anderson (1978) and Tong (1990). The paper also deals with other methodological aspects, namely: the implemetation of test for linearity (Sec. 2), the analysis of the conditions of stability (Sec. 3), techniques of subset identification (Sec. 4) and adaptive estimation (Sec. 5).

2. TESTING

The preliminary step that must be taken in modeling a nonlinear time series, is testing for its linearity. This means checking the existence of nonlinear relationships in general, and next tentatively identify their typology in terms of classes of models. Over the last ten years many linearity tests have been proposed in literature; their general features can be summarized as follows: 1) Tests are usually based on statistics of residuals generated by linear AR models, which represents the null hypothesis H_0. 2) Tests can be classified into two groups, according on the fact that they do or do not assume a specific class of nonlinear models (e.g. bilinear, exponential, threshold) under the alternative.

It should be noted, however, that many of the tests that do not assume a specific model under H1 implicitly refer to quadratic processes, i.e. to systems that admit at most a second order Volterra expansion (see Keenan, 1985). Moreover, many tests cannot be applied when no linear model can be specified under H_0, i.e. when the series to be investigated is white noise (see Lukkonen *et al.*, 1988). In this case, the resort to "non-parametric" methods becomes necessary. Such methods are traditionally based on the analysis of higher order moments in the frequency domain (see Brockett *et al.*, 1988), which causes computational complexity, difficulty of interpretation and low power. There exists, however, the possibility to develop time-domain versions of such tests that may avoid these drawbacks. Some results have been already outlined by Granger and Andersen (1978, p.87), Maravall (1983) and Li (1984), using sample autocorrelations of squared whitened series. More general developments are still possible.

A general test. A fairly general starting point for testing linearity stems from the relationship between non-gaussianity and non-linearity. Linear non-gaussian processes simply arise by passing an input $a_t \sim$ IID $(0,\sigma^2)$ through a linear filter $\Psi(B) = \sum_{k=0}^{\infty} \Psi_k B^k$. Now assuming $\mu_3^a = \mathrm{E}(a_t^3) \neq 0$, the typical situation is that $z_t = \Psi(B)a_t$ have a third order cumulant (bicovariance) function which is uniformly non-zero, namely

$$\mu_3^z(i,j) = \mathrm{E}\left[(z_t - \mu_z)(z_{t-i} - \mu_z)(z_{t-j} - \mu_z)\right] = \mu_3^a \sum_{k=0}^{\infty} \Psi_k \Psi_{k+i} \Psi_{k+j} \neq 0.$$

To test for the non-gaussianity of $\{z_t\}$ it is then sufficient to check if $\mu_3^z(i,j) \neq 0$ for *any* (i,j). On the other hand, a test for linearity may be developed on the linear innovations $\tilde{a}_t = \left[z_t - \mathrm{E}(z_{t-1}, z_{t-2} \ldots)\right]$, by checking if $\tilde{\mu}_3^a(i,j) = 0$ for *all* $(i,j) \neq 0$.

More precisely, a general linearity test may refer to the *multicorrelation function*

$$\tilde{\rho}_{k+1}^a(i_1,\ldots,i_k) = \frac{\tilde{\mu}_{k+1}^a(i_1,\ldots,i_k)}{|\Delta_{k+1}|} = \frac{\mathrm{E}\left[\tilde{a}_t\left(\tilde{a}_{t-i_1} \ldots \tilde{a}_{t-i_k}\right)\right]}{|\Delta_{k+1}|}$$

for any $(i_1,\ldots,i_k) \neq (0,\ldots,0)$, where Δ_{k+1} is a normalizing quantity for the $(k+1)$-th order cumulant function. Various choices are available for Δ, such as, in increasing order

$$\Delta_{k+1} = \begin{cases} \mu_{k+1}^a(0,\ldots,0) = \mathrm{E}\left(\tilde{a}_t^{k+1}\right) \\ \left\{\mathrm{E}\left(\tilde{a}_t^2\right)\mathrm{E}\left[A_t - \mathrm{E}(A_t)\right]^2\right\}^{1/2}, \quad A_t = \prod_{j=1}^{k} \tilde{a}_{t-i_j} \\ \mathrm{E}\left(\tilde{a}_t^2\right)^{(k+1)/2} \end{cases}$$

the most suitable in terms of cross-correlation interpretation is the second one.

The meaning of the above approach is that of reducing tests for linearity to tests for the independence of linear innovations. This involves the estimation of the filter $\hat{\Psi}(B)$, the generation of residuals $\hat{a}_t = \hat{\Psi}(B)^{-1} z_t$, the computation of the statistics $\hat{A}_t = \prod_{j=1}^{k} \hat{a}_{t-i_j}$, $\bar{A}_T = T^{-1} \sum_{t=1}^{T} \hat{A}_t$ and finally one may obtain the sample multicorrelations

$$r_{k+1}(i_1,\ldots,i_k) = \frac{\sum_{t=1}^{T} \hat{a}_t \hat{A}_t}{\left[\left(\sum_t \hat{a}_t^2\right)\sum_t \left(\hat{A}_t - \bar{A}_T\right)^2\right]^{1/2}} \underset{H_0}{\overset{L}{\approx}} \mathrm{N}\left(0, \frac{1}{\left[T - \max(i_j)\right]^{1/2}}\right) \quad (1.1)$$

The approximate distribution (1.1) holds under the null $H_0: z_t = a_t \sim$IID, with $\mathrm{E}\left(a_t^{k+1}\right) < \infty$ and for T sufficiently large. The proof is similar to that for the autocorrelations and may be obtained as in Hinich and Patterson (1985) or Grillenzoni (1993), see the Appendix.

The case study. To clear the above ideas we begin here the numerical application of the paper. It deals with the IBM stock price series $\{Z_t\}$ published by Box and Jenkins (1976, p. 526) which consists of $T = 369$ daily observations during the period May 17, 1961 - Nov 2, 1962. For such series, Box and Jenkins (1976, p.239) have substantially confirmed the hypothesis of random walk, which is widely diffused in financial data. Indeed, the ML estimation of the tentatively identified IMA(1,1) model provided a non-significant MA parameter. For the differenced series $z_t = (Z_t - Z_{t-1})$, displayed in Figure 1, Tong (1990, p.260) has detected non-linearity over the first 218 observations. He applied CUSUM-type tests of Petruccelli and Davis (1986), assuming the existence of an AR(1) model under the null. Now, since the series is nearly white noise there is the need for further checks.

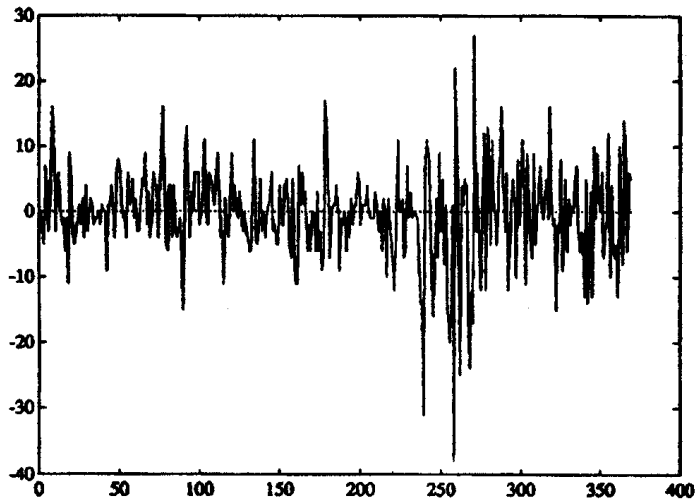

Fig. 1: Plot of first differences of IBM stock price series.

Sample autocorrelations of squared series (in Table 1) has provided strong evidence of non-linearity on the whole sample. Recall, in fact, that the approximate distribution of these coefficients under the null is $N[0,(1/368)^{1/2}] \approx .053$. Table 1 also provides the portmanteau test of McLeod and Li (1983) which adopts the Box-Pierce statistic $Q_T(K) = \sum_{k=1}^{K} (T-k) r^2(k)$; given the result (1.1) its approximate distribution under the null is $\chi^2(K)$. Finally, from the sample bi-correlations $r_3(i,j)$ in Table 2, it is possible to obtain an evidence of nonlinearity which may be useful in the identification of quadratic AR models.

Tab. 1: Sample bi-tri autocorrelations of differenced IBM series.

Lag k	1	2	3	4	5	6	7	8	9	10	11	12	
$r(z_t^2, z_{t-k})$	-.22	-.19	-.18	-.19	-.04	-.11	-.06	-.01	-.03	-.11	-.02	-.08	$Q(20) = 106$
$r(z_t, z_{t-k}^2)$	.13	.01	-.02	-.07	.01	-.07	-.07	-.02	-.01	-.11	-.01	-.03	$Q(20) = 29$
$r(z_t^2, z_{t-k}^2)$	.26	.16	.25	.26	.01	.11	.06	.05	.20	.20	.08	.15	$Q(20) = 207$

Tab. 2: Significant sample bi-autocorrelations $r[z_t\,(z_{t-i} - z_{t-j})]$.

$i\downarrow, j\rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	
1	.12												
2	-.10	.											
3	.15	.	.										
4	.12	.	.17	.									
5	.	.	.	.	.								
6	.	.	.	.	.	.							
7	.	.	.14	.	.	-.16	.						
8	-.15	.	.	.	.	.	.	.					
9	.22	.	.	.18	.	.	.12	.	.				
10	-.11	.	.10	-.11	.	-.19	.	.	.	-.11			
11	.	.	.	.	.	.	.	.	.	.	.		
12	-.16	-.15	-.14	.13	.	.	.	.	-.12	.	.12	.	

3. REPRESENTATION

A natural extension of the ARMA model $z_t = (\phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}) + a_t$, $a_t \sim \mathrm{IN}(0,\sigma^2)$ toward a representation nonlinear in the variables, can be obtained by taking a general function $z_t = f(\boldsymbol{x}_t) + a_t$ of its "regressors" $\boldsymbol{x}'_t = \left[z_{t-1}\ldots,\ldots a_{t-q}\right] = \{x_{it}\}$. Now, assuming $f(\cdot)$ analytic (i.e. differentiable of every order) around the origin $\boldsymbol{x}_t = \boldsymbol{0}$, we may expand it in *Volterra series* (see Priestley, 1988 p. 92) obtaining

$$\text{MARMA}\quad z_t = \beta_0 + \sum_{j=1}^{n}\left[\sum_{i_1=1}^{m}\ldots\sum_{i_j=1}^{m}\beta_{i_1\ldots i_j}\left(x_{i_1 t}\ldots x_{i_j t}\right)\right] + a_t,\ a_t \sim \mathrm{IID}\left(0,\sigma^2\right) \qquad (3.1)$$

$$x_{it} = \left\{z_{t-k}, a_{t-j}\right\}_{k=1,2\ldots p;\ j=1,2\ldots q}, \qquad \beta_{i_1\ldots i_j} = \left.\frac{\partial^j f(\boldsymbol{x}_t)}{\partial x_{i_1 t}\ldots\partial x_{i_j t}}\right|_{\boldsymbol{x}_t=\boldsymbol{0}}$$

where $m = p + q$. The order of the expansion n depends on the shape of $f(\cdot)$ and requirements of accuracy, the coefficients $\{\beta_{i_1 \ldots i_j}\}$ are the Volterra *kernels* and $\beta_0 = f(\mathbf{0})$.

Constraining $n = 2$ we obtain a quadratic ARMA model which encompasses many of the existing nonlinear processes having a "regression" structure

$$z_t = \beta_0 + \left(\sum_{i=1}^{p} \phi_i z_{t-i} + \sum_{j=1}^{q} \theta_j a_{t-j} + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} z_{t-i} a_{t-j} + \sum_{i=1}^{P} \sum_{j=1}^{Q} \alpha_{ij} z_{t-i} z_{t-j} + \sum_{i=1}^{R} \sum_{j=1}^{S} \delta_{ij} a_{t-i} a_{t-j} \right) + a_t \quad (3.2)$$

where $(r,P,R) \le p$, $(s,Q,S) \le q$ and $[\phi_i, \theta_j; \alpha_{ij}, \delta_{ij}]$ are subsets of the coefficients $[\beta_i; \beta_{ij}]$ in (3.1). The above includes the bilinear models of Granger and Andersen (1978), the polynomial MA and AR models of Hinich and Patterson (1985) and Mittnik (1990).

A further extension of the multilinear system (3.1) concerns the time-variability of the regression coefficients $\beta_{i_1 \ldots i_j}(t)$. This feature arises from the nonlinear representation whenever the function $f(\cdot)$ changes over time, either in terms of its structure (heterogeneity) or variability of its parameters (evolution). In both cases we would have $z_t = f_t(\boldsymbol{x}_t) + a_t$, and expanding $f_t(\cdot)$ around $\boldsymbol{x}_t = \mathbf{0}$ for each t we obtain a representation (3.1) with deterministically varying coefficients. A necessary constrain for modeling the process is that $\beta_{i_1 \ldots i_j}(t)$ be smooth functions of the time.

Stochastic Stability. Stability properties are suitable features for dynamic models, since they determine the reliability of parameter estimators and forecasting algorithms. As a 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. While this condition may allow for the existence of some moments, asymptotic stationarity is a stronger concept because anables the same moments to have a constant expression.

To study these properties for complex nonlinear models, such as (3.1)-(3.2), simulation experimemts are recommended. The simulation approach may also be useful for checking the validity of parametric conditions established on simpler models. In the following we summarize the results of Monte Carlo experiments consisting of 30 indepentent replications of size 100,000 with input $a_t \sim$ IN (0,1) and initial value $z_0 = a_0$. All computations were carryed out with the MATLAB package on a personal computer.

Simulation 1. It is widely recognized that polynomial AR models are explosive, that is their region of stability in the parameter space is empty. Granger and Andersen (1978, p. 28) have shown this situation for the process $z_t = \alpha_t z_t^2 a_{t-1} + a_t$; however, in the simulation we have cheked that the model is stable for $|\alpha| \leq .151$ and *may not* diverge for $|\alpha| < .185$. Specifically, the mean value of the stability frontier was $\bar{\alpha} = \pm .171$ with SE = .0085.

Simulation 2. It is well known that the bilinear process $z_t = \beta z_{t-1} a_{t-1} + a_t$ is second order stationary if $\beta^2 \sigma_a^2 < 1$, or $|\beta| < 1$ if $\mathrm{E}(a_t^2) = 1$. Now, in our experiment we have cheked that the model is stable for $|\beta| \leq 1.88$ and may not diverge for $|\beta| \leq 1.91$. In particular, the mean value of the stability boundary was $\bar{\beta} = \pm 1.896$ with SE =.0077.

These findings partially contradicts well established results. Their motivation arises from the fact that the input process is standard normal, therefore its realizations are actually confined within finite bounds (±5.5). In general, given any bounded input sequence with mean zero, if the "regression" coefficients are sufficiently small and do not have the same sign, then nonlinear systems of type (3.1) may be stable. Of some interest is the different way in which the models in experiments 1 and 2 diverge: while $z_t = \alpha z_{t-1}^2 + a_t$ explodes suddenly (see Figure 3a), the process $z_t = \beta z_{t-1} a_{t-1} + a_t$ diverges gradually and for $|\beta| > 1$ it may take large values. Graphs of realizations of $\log|z_t|$ in Figure 2 show, however, that the critical value $\beta = 1.90$ has the same role as the unit circle in linear AR(1) models.

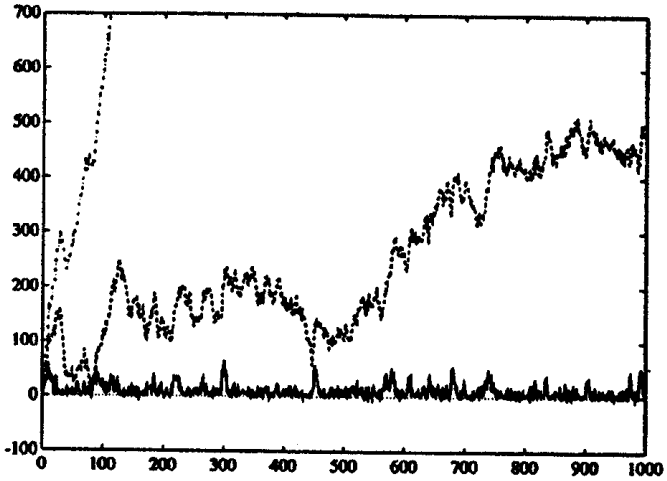

Fig. 2: Realizations of $\log|z_t = \beta z_{t-1} a_{t-1} + a_t|$ for $\beta = 1.8$ (——), 1.9 (– – –), 2.0 (– · –).

Apart from previous simulations results, there are sensible considerations that may relax the need for finding conditions of stability: *i*) Stationarity properties are certainly suitable features, but they are concerned with the asymptotic behaviour of the models. In modeling real data, users typically deal with finite sample intervals; moreover, many time series, mostly in economics, are non-stable (explosive) in nature. *ii*) As shown in the analysis of linear AR models (see Rao, 1961), explosiveness has useful consequences for standard LS-ML estimators since it increases their speed of convergence in probability. This property, named *super consistency,* might be extended to nonlinear AR models of polynomial type, since they retain a regression structure. *iii*) If the parameters vary with time (in a deterministic or stochastic fashion), issues of convergence and stationarity do not arise by definition. What becomes important is the non-divergence of the models, which may be accomplished by the variability of the "regression" coefficients itself. Specifically, as it may be shown in the case of linear models, or by simulation 2 with $\beta_t = (\beta a_{t-1})$, the stability region of time-varying models is larger than that of their stationary versions. These statements may be cheked with further experiments.

Simulation 3. Under the same experimetal conditions as before, we have cheked that the process $z_t = \alpha_t z_t^2 a_{t-1} + a_t$ with parameter $\alpha_t = -\alpha \text{sign}(a_t)$, where the function $\text{sign}(x) = +,-1$ for $x >, <0$, is stable for $\alpha \leq .20$ and may not diverge for $\alpha \leq .29$. In practice, the mean value of the stability frontier becomes $\bar{\alpha} = .254$ with SE = .021.

Simulation 4. To show the consistency of the LS estimator in the case of explosive non-linear models we have applied its recursive version (RLS) to 4 independent realizations of $z_t = .3z_{t-1}^2 + a_t \sim \text{IN}(0,1)$. The graphs of a realization and of the recursive estimates are given in Figure 3 (a,b) respectively. It is interesting noting that the convergence of $\{\hat{\alpha}_t\}$ toward 0.3 is not disturbed by the sudden divergence of $\{z_t\}$.

Also conditions of invertibility may be investigated by simulations. This property enables to obtain the input series $\{a_t\}$ from any bounded sequence of $\{z_t\}$, and is assential to the estimation of MA parameters. Anyway, even if conditions of invertibility are not met, consistent estimators may still be implemented (see Section 5).

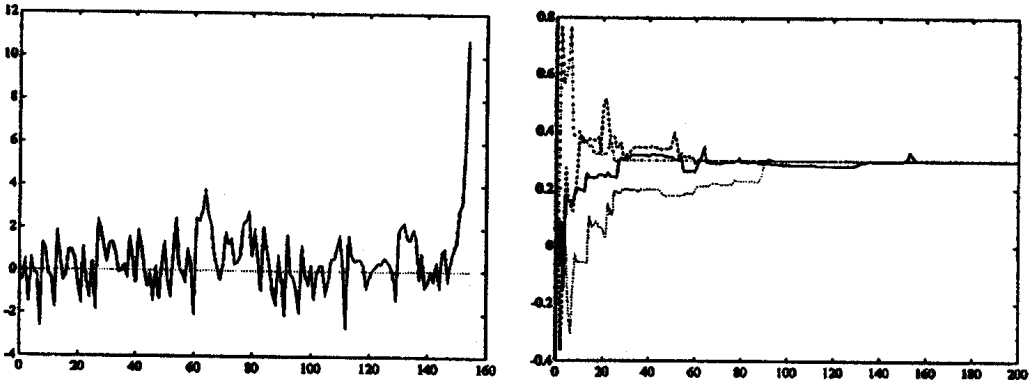

Fig. 3: (a) A realization of $z_t = 0.3z_{t-1}^2 + a_t$; (b) Four RLS estimates of $\alpha = 0.3$.

4. IDENTIFICATION

A crucial phase in modeling a nonlinear time series is the specification of its structure. With respect to the class (3.1), this requires the identification of the monomials $\left\{ y_{jt} = \prod_{i=1}^{p} z_{t-i}^{k_{ij}} \prod_{i=1}^{q} a_{t-i}^{h_{ij}} \right\}_{j=1...n}$, i.e. of the powers $\{k_{ij}, h_{ij}\}$. For the sub-class (3.2), which is more regular, two techniques developed for linear and bilinear models might be used. In the first, by assuming $a_t \sim$ IN (independent normal), the orders are selected by minimizing some information criterion (IC). The second one simply requires $a_t \sim$ IID and selects models by investigating the sample behaviour of same cumulant function.

Both these approaches are of limited practical value since they rely heavily on the assumption that a regular multilinear system exists. By contrast, data are often generated by *subset models*, which have coefficients spread at various lags. The main consequences are that the estimation of information criteria may fail owing to the presence of many non-significant and collinear terms, which make the Hessian matrix associated with the non-linear estimator, ill-conditioned. Secondly, analysis of the theoretical multicovariance functions, related to all the subset alternatives of (3.1), is practically impossible and some patterns are shared by different model structures.

The identification procedure that we now propose is inspired to the method of *subset regression* and treats the multilinear model as the sum of independent components to be identified separately and next recomposed. This approach has been applied by Grillenzoni (1991) in the identification of other complex systems, such as simultaneous transfer functions. With reference to (3.2) the method is as follows:

Step 1. The model is decomposed into five *submodels:* 1) linear $[\phi_i, \theta_j]$; 2) bilinear subdiagonal $[\beta_{ij}, i \le j]$; 3) bilinear superdiagonal $[\beta_{ij}, i > j]$; 4) quadratic AR $[\alpha_{ij}]$; 5) quadratic MA $[\delta_{ij}]$. Since the nonlinear parts have the same parametric complexity, a common maximum order $D = (r = s = P = Q = R = S)$ is defined depending on the available data and requirements of parsimony.

Step 2. The linear submodel is identified in the usual way. For the others, all possible elementary models of the type $z_t = \alpha_{ij} z_{t-i} z_{t-j} + a_t$, $z_t = \beta_{ij} z_{t-i} a_{t-j} + a_t$, $z_t = \delta_{ij} a_{t-i} a_{t-j} + a_t$ with $i, j = 1,2 \ldots D$, are estimated with nonlinear algorithms.

Step 3. The nonlinear submodels are identified by *assembling* the most significant of the elementary models and then by estimating the resulting structure and dropping all non-significant terms. Finally, the global system (3.2) is identified by recomposing the 5 submodels.

The above procedure induces a moderate overparametrization, but it drastically reduces the number of terms to be considered and is suitable for subset models. In any event, in the final estimation all non-significant monomial are identified and deleted.

There are some questions related to the properties of estimates and the reduction of intermediate computations, that must be discussed. *i*) Subba Rao and Gabr (1984, p.287) have shown, by simulations, that nonlinear least squares (NLS) estimates of the parameters of bilinear models are not normally distributed. This rises a difficulty in evaluationg the significant monomials to be included (or dropped) in the various phases. At Step 2, however, the significance may be easily evaluated with an F-test on the reduction of the residual sum of squares (RSS). *ii*) In order to reduce computations, the estimates at Step 2 of the models $z_t = \alpha_{ij} z_{t-i} a_{t-j} + a_t$ may be replaced by the sample multicorrelations $r_3(i,j)$ (see Table 2). This approach might be extended to the bilinear components by focusing on the "cross-correlations" between z_t and $\tilde{y}_{ijt} = z_{t-i} \tilde{a}_{t-j}$, where $\tilde{a}_t$ is the residual of the model formed by the linear plus quadratric AR components.

The Application. In the differenced IBM series the linear component is absent, because it is nearly a white noise. Letting the maximum order of the nonlinear submodels be $D = 10$, by means of Table 2 we have identified the quadratic AR system

$$z_t = \underset{(1.5)}{-.584} + \underset{(2.0)}{.006} z_{t-1}^2 - \underset{(2.1)}{.010} z_{t-1} z_{t-2} - \underset{(2.7)}{.017} z_{t-1} z_{t-8} + \underset{(3.1)}{.020} z_{t-1} z_{t-9} + \underset{(1.9)}{.009} z_{t-3} z_{t-4}$$
$$+ \underset{(2.4)}{.017} z_{t-4} z_{t-9} - \underset{(3.4)}{.018} z_{t-4} z_{t-10} - \underset{(4.1)}{.019} z_{t-6} z_{t-10} + \hat{a}_t, \qquad \sum_{t=11}^{369} \hat{a}_t^2 = 15,428 \tag{4.1}$$

where values in parentheses are t-statistics and the total sum of squares to be reduced was $\sum_{t=11}^{369} z_t^2 = 18,740$. In a simulation exercise consisting of 30 independent replications of size 10,000 with input $a_t \sim \text{IN}\left(0, \sqrt{15428/349}\right)$, we have checked that the above model is stable, with realizations lying in the band ±45. This property may be explained by the fact that the regression coefficients are small enough and have not the same sign. A useful consequence is that the model can be used in forecasting.

5. ESTIMATION

The identification procedure outlined in the previous section has implicitly assumed that an efficient estimator for multilinear models is available. If the distribution of $\{a_t\}$ is known a-priori a natural candidate is the maximum likelihood method; however, this assumption contradicts the formulation (3.1) where $a_t \sim$ IID. In this section we adopt the nonlinear least squares (NLS) and the pseudo-linear regression (PLR) approaches.

First, note that any subset MARMA model can be rewritten in "regression" form as $z_t = \beta_0 + \sum_{j=1}^{n} \beta_j y_{jt} + a_t$ with $y_{it} = \prod_{i=1}^{m} (x_{it})^{w_{ij}}$; now, setting $\boldsymbol{\beta}' = [\beta_0, \beta_1 \ldots \beta_n]$ the NLS estimator is defined as

$$\hat{\boldsymbol{\beta}}_T = \underset{\boldsymbol{\beta}}{\text{argmin}} \left[Q_T(\boldsymbol{\beta}) = \sum_{t=1}^{T} a_t^2(\boldsymbol{\beta}) \right], \qquad a_t(\boldsymbol{\beta}) = \left[z_t - \boldsymbol{\beta}' \mathbf{y}_t(\boldsymbol{\beta}) \right] \tag{5.1}$$

where $\mathbf{y}_t(\boldsymbol{\beta})' = [1, y_{1t} \ldots y_{nt}]$ is the vector of "regressors". The computation of $\hat{\boldsymbol{\beta}}_T$ (5.1) in terms of Gauss-Newton algorithm, requires the analytical expression of the gradient $\boldsymbol{\xi}(\boldsymbol{\beta}) = \partial a_t / \partial \boldsymbol{\beta}$. Unlike linear and bilinear ARMA models, however, this cannot easily be derived and optimization (5.1) must proceed by strictly numerical methods .

Statistical properties of the NLS estimator applied to MARMA models are difficult to analyse, even more than the probabilistic properties of the underlying processes. Since higher order moments are involved, existence of stochastic stability may not be sufficient to yield consistency; in any case extensive Monte Carlo simulations are recommended. The problem of invertibility is more urgent than that of stationarity, because it enables the iterative estimates to be computed. As mentioned in the Section 2, however, there exists the possibility of implementing alternative algorithms. As in the linear case (see Hannan and McDougall, 1988), the

basic idea is to make the input sequence $\{a_t\}$ observable, by estimating a "long" polynomial AR model, next the monomials $\{y_{ij}\}$ are generated and the vector $\boldsymbol{\beta}$ is estimated by OLS. Since the initial estimation tends to converge even under non-stability (see Figure 2), the resulting procedure may be consistent.

Pseudolinear Algorithms. The previous two-stage procedure introduces to pseudo-linear regression. Following Grillenzoni (1993), this method arises from approximating the gradient as $\xi(\boldsymbol{\beta}) \approx y_t(\boldsymbol{\beta})$ and inserting its iterative expression $\hat{y}_t(k)$ in the Gauss-Newton estimator; the final result is an iterative OLS algorithm applyed to the model $z_t = \boldsymbol{\beta}' y_t + a_t$

$$\text{PLR} \quad \hat{\boldsymbol{\beta}}_T(k+1) = \left[\sum_{t=1}^{T} \hat{y}_t(k)\hat{y}_t(k)\right]^{-1} \sum_{t=1}^{T} \hat{y}_t(k) z_t \tag{5.2}$$

In the estimation of nonlinear models, this approach significantly reduces the order of the moments that need to exist, on the other hand it does not provide a minimization method. In practice, as shown in the linear context by Hannan and McDougall (1988) or Grillenzoni (1991), the approximation of the gradient makes the resulting algorithms not always consistent and generally inefficient.

Utilization of the PLR method should then be limited to the initial estimation phase. Here, it represents a very flexible tool that may handle models of large dimensions and quickly shows which of the monomials selected at Step 2 are significant. There are other practical advatages that it is worth to discuss. *i*) The estimator (5.2) may avoid the need for initial parameter values $\hat{\boldsymbol{\beta}}_T(0)$; what it really needs is the availability of a vector $\hat{y}_t(0)$ which may be generated by a "long" polynomial autoregression. *ii*) There are several *stepsize mechanisms* that may improve the convergence of (5.2) with respect to $k \to \infty$. These include a moving average of the parameter estimates $\hat{\beta}^*_T(k) = \frac{1}{2}\left[\hat{\beta}_T(k) + \hat{\beta}^*_T(k-1)\right]$ or the use of the "residuals of regression" $\hat{a}^*_t(k) = z_t - \hat{\boldsymbol{\beta}}_T(k)' \hat{y}^*_t(k-1)$ in the generation of monomials. The second mechanism is computationally simpler.

The PLR approach is also useful in the adaptive estimation of the parameters of time-varying models. In this case, questions of convergence with respect to $T \to \infty$, do not matter and the tracking properties of PLR, allowed by the greater computational speed, are preferable to those of accuracy of NLS. An *adaptive* version of (5.2) may be obtained by discounting observations with exponential

weights $0<\lambda^{T-t}<1$; the *recursive* version follows by equating number of observations and iterations $(k = T) = t$ and proceeding as in the derivation of the RLS algorithm (see Grillenzoni, 1991).

$$\tilde{a}(t) = z_t - \hat{\boldsymbol{\beta}}(t-1)'\hat{y}(t) \tag{5.3a}$$

$$\boldsymbol{R}(t) = \lambda \cdot \boldsymbol{R}(t-1) + \hat{y}(t)\hat{y}(t)' \tag{5.3b}$$

R-PLR

$$\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\beta}}(t-1) + \boldsymbol{R}(t)^{-1}\hat{y}(t)\tilde{a}(t) \tag{5.3c}$$

$$\tilde{a}(t) = z_t - \hat{\boldsymbol{\beta}}(t)'\hat{y}(t) \tag{5.3d}$$

$$\hat{\sigma}(t)^2 = \lambda \cdot \hat{\sigma}(t-1)^2 + (1-\lambda)[\hat{a}(t)\tilde{a}(t)] \tag{5.3e}$$

$$\hat{y}(t+1) = \left\{\prod_{i=1}^{p}(z_{t+1-i})^{k_{ij}} \prod_{i=1}^{q}\hat{a}(t+1-i)^{h_{ij}}\right\} \tag{5.3f}$$

The terms $\tilde{a}_t, \hat{a}_t$ are the prediction error and the recursive residual respectively; the factor $0 < \lambda < 1$ by preventing $\boldsymbol{R}(t)$ from vanishing, enables parameter changes $(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})$ to be tracked. Finally, $\hat{\sigma}(t)^2$ provides an adaptive estimator for the variance σ_t^2 of the input sequence, because $\hat{u}(t) = [\hat{a}(t)\tilde{a}(t)]^{1/2}$ are standardized innovations.

In order to avoid numerical problems in the algorithm (5.3), the direct calculation of the inverse $\boldsymbol{P}(t) = \boldsymbol{R}(t)^{-1}$ may be replaced by

$$\boldsymbol{P}(t) = \frac{1}{\lambda}\left[\boldsymbol{P}(t-1) - \frac{\boldsymbol{P}(t-1)\hat{y}(t)\hat{y}(t)'\boldsymbol{P}(t-1)}{\lambda + \hat{y}(t)'\boldsymbol{P}(t-1)\hat{y}(t)}\right] \overset{\text{def}}{=} \frac{1}{\lambda}[\boldsymbol{P}(t-1) - \boldsymbol{\Delta}(t)] \tag{5.4}$$

This implementation avoids blowing up of $\boldsymbol{P}(t)$ and for $\lambda = 1$ it explains why in Section 2 the RLS estimates of the unstable model $z_t = \alpha z_{t-1}^2 + a_t, |\alpha| \geq 0.2$ converged despite the fact that the corresponding OLS estimates were not asymptotically definite.

The implementation (5.4) also enables to extend the algorithm (5.3) toward a more flexible and adaptive design. Following Grillenzoni (1994), the estimator which unifies weighted RLS, Kalman filter and stochastic approximation schemes is given by

$$\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\beta}}(t-1) + \mu \boldsymbol{P}(t)\hat{y}(t)\tilde{a}(t), \qquad \hat{\boldsymbol{\beta}}(0) = \boldsymbol{\beta}_0 \tag{5.5a}$$

$$\boldsymbol{P}(t) = \frac{1}{\gamma}\boldsymbol{P}(t-1) - \alpha\boldsymbol{\Delta}(t) + \gamma_1 \boldsymbol{I}_n, \qquad \boldsymbol{P}(0) = \gamma_0 \boldsymbol{I}_n \tag{5.5b}$$

where $0<\alpha,\lambda<1$ and $0<\gamma_1,\mu<\infty$ are tracking coefficients, γ_0,β_0 are parametric initial values and $\Delta(t)$ is defined as in (5.4) with $\lambda=1$. As for the algorithm (5.3), it is important noting that no assumption of parameter evolution is made, therefore (5.5) is consistent with approach of non-parametric regression.

Algorithm (5.5) involves (5+n) unknown coefficients $\boldsymbol{\delta}=[\mu,\lambda,\alpha,\gamma_1;\gamma_0,\boldsymbol{\beta}_0']$ whose heuristic design may decrease the statistical performance of the adaptive MARMA model with respect to its version with constant parameters. An estimation criterion for $\boldsymbol{\delta}$ analogous to (5.1), is based on the prediction errors

$$\text{CLS}\quad \hat{\boldsymbol{\delta}}_T=\underset{\boldsymbol{\delta}}{\operatorname{argmin}}\left[\tilde{Q}_T(\boldsymbol{\delta})=\sum_{t=1}^{T}\tilde{a}(t)^2\right],\quad \tilde{a}(t)=[z_t-\hat{y}(t)'\boldsymbol{\beta}(t-1)] \tag{5.6}$$

where calculation is provided by (5.5) itself. This framework sets up a highly nonlinear estimation problem, whose properties may be investigated in the context of the conditional least squares (CLS) theory (see Hall and Heyde, 1980, p.172). From this theory, consistency of $\hat{\boldsymbol{\delta}}_T$ may be established if the filter (5.5) is exponentially stable, that is if $0<\alpha,\lambda<1$ and $0<\gamma,\mu<\infty$. Finally, to avoid problems of parametric identifiability, it may be usefull to reduce the dimension of $\boldsymbol{\delta}$ by introducing the constraints $\alpha=\lambda$, $\gamma_1=\gamma_0/10^c$.

The application. CLS estimates of the coefficients of algorithm (5.5) with the constraint $\gamma_1=\gamma_0/100$ and applied to the model (4.1) without the constant β_0, are given in Table 3. Calculations were carried out with the Gauss package.

Tab. 3: CLS estimates of the coefficients of the algorithm (5.5) applied to (4.1).

γ_0	λ	α	μ	β_{11}	β_{12}	β_{18}	β_{19}	β_{34}	β_{49}	β_{410}	β_{610}	$\tilde{Q}_T$
.0114	.9596	.8139	−.325	.0471	−.0141	.0466	.0096	.0103	−.001	.0472	−.026	12,653
(2.9)	(71.4)	(7.5)	(12.3)	(8.2)	(1.5)	(6.9)	(1.2)	(1.4)	(0.9)	(7.5)	(3.1)	

Since the Q-statistic in (5.6) has the same nature as that in (5.1), under the assumption $a_t \sim$ IN we may use F-tests for checking the time-variability of MARMA parameters. In our application, the reduction of Q_T, from equation (4.1) to Table 3 (about -18%), yields a statistic $\hat{F}_{3,346}=16.3$ which is 1% significant. Finally, Figure 4 shows the trajectories of the recursive estimates generated with the algorithm (5.5) and the coefficients of Table 3. Specifically, (a,b) show $\hat{\boldsymbol{\beta}}(t)$, while (c,d) display the diagonal elements of $\boldsymbol{P}(t)$.

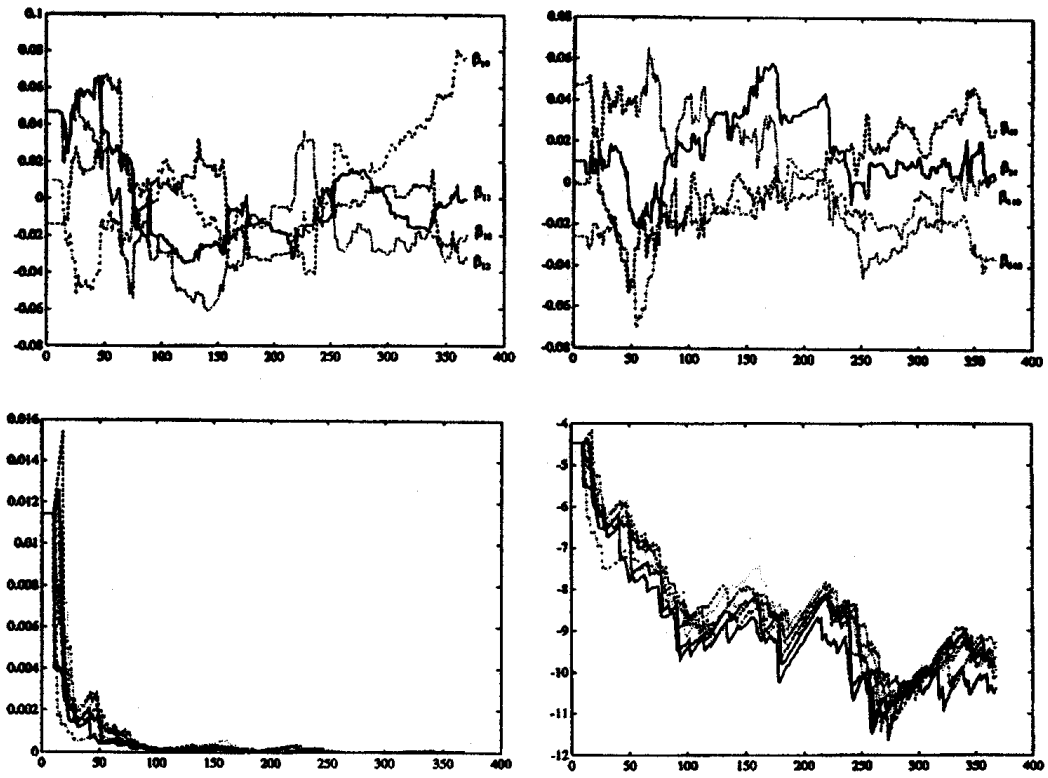

Fig. 4: Recursive estimates of model (4.1). (a,b) $\hat{\beta}(t)$, (c) diag$P(t)$, (d) log[*diagP*(t)].

6. FORECASTING

It is well known that the optimal predictor (in MSE sense) of a future value $z_{t+l}, l>0$ of a process is given by $\hat{z}_t(l) = \mathrm{E}\left[z_{t+l}|z_t, z_{t-1}\ldots\right]$; this conditional expectation, however, is linear in z_{t-k} only under Gaussianity. In general, the derivation of the exact expression of the optimal multistep $(l>1)$ predictor of a nonlinear model is not a feasible task and one has to resort to suboptimal solutions.

In the general representation $z_t = f(\boldsymbol{x}_t) + a_t$ we easily find that $\hat{z}_t(1) = f(\boldsymbol{x}_t)$; therefore a *simplified* multistep predictor may be obtained by extrapolating the identified function in the form of a deterministic difference equation, namely

$$\hat{z}_t(l) = f_l\left\{\ldots f_2\left[f_1(\boldsymbol{x}_t)\right]\ldots\right\} = f\left[\tilde{z}_t(l-1)\ldots z_{t+l-p}, a_t\ldots a_{t+l-q}\right] \tag{6.1}$$

The application of this approach to the multilinear model (3.1) involves approximations of the type $\mathrm{E}\left[z_{t+l}z_{t+h}z_{t+k}|z_t z_{t-1}\ldots\right] \approx \hat{z}_t(l)\hat{z}_t(h)\hat{z}_t(k), (l,h,k)>1$, such

as in the one-step-ahead forecast. Hence, when lagged values of $\{a_t\}$ are present in the "regressors", very biased results may be generated and other solutions should be attempted. A possible strategy consists of combining optimal forecasts generated by sub-models.

A practical solution must also be adopted for the variances of prediction errors $\hat{\sigma}^2(l) = \mathrm{E}\left[a_{t+l}^2 | z_t z_{t-1} \ldots\right]$. For $l = 1$ we clearly have $\hat{\sigma}^2(1) = \sigma^2$, but for general prediction horizons we must resort to empirical estimators based on past forecasts $\tilde{z}_\tau(l)$, namely

$$\tilde{\sigma}^2(l) = \sum_{\tau=l}^{t} \left[z_{\tau+l} - \tilde{z}_\tau(l)\right]^2 / (t-l) \qquad l > 1 \tag{6.2}$$

The Application. Given the variability of parameters of the previous models, their forecasting performance turned out to be disappointing. To account for the latent non-stationarity we have then developed the nonlinear modeling on the last part of the IBM series, starting from $t = 272$. Specifically, model identification and estimation were carried out on $\{z_{272} \ldots z_{282} \ldots\ldots z_{355}\}$ and out-of-sample forecasting on $\{z_{351} \ldots z_{369}\}$; the small overlapping of the two sets is motivated by the need of avoiding re-estimation of the models at each change of forecast origin (see below).

Applying the procedures of Sections 3 and 4 we have obtained four submodels which are reported in Table 4 with their iterative estimates. Since NLS estimation substantially confirms the results of the PLR one, we have a concrete evidence of the validity of the latter (implemented with stepsize).

Tab. 4: Sub-models identified on the sub-sample $\{z_{355} \ldots\ldots z_{282} \mid z_{281} \ldots\ldots z_{272}\}$.

a) Quadratic AR	α_{310}	α_{47}	α_{48}	α_{410}	α_{55}	α_{710}	α_{910}	RSS
OLS	.0409	.0484	.0233	-.0324	-.0225	.0329	-.0448	2030
t-stat.	(2.7)	(3.9)	(2.1)	(-2.4)	(-2.7)	(2.8)	(3.6)	
b) Quadratic MA	δ_{16}	δ_{37}	δ_{48}	δ_{55}	δ_{68}	δ_{910}	·	RSS
PLR	.0221	-.0363	.0213	-.0192	.0383	-.0422	·	2825
NLS	.0243	-.0258	.0143	-.0324	.0426	-.0634	·	2750
c) Bilinear SuP.	β_{12}	β_{47}	β_{48}	β_{410}	β_{55}	β_{710}	β_{910}	RSS
PLR	.0557	-.0243	.0428	.0412	-.0340	.0246	-.0467	2360
NLS	.0706	-.0339	.0535	.0342	-.0170	.0255	-.0411	2280
d) Bilinear SuB.	β_{12}	β_{110}	β_{310}	β_{910}	β_{32}	β_{710}	β_{910}	RSS
PLR	-.0221	.0257	-.0505	-.0333	.0227	.0332	.0390	2640
NLS	-.0272	.0155	-.0575	-.0140	.0290	.0301	.0457	2580

Analysis of the prediction ability of the various models has been done with mean absolute forecast errors (MAFE), defined as $\mathrm{MAFE}_n(l|t) = n^{-1}\sum_{i=1}^{n}|\tilde{z}_{t+i}(l) - z_{t+i+l}|$, where $n = 10$ is the size of the mean, $t = 350$ is the first forecast origin, $l = 1, 2 \ldots 10$ are steps ahead. In practice, the forecast origin was shifted 10 times, starting from $t = 350$, and each time 10 predictions were computed. To avoid model re-estimation at each shift, we have included 5 out-of-sample observations in the identification sub-sample.

The graphs of $\mathrm{MAFE}_1(l|350)$, $l = 1 \ldots 10$, corresponding to the models in Table 4 with suboptimal predictor (6.1), are given in Figure 5. Consistently with the value of the residual sum of squares (RSS), the best performance is provided by the quadratic AR model.

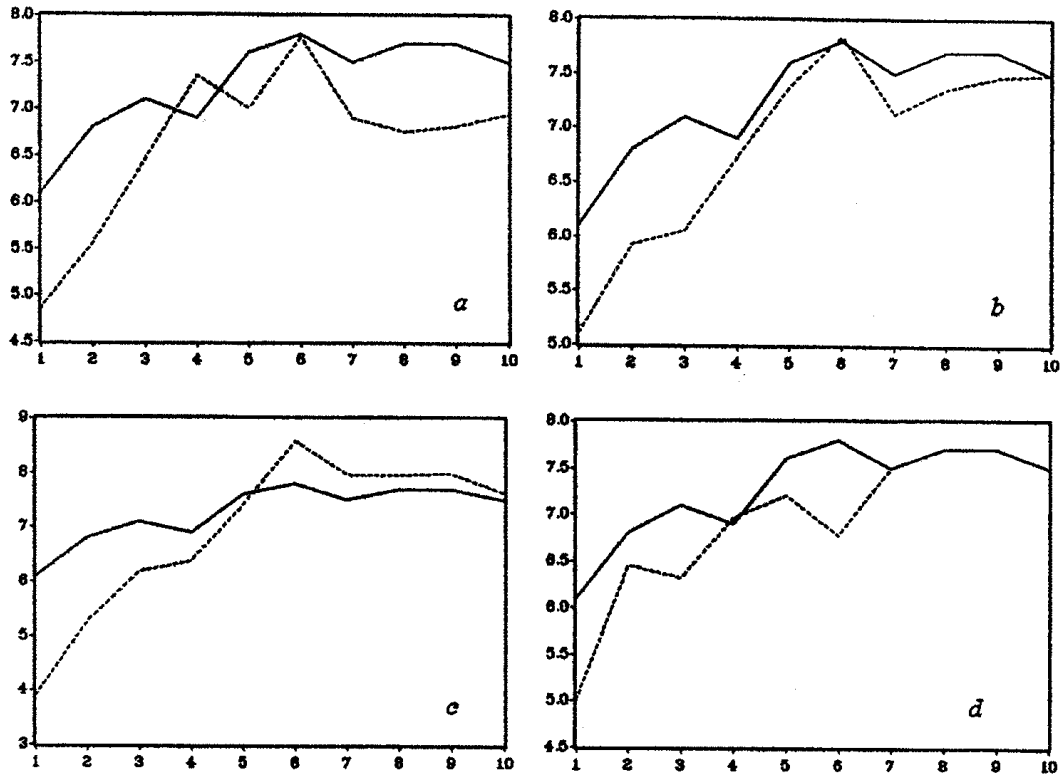

Fig. 5: $\mathrm{MAFE}_{10}(l \mid 350)$ of the Random Walk (——) and of Models in Table 4 (- - -).

In order to improve the above statistics and to approach the nonlinear ARMA model (3.2), we have assembled the quadratic and the bilinear components in Table 4. Results in Table 5 show that the quadratic MA model is nearly "absorbed" by the

AR one. This is due to the lag structure of the two schemes, which is similar (so that the terms $z_{t-k}z_{t-h}$, $a_{t-k}a_{t-h}$ may be competitive), and to the optimal order of the models in Table 4, that approaches 7.

Tab. 5: Partial unification of the models in Table 4.

a) Quadr. ARMA	β_{12}	β_{110}	β_{310}	β_{47}	β_{410}	β_{32}	β_{68}	β_{910}	RSS
PLR	.0393	.0371	-.0383	-.0231	.0356	-.0506	-.0348	.0748	1830
NLS	.0408	.0445	-.0378	-.0208	.0327	-.0449	-.0423	.0661	1780
b) Bilinear	β_{12}	β_{110}	β_{310}	β_{910}	β_{32}	β_{96}	·	·	RSS
PLR	.0470	-.0432	.0638	-.0568	-.0499	.0482	·	·	2375
NLS	.0577	-.0849	.0484	-.0660	-.0706	.0455	·	·	2275

The forecasting ability of the models in Table 5 is shown by Figure 6 . For the bilinear scheme we have used PLR estimates since the NLS ones significantly worsen the long term performance. This result is similar to *overfitting* and is due to the capability of efficient numerical methods to reach parametric solutions that are near the boundary of the invertibility and stationarity regions. By contrast, PLR algorithms provide estimates that are suboptimal, but have the advantage of stability. Moreover, as shown by Granger and Teräsvirta (1991), in nonlinear models the fitting (in-sample) performance may not be indicative of the forecasting (out-of-sample) ability.

Finally, the general model (3.2) is identified by combining the submodels in Table 5; parameter estimates are given in Table 6. Even in this case, the quadratic AR component tends to absorbe the bilinear one and the optimal model order is 7.

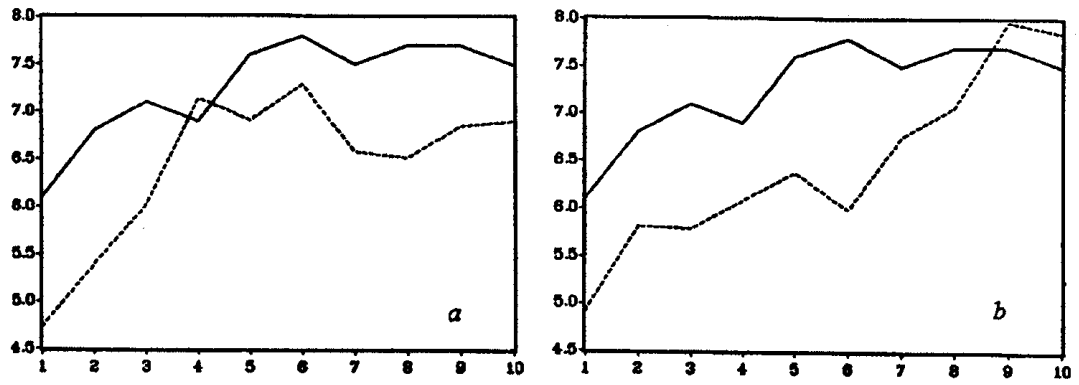

Fig. 6: $MAFE_{10}(l \mid 350)$ of the Random Walk (——) and of Models in Table 5 (- - -).

Since NLS estimation is developed on the results of the PLR identification, one may suspect that the dominance of the quadratic AR model is due to the suboptimal properties of the latter. In other words, there is the possibility that PLR accords priority to the "linear regressors" $z_{t-i}z_{t-j}$. To check this impression we have re-introduced into the model of Table 6, the bilinear monomials discarded in the PLR identification and we have re-estimated by NLS. This experiment has confirmed the non-significance of many $z_{t-i}a_{t-j}$ and the capability of the term $z_{t-3}a_{t-2}$ to yield non-invertibility.

Tab. 6: Unification of the models in Table 5.

Model (3.2)	β_{310}	α_{47}	α_{410}	α_{55}	α_{710}	α_{910}	δ_{37}	δ_{48}	RSS
PLR	.0544	.0386	-.0467	-.0221	.0386	-.0544	-.0512	.0793	1832
NLS	.0519	.0394	-.0418	-.0205	.0386	-.0441	-.0476	.0740	1788

The forecasting ability of the model in Table 6 is shown by Figure 7a . Since it seems inferir to that of Figure 6a, there is the doubt that the strategy of recomposing sub-models is not suitable in forecasting. In other words, it may be preferable to directly combine the forecasts of quadratic and bilinear models. This is done in Figure 7b and the results are significantly better.

The forecasting results displayed in the previous figures are good enough, especially because they are produced by the suboptimal agorithm (6.1). This has also the advantage of yielding stability when the underlying process is stochastically unstable. In particular, in a simulation exercise we have cheked that the model in Table 6 explodes for $a_t \sim IN(0,5)$ where $5 \approx (1832/73)^{1/2}$; however, the corresponding forecasting function (6.1) converges (see Figure 8).

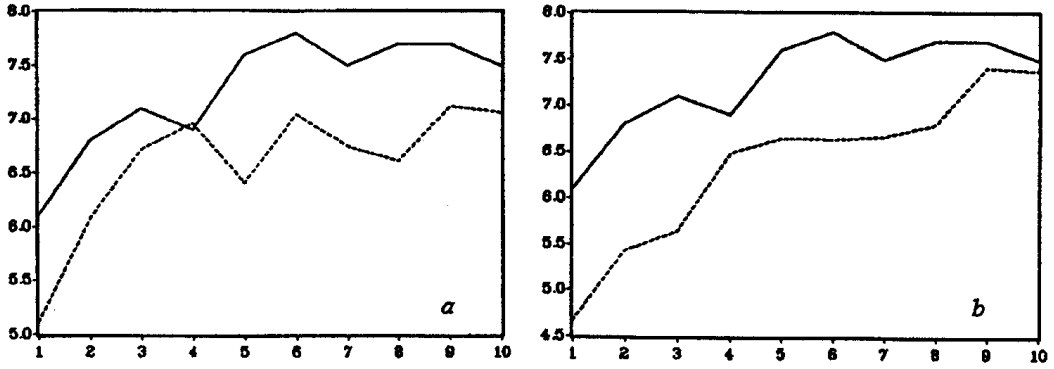

Fig. 7: $MAFE_{10}(l \mid 350)$ generated by the Random Walk (——) and by : (a) the model in Table 6; (b) the combined forecasts of the models in Table 5.

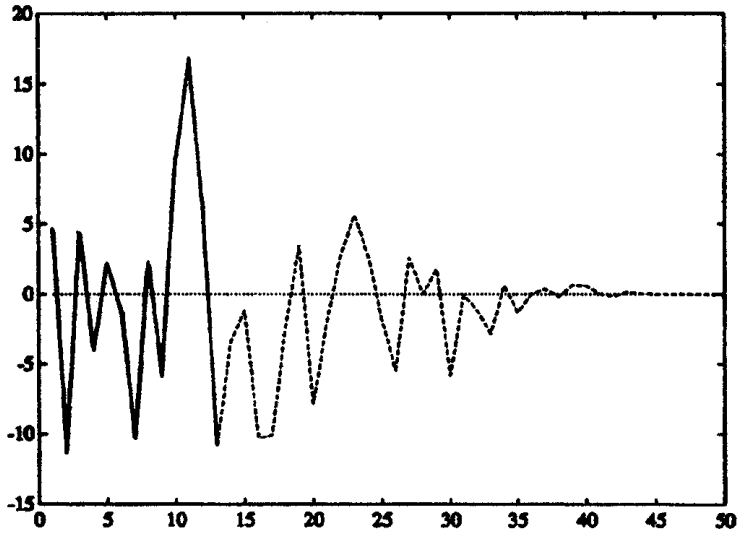

Fig. 8: Convergence of the forecasting function of the model in Table 6.

7. CONCLUSIONS

From previous fitting and forecasting results, we have seen that quadratic AR, quadratic MA and "pure" bilinear models have a similar performance. Given the numerical complexity of the nonlinear estimators (NLS, PLR), and their uncertain statistical properties, the general suggestion of this paper is that nonlinear time series modeling must preferably be developed in terms of polynomial AR models. In particular, it is worth recalling that identification and estimation of these schemes only require correlation functions and linear algorithms (OLS, RLS). Further, since a general modeling must also face the problem of the time variability of the parameters (and in Section 5 we have seen that this involves highly nonlinear estimators (CLS)), it is sensible to keep the basic representation linear in the parameters. Finally, the forecasting functions (exact and approximate), of polynomial AR models are persistently non-zero, allowing for long-run predictions. However, since this may not imply a good forecasting ability, in the case of stable time series it is necessary to check and control convergence of the forecasts.

APPENDIX: PROOF OF RESULT (1.1)

Assume $\{z_t\}$ zero mean and consider its sample bicovariance function $c_3(i,j) = r_{xy}(i,j)\cdot\hat{\sigma}_z\hat{\sigma}_y$, with $y_{t-i,j}=(z_{t-i}z_{t-j})$. For $l = \max(h,k,i,j) > 0$ small we may set $(T-l) \approx T$, so that

$$\begin{aligned}
\mathrm{Cov}[c_3(h,k),c_3(i,j)] &= \mathrm{Cov}\left[\left(\frac{1}{T}\sum_{t=l}^{T} z_t z_{t-h} z_{t-k}\right)\left(\frac{1}{T}\sum_{s=l}^{T} z_s z_{s-i} z_{s-j}\right)\right] \\
&= \frac{1}{T^2}\sum_{t=l}^{T}\sum_{s=l}^{T}\mathrm{Cov}\left[\left(z_t z_{t-h} z_{t-k}\right)\left(z_s z_{s-i} z_{s-j}\right)\right] \\
&= \frac{1}{T^2}\sum_{t=l}^{T}\sum_{s=l}^{T}\mathrm{E}\left\{\left[\left(z_t z_{t-h} z_{t-k}\right)-\mu_3(h,k)\right]\left[\left(z_s z_{s-i} z_{s-j}\right)-\mu_3(i,j)\right]\right\} \\
&= \frac{1}{T^2}\sum_{t=l}^{T}\sum_{s=l}^{T}\mathrm{E}\left[z_t z_{t-h} z_{t-k} z_s z_{s-i} z_{s-j}\right]-\mu_3(h,k)\mu_3(i,j)
\end{aligned}$$

By 6-th order stationarity, and changing the summation variables as $r = (t-s) \in (-T,+T)$ and $\tau = t \in (1,T)$ we may get

$$\begin{aligned}
\mathrm{Cov}[c_3(h,k),c_3(i,j)] &= \frac{1}{T^2}\sum_{t=l}^{T}\sum_{s=l}^{T}\mu_6(t-s,h,k,t-s+i,t-s+j)-\mu_3(h,k)\mu_3(i,j) \\
&= \frac{1}{T^2}\sum_{r=-T}^{T}\left[\mu_6(r,h,k,r+i,r+j)-\mu_3(h,k)\mu_3(i,j)\right]\cdot\sum_{\tau=1}^{N}(1)
\end{aligned}$$

To obtain the variance of the bicovariance estimators, set $h=i$, $k=j$ then

$$\mathrm{Var}[c_3(i,j)] = \frac{1}{T}\sum_{r=-T}^{T}\mu_6(r,i,j,r+i,r+j)-\mu_3(i,j)^2$$

and under the null hypothesis $\mathrm{H}_0: z_t = a_t$, one may easily obtain

$$\mathrm{Var}[c_3(i,j)] = \frac{1}{T}\mu_6(0,i,j,i,j) = \frac{1}{T}\mathrm{E}\left[z_t^2 z_{t-i}^2 z_{t-j}^2\right] = \frac{1}{T}\sigma^6$$

where, by stationarity, a 0-lag term in $\mu_6(\cdot)$ was omitted.

The distribution of (1.1) easily follows by noting that $\sigma_z\sigma_y = \sigma^2$ and that series $\{z_t z_{t-i} z_{t-j}\}$ is an asymptotically independent sequence under the same condition for z_t (see White, 1984). Hence, the central limit theorem for asymptotically independent random variables can be applied to show that

$$\sqrt{T - \max(i,j)} \cdot c_3(i,j) \xrightarrow[H_0]{L} \text{IN}(0,\sigma^3) \quad \text{as} \quad T \to \infty | i,j < \infty$$

Finally, as a corollary, the distribution of the portmanteau statistic $Q_T(\cdot)$ easily follows.

REFERENCES

BOX, G.E.P., JENKINS, G.M., 1976, Time series analysis: forecasting and control (2nd Edition). San Francisco: Holden Day.

BROCKETT, P.L., HINICH, M.J., PATTERSON, D., 1988, Bispectral based tests for the detection of gaussianity and linearity in time series. Jour. of Americ. Stat. Ass., 83, 657-664.

GRANGER, C.W.J., ANDERSEN, A.P., 1978, An introduction to bilinear time series models. G"ttingen: Vandenhoeck and Ruprecht.

GRANGER, C.W.J., TERASVIRTA, T., 1991, Experiments in modeling nonlinear relationships in time series; in CASDAGLI, M., EUBANK, S., (eds.), Nonlinear modeling and forecasting, 189-197. New York: Addison Wesley.

GRILLENZONI, C., 1991, Simultaneous transfer functions versus vector ARMA models. Journal of Forecasting, 10, 477-499.

GRILLENZONI, C., 1993, Multilinear models for nonlinear time series. EUI Working Papers in Economics n.93/2. Florence: European University Institute

GRILLENZONI, C., 1994, Optimal recursive estimation of dynamic models. Jour. of American Statistical Association, 89, 777-787.

HALL, P., HEYDE, C.C., 1980, Martingale limit theorems and their applications. New York: Academic Press.

HANNAN, E.J., McDOUGALL, A.J., 1988, Regression procedures for ARMA estimation. Jour. of American Statistical Association, 83, 490-498.

HINICH, M.J., PATTERSON, D.M., 1985, Identification of the coefficients in a non-linear time series of the quadratic type. Jour. of Econometrics, 30, 269-288.

LI, W.K., 1984, On the autocorrelation structure and identification of some bilinear time series. Jour. of Time Series Analysis, 5, 173-181.

LUKKONEN, R., SAIKKONEN, P., TERASVIRTA, T., 1988, Testing linearity in univariate time series models. Scand. Jour. of Statistics, 15, 161-175.

KEENAN, D.M., 1985, A Tukey non-additivity-type test for nonlinearity. Biometrika, 72, 39-44.

MARAVALL, A., 1983, An application of nonlinear time series forecasting. Jour. of Business and Economic Statistics, 1, 66-74.

MITTNIK, S., 1990, Modeling nonlinear processes with generalized autoregressions. Applied Mathematics Letters, 3, 71-74.

McLEOD, A.I., LI, W.K., 1983, Diagnostic checking ARMA time series models using squared-residuals autocorrelations. Jour. of Time Series Analysis, 4, 269-273.

PETRUCCELLI, J.D., DAVIES, N., 1986, A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. Biometrika, 73, 687-694.

PRIESLEY, M.B., 1988, Nonlinear and nonstationary time series analysis. London: Academic Press.

RAO, M.M., 1961, Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. Annals of Math. Statistics, 32, 195-218.

SUBBA-RAO, T., GABR, M., 1984, An introduction to bispectral analysis and bilinear time series models. Berlin: Springer-Verlag.

TONG, H., 1990, Nonlinear time series: a dynamical system approach. Oxford: Claredon Press.

WHITE, H., 1984, Asymptotic Theory for Econometricians. New York: Academic Press.

PREVISIONI NONLINEARI DI SERIE TEMPORALI FINANZIARIE

Riassunto

Questo articolo sviluppa tecniche statistiche per costruire modelli ARMA multilineari (o polinomiali) per serie temporali nonlineari. In particolare: i test di linearità usano funzioni di multicorrelazione campionarie; le proprietà di stabilità sono investigate per mezzo di simulazioni; la identificazione della struttura dei modelli è basata su regressioni parziali; gli algoritmi di previsione adottano funzioni di estrapolazione deterministiche. Attraverso il lavoro, una estesa applicazione numerica al data-set IBM di Box e Jenkins (1976) illustra e verifica le varie soluzioni.

Parole Chiave: Serie Temporale IBM, Funzioni di Multicorrelazione, Modelli Nonlineari, Regressione Pseudolineare, Algoritmi Recursivi.