

MULTILINEAR MODELS WITH TIME-VARYING PARAMETERS

Carlo Grillenzoni

1. INTRODUCTION

There is no compelling reason to expect social and environmental time series to conform to dynamic models which are linear in the variables. Such schemes, usually having a finite number of nonlinear parameters, are used for ease of statistical analysis, just as the assumption that variates are normally distributed is made for the convenience of mathematical treatment and interpretation.

Time series analysts have begun to turn their attention to the study of nonlinear stochastic processes. Granger and Andersen (1978) and Subba-Rao and Gabr (1984) have developed a class of models, called Bilinear ARMA, that extends the ARMA representation in the same way as the dynamic bilinear systems (see Rugh, 1981). In mathematical terms, the rationale of the approach is given by taking a second order Volterra expansion of the unknown stochastic function that generates the data.

Other models for nonlinear time series exist, *e.g.* exponential and threshold autoregressions (see Priestley, 1988; Tong, 1990, for surveys and comparisons) and, recently, neural networks (see White, 1989). However, the advantage of the bilinear approach is that the resulting equations retain a regression structure so that many algorithms designed for linear models can be applied. On the other hand, its fundamental limit does consist in excluding from the representation nonlinear terms which are produced by Volterra series expansions of higher order.

An attempt to fill this gap has been provided by Hinich and Patterson (1985 a,b) with a class of quadratic innovation models having a nonzero bicovariance (third order cumulant) function. Another extension will be proposed in this paper with a multilinear representation that includes all possible monomial combinations of lagged input and output. This approach is suitable for covering, in a parsimonious manner, the higher order moment information contained in a nonlinear time series. At a theoretical level, its derivation has the same starting point as the state dependent models of Priestley (1988).

The central purpose of the paper is that of providing a model-building framework for multilinear ARMA models. Special attention will then be devoted to

the problems of identification of the dynamic structure (order selection) and estimation of the parameters from sample data. Since empirical models often have irregular (subset) structure and their coefficients are time-varying, suitable technical solutions are given by *partial multicorrelation functions* (in identification) and *recursive pseudolinear regression* in estimation (see Solo, 1978). Loosely speaking, these are generalized moment methods which possess suboptimal properties and are easily implementable.

The paper is organized as follows: section 2 derives a multilinear representation for nonlinear time-series and discusses problems related to its stationarity and stability. Section 3 deals with methodological problems related to nonlinear estimation in the case of time-varying parameters. In section 4 techniques of structure identification based on the inspection of cumulant functions are developed. Finally, section 5 performs an extended numerical application on a real data-set and makes comparisons with other nonlinear models.

2. REPRESENTATION

A natural extension of the autoregressive moving average (ARMA) model $z_t = (\phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}) + a_t$, $a_t \sim \text{IN}(0, \sigma^2)$, toward a representation nonlinear in the variables, can be obtained by considering a general nonlinear function $f(\cdot)$ of the vector of "regressors" $\mathbf{x}'_t = [z_{t-1}, \dots, z_{t-p}, a_{t-1}, \dots, a_{t-q}] = \{x_{it}\}$ (see Priestley, 1988, p. 92)

$$\text{NARMA} \quad z_t = f(z_{t-1}, \dots, z_{t-p}, a_{t-1}, \dots, a_{t-q}) + a_t, \quad a_t \sim \text{IID}(0, \sigma^2) \quad (2.1)$$

With this formulation $\{a_t\}$ play the role of *innovations* of the process $\{z_t\}$ and $f(\mathbf{x}_t)$ that of *projection* on the past history: $E[z_t | \mathcal{F}_{t-1} = (z_{t-1}, z_{t-2}, \dots)]$, in which $\{\mathbf{x}_t\}$ has the role of *state-vector* since contains all the relevant information necessary for prediction.

Instead of proceeding as in the derivation of the state dependent model (SDM, Priestley, 1988, p. 93) - *i.e.* by expanding $f(\cdot)$ in a first order Taylor series about any fixed point \mathbf{x}_{t_0} - we now consider a general expansion about the origin in terms of *Volterra series*. By assuming $f(\cdot)$ analytic (*i.e.* differentiable of every order) around $\mathbf{x}_t = \mathbf{0}$, we may get the expansion

$$z_t = \beta_0 + \sum_{i=1}^m \beta_i x_{it} + \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} x_{it} x_{jt} + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \beta_{ijk} x_{it} x_{jt} x_{kt} + \dots$$

the various sums define, respectively, linear, bilinear, trilinear, ... forms in the pseudoregressors x_{it} , $i = 1, 2, \dots, (p+q) = m$. The explicit constant term $\beta_0 = f(\mathbf{0})$ may have a significant role in the dynamic behaviour of the resulting *multilinear* framework, especially when a finite series has to be used in practice. Here, the Multilinear ARMA model simply arises by truncating the above expansion and taking any subset structure:

$$\begin{aligned}
 \text{MARMA} \quad z_t &= \beta_0 + \sum_{j=1}^n \left[\sum_{k_1 + \dots + k_m = j} \beta_{k_1, \dots, k_m} \prod_{i=1}^m (x_{it})^{k_i} \right] + a_t, \\
 a_t &\sim \text{IID}(0, \sigma^2) \\
 x_{it} &= \{z_{t-i}, a_{t-(i-p)}\} \quad i = 1, 2, \dots, p + q, \\
 \beta_{k_1, \dots, k_m} &= \frac{\partial^j f(\mathbf{x}_t)}{\partial^{k_1} x_{1t} \dots \partial^{k_m} x_{mt}} \Big|_{\mathbf{x}_t = 0}
 \end{aligned} \tag{2.2}$$

the degree of the expansion n depends on the shape of $f(\cdot)$ and requirements of accuracy, the coefficients $\{\beta_{k_1, \dots, k_m}\}$ are subsequences of the Volterra kernels. In certain cases, the justification of (2.2) also stems from ordinary series expansions, as in the following.

Example 1. Consider the simple model $z_t = z_{t-1} \exp(-a_{t-1}) + a_t$. From the expansion $\exp(-a) = \sum_{k=0}^{\infty} (-1)^k a^k / k!$ we get, by truncation, $z_t = z_{t-1} - z_{t-1} a_{t-1} + z_{t-1} a_{t-1}^2 / 2 - z_{t-1} a_{t-1}^3 / 6 + a_t$ which is a multilinear model.

Model (2.2) can also be viewed as a direct generalization of the Bilinear ARMA system, e.g. taking $n = 2$ we may get the quadratic model

$$\begin{aligned}
 z_t = \beta_0 + \left(\sum_{i=1}^p \phi_i z_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + \sum_{i=1}^r \sum_{j=1}^s \beta_{ij} z_{t-i} a_{t-j} + \right. \\
 \left. \sum_{i=1}^P \sum_{j=1}^Q \alpha_{ij} z_{t-i} z_{t-j} + \sum_{i=1}^R \sum_{j=1}^S \delta_{ij} a_{t-i} a_{t-j} \right) + a_t
 \end{aligned} \tag{2.3}$$

where $(r, P, R) \leq p$, $(s, Q, S) \leq q$ and $[\phi_i, \theta_j, \alpha_{ij}, \delta_{ij}]$ are subsets of the coefficients $[\beta_i, \beta_{ij}]$. Quadratic terms like $(z_{t-i} z_{t-j})$, $(a_{t-i} a_{t-j})$ were excluded from the bilinear representation by Granger and Andersen (1978) on the basis of the fact that they may raise difficult problems of stationarity and invertibility. However, Hinich and Patterson (1985 a,b) have shown, on real economic data, the effectiveness of quadratic innovation models of the type

$$z_t = \sum_{i=1}^R \sum_{j=1}^S \delta_{ij} a_{t-i} a_{t-j} + a_t.$$

Unlike the Bilinear ARMA (see Kumar, 1986), this class always generates nonzero third order cumulants $\mu_3(i, j) = E[(z_t - \mu)(z_{t-i} - \mu)(z_{t-j} - \mu)]$, $\mu = E(z_t)$ (bicovariances) and thus it is a candidate to represent more complex processes.

A further generalization of the multilinear framework (2.2) concerns the variability of the regression coefficients over time: $\beta_0(t)$, $\beta_{k_1, \dots, k_m}(t)$. This feature arises from the nonlinear representation (2.1) by assuming that 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 $z_t = f_t(\mathbf{x}_t) + a_t$ and expanding in Volterra series $f_t(\cdot)$ for each t we may get a

representation (2.2) with deterministically varying coefficients. It should be recalled that deterministic is not synonymous of smooth, since $\beta_{k_1, \dots, k_m}(t)$ may not be functions of t , being simply sequences in t .

Stationarity. Stability properties are suitable features for dynamic models, since they determine the reliability of the forecasting and control rules as well as consistency of the parameter estimators. As a general definition of *stochastic stability* we adopt the principle that to inputs $\{a_t\}$ bounded in probability there must correspond outputs $\{z_t\}$ uniformly bounded in probability. While this condition may allow for the existence (finiteness) of some moments, *asymptotic stationarity* is a stronger concept since it enables the same moments to have constant asymptotic expressions. On the other hand, *strict stationarity* does not presume the existence of any moment, since the densities associated to $\{z_t\}$ may be Cauchy with constant parameters.

In past years, most of the theoretical research in nonlinear time series has been concerned with finding *parametric conditions* for the existence of convergent solutions to the various models. Specifically, given $a_t \sim \text{IID}$, if there exists a unique measurable function $g: \mathfrak{R}^\infty \rightarrow \mathfrak{R}$ such that $z_t = g(a_t, a_{t-1}, a_{t-2}, \dots)$ almost surely for all $t = 0, 1, 2, \dots$, then the process $\{z_t\}$ is strictly stationary and ergodic (see Stout, 1974, p. 182). Note, by contrast, that stochastic stability simply requires that for any input bounded in probability the solution $g_t(a_t, a_{t-1}, \dots)$ does not diverge. Recently, a general technique of analysis of the ergodicity has been exploited by Tong (1990, chap. 4), assuming that the nonlinear models be representable in terms of a vector Markov chain $\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{e}_t$, with $f(\cdot)$ analytic and $\mathbf{e}_t \sim \text{IID}(\mathbf{0}, \Sigma < \infty)$. Unfortunately, this feature does not hold for multilinear models, and therefore the Tong's method cannot be applied to equations (2.2)-(2.3).

Even restricting the treatment to bilinear ARMA models (*i.e.* $P = Q = R = S = 0$ in (2.3)), *compact* parametric conditions of stationarity have been established only for particular orders (p, q, r, s) . For the *superdiagonal* model $z_t = \sum_{i=1}^p \phi_i z_{t-i} + \sum_{j=1}^p \beta_{ji} z_{t-j} a_{t-1} + a_t$ with $E(a_t^2) = \sigma^2 < \infty$, that written with a vector notation becomes $z_t = \Phi' z_{t-1} + \beta_1' z_{t-1} a_{t-1} + a_t$, Bhaskara-Rao, Subba-Rao and Walker (1983) have obtained the following conditions

$$\text{stationarity: } \lambda_{\max}(\Phi \otimes \Phi + \mathbf{B} \otimes \mathbf{B} \sigma^2) < 1 \quad (2.4a)$$

$$\text{invertibility: } \beta_1' E(z_{t-1} z_{t-1}') \beta_1 < 1 \quad (2.4b)$$

where $\Phi' = [\phi : \mathbf{I}_{p-1}]$, $\mathbf{B}' = [\beta_1 : \mathbf{O}_{p-1}]$ and $\lambda_{\max}(\cdot)$ denotes the spectral radius. If (2.4a) is satisfied, then the equivalent markovian representation $\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \mathbf{B} \mathbf{z}_{t-1} a_{t-1} + \mathbf{c} a_t$, with $\mathbf{c}' = (1 : \mathbf{0}_{p-1})$, admits the multilinear MA solution $\mathbf{z}_t = \sum_{k=1}^{\infty} \prod_{j=1}^k (\Phi + \mathbf{B} a_{t-j}) \mathbf{c} a_{t-k} + \mathbf{c} a_t$ that converges in mean square for all t . Analo-

gously, under (2.4b) there exists a univ function $b : \mathfrak{R}^\infty \rightarrow \mathfrak{R}$ such that $a_t = b(z_t, z_{t-1}, z_{t-2}, \dots)$ converges with probability one for all t .

Following this approach, Pham (1985) and Liu and Brockwell (1988) have derived stationarity constraints for a general bilinear model, which are very complicated and difficult to check. Anyway, even though conditions (2.4) seem more transparent, their importance remains of "limited" practical value in view of the following remarks: (i) it is not clear what they actually mean in terms of system parameters $\{\phi_i, \beta_{j1}\}$. For example, it can be shown that stationarity in mean is ensured by $\lambda_{\max}(\Phi) < 1$, and this is clearly equivalent to the stability of the AR polynomial $\phi(B)$, but what are the parametric consequences of (2.4b)? (ii) They are not concerned with cumulants and higher order moments involved by nonlinear algorithms of estimation. Conditions for the k -th order stationarity of bilinear models could lead to severe requirements on their parameters, difficult to fulfill in practice. Two examples better illustrate these points.

Example 2. Let $z_t = \beta z_{t-k} a_{t-b} + a_t$ with $k > b > 0$; in this case condition (2.4b) means $\beta^2 \cdot E(z_{t-k}^2) < 1$. Squaring z_{t-k} we get the difference equation $E(z_{t-k}^2) = \beta^2 \sigma^2 E(z_{t-2k}^2) + \sigma^2$; if the condition of stationarity (2.4a) holds, i.e. $\beta^2 \sigma^2 < 1$, the asymptotic solution of this equation leads to the invertibility requirement $\beta^2 \sigma^2 / (1 - \beta^2 \sigma^2) < 1$, i.e. $\beta^2 \sigma^2 < .5$.

Example 3. Consider the above example with $a_t \sim \text{IN}(0, \sigma^2)$ gaussian; having $E(z_{t-k} a_{t-b}) = 0$ taking fourth moments, we get $E(z_t^4) = (\beta^4 3 \sigma^4) E(z_{t-k}^4) + 3 \sigma^4$. Solving for this difference equation, a necessary condition for the 4-th order stationarity of $\{z_t\}$ becomes $\beta^2 \sigma^2 < 1/\sqrt{3} \approx .6$. As for invertibility this condition is stronger than (2.4a).

When autoregressive components are present, the constraints tend to become even more severe. For example, in the model $z_t = \phi z_{t-1} + \beta z_{t-1} a_{t-1} + a_t$ the existence of 4-th order moments requires $[\phi^4 + 6(\phi\beta\sigma)^2 + 3(\beta\sigma)^4] < 1$ (see Sesay and Subba-Rao, 1988). It is then clear that conditions of stationarity for complex non-linear models not only are difficult to establish but may not exist at all - that is, the *region of stationarity* in the parameter space might be empty, or nearly so.

These comments tend to discourage the analysis of the stability properties of multi-linear models (2.2)-(2.3). Granger and Andersen (1978) have heuristically shown the virtual nonstationarity and non-invertibility of the schemes $z_t = \alpha z_{t-1}^2 + a_t$, $z_t = \delta a_{t-1}^2 + a_t$; but, unless the contrary is proved, one cannot exclude that suitable properties may hold, even though locally and for particular realizations, for the quadratic model (2.3). With respect to the nonlinear system (2.1), if $|f(\mathbf{x}_t)| < \|\mathbf{x}_t\|$, i.e. $f(\cdot)$ is a contraction mapping, the process $\{z_t\}$ is stochastically stable. Examples of this kind are given by systems that can be reduced in the form of rational transfer functions, e.g.

$$(1 + \delta z_{t-1} a_{t-1}) z_t = (\alpha - \beta z_{t-1} a_{t-1}) z_{t-b} + a_t (1 + \delta z_{t-1} a_{t-1}), \quad \left| \frac{\alpha - \beta z_{t-1} a_{t-1}}{1 + \delta z_{t-1} a_{t-1}} \right| < 1.$$

However, these models are nonlinear in the parameters and thus require complicated algorithms of identification, estimation and forecasting. The peculiar feature of representation (2.2) is given by its regression structure which allows for direct application of many recursive procedures of standard time series analysis.

Nonstationarity. Realistic considerations that may mitigate the picture of uncertainty outlined so far, are given by the following remarks: (a) Stationarity properties are certainly suitable features, but they are concerned with abstract asymptotic behaviour of the output of the models. In modeling real data, users typically deal with finite sampling intervals and many observational time series (mostly in economics) are nearly explosive in nature. (b) As shown in the analysis of linear AR models, stability properties are sufficient but non-necessary conditions for the existence of consistent estimators of the parameters. Paradoxically, conventional least squares and maximum likelihood methods improve their speed of convergence in presence of roots on or outside the unit circle (see Rao, 1961), although they do not retain asymptotic normality. (c) If parameters of the models are time-varying, issues of convergence and stationarity do not arise by definition; moreover, the change of the "regression" coefficients may have a stabilizing effect on the behaviour of the output $\{z_t\}$. Specifically, even if the region of stationarity of the constant parameter model is empty, or nearly so, there may exist sequences of time-varying coefficients $\{\beta_t\}$ which force $\{z_t\}$ to be bounded in probability.

Example 4. Situations of this kind can be illustrated by simulations. In 10 experiments of sample size 100,000 (which are equivalent to 1000 replications of length 1000) we have assessed that the quadratic process $z_t = \alpha z_{t-1}^2 + a_t$, with $z_0 = 0$ and $a_t \sim \text{IN}(0, \sigma^2)$ tends to overflow for $|\alpha| > .16$. However, if the model is time-varying with parameter function $\alpha_t = \alpha_0 \text{sign}(a_t)$, where $\text{sign}(x) = +, -1$ for $x >, < 0$, the critical value becomes $\alpha_0 > .26$. In practice, there are infinite trajectories of $\{\alpha_t\}$, completely laying outside the region of stability, which confine $\{z_t\}$ within finite bounds.

In the sequel we shall use the additional degree of freedom represented by the variability of the coefficients for enlarging the region of stability of the multilinear models and for weakening the parametric conditions of existence of their moments.

Assumptions A. Consider the class of processes (2.2) with deterministically varying coefficients $\{\beta_t\}$ and distribution functions $F_t(\cdot)$. Then, for every input $\{a_t\}$ bounded in probability [i.e. $\sup_t P(|a_t| = \infty) = 0$], there exist trajectories of $\{\beta_t\}$ such that the output $\{z_t\}$ is an *asymptotically independent* and a *K-th order process*. In notation this means:

$$\phi(m) = \sup_t \sup_n |F(z_{t_1}, \dots, z_{t_n} | z_{t_1-m}, \dots, z_{t_n-m}) - F(z_{t_1}, \dots, z_{t_n})| \rightarrow 0$$

as $m \rightarrow \infty$ (A1)

$$\sup_t P(|z_t| \geq z) = o(1/z^{K+\delta}) \text{ as } |z| \rightarrow \infty \text{ with } \delta > 0. \tag{A2}$$

Condition (A1) is a particular version of the so-called ϕ -mixing property (see Stout, 1974) and is satisfied by taking $\phi(m) = O(1/m^b)$ with $b > 0$. Condition (A2) enables the existence of moments of order $k \leq K$ (see Laha and Rohatgi, 1979, p. 38), that is $\sup_t E|z_t|^k < \infty$ for all $k \leq K$. For identification purposes we also need the following regularity conditions.

Assumptions B. For any model that satisfies Conditions (A), the sequence of coefficients $\{\beta_t\}$ is bounded and has a well defined *time-average behaviour*, namely:

$$0 < |\bar{\beta} = \bar{E}(\beta_t)| < \infty, \text{ with } \bar{E}(\beta_t) = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{t=1}^N \beta_t \right), \quad t = 1, 2, \dots \tag{B1}$$

Moreover, it enables the process $\{z_t\}$ to be *quasi-stationary of order K*, that is:

$$0 < |\bar{\mu}_k = \bar{E}(z_t^k)| < \infty, \text{ with } \bar{E}(z_t^k) = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum_{t=1}^N E(z_t^k) \right], \quad k \leq K \tag{B2}$$

It is important to stress the structure of the asymptotic average operator $\bar{E}(\cdot)$ since it has a fundamental role in defining suitable parameters for off-line inference.

Forecasting. We conclude the section with some issues of forecasting in multilinear models. Given a model and the set of information \mathfrak{S}_t up to time t , the task is to find the expression of $z_t(l)$, the predictor of z_{t+l} optimal in MSE sense. It is well known that

$$\hat{z}_t(l) = \operatorname{argmin} E\{[z_{t+l} - z_t(l) | \mathfrak{S}_t]^2\} = E[z_{t+l} | \mathfrak{S}_t]$$

but as shown in the bilinear context the calculation of this expectation is strongly affected by the presence of nonlinear regressors such as $(z_{t-b} a_{t-k})$, $b < k$. These difficulties increase in multilinear systems, so that suboptimal and pragmatic solutions must be sought.

In the general nonlinear process (2.1) we easily find $\hat{z}_t(1) = f(\mathbf{x}_t)$, therefore a simplified multistep predictor can be obtained by extrapolating the identified function in the form of a deterministic difference equation, namely

$$\hat{z}_t(l) = f_l\{ \dots f_2[f_1(\mathbf{x}_t)] \dots \} = f[\hat{z}_t(l-1), \dots, z_{t+l-p}, a_t, \dots, a_{t+l-q}].$$

The application of this approach to the multilinear model (2.2) involves approximations of the kind $E[z_{t+l} z_{t+h} z_{t+k} | \mathfrak{S}_t] \approx \hat{z}_t(l) \hat{z}_t(h) \hat{z}_t(k)$, $(l, h, k) > 1$, such as in the one step a head forecast. However, when lagged values of $\{a_t\}$ are present in the "regressors", bad results may be generated and other solutions must be attempted. In practice, the above strategy corresponds to combining forecasts generated by sub-models.

Example 4. Consider the model $z_t = \alpha z_{t-1}^2 + \beta z_{t-1} z_{t-2} + a_t$. The optimal one step ahead predictor is $\hat{z}_t(1) = (z_{t+1} - a_{t+1})$; similarly $\hat{z}_t(2) = \alpha E[z_{t+1}^2 | \mathcal{S}_t] + \beta E[z_{t+1} | \mathcal{S}_t] \cdot z_t$, where $E[z_{t+1}^2 | \mathcal{S}_t] = E[(\alpha z_t^2 + \beta z_t z_{t-1})^2 + 2(\alpha z_t^2 + \beta z_t z_{t-1})a_{t+1} + a_{t+1}^2 | \mathcal{S}_t] = (\alpha z_t^2 + \beta z_t z_{t-1})^2 + \sigma^2$, hence $\hat{z}_t(2) = \alpha \hat{z}_t(1)^2 + \alpha \sigma^2 + \beta \hat{z}_t(1) \cdot z_t$. Prediction based on sub-models gives $\tilde{z}_t(2) = \hat{z}_t(2) - \alpha \sigma^2$.

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

$$\tilde{\sigma}^2(l) = \sum_{\tau=1}^t [z_{\tau+l} - \tilde{z}_\tau(l)]^2 / (t-l), \quad l > 1$$

3. ESTIMATION

The identification procedure that we outline in the next section will require the existence of an efficient estimator for the parameters of multilinear models. If the distribution of the input process $\{a_t\}$ is known *a priori* a natural candidate is the maximum likelihood method; however this condition contradicts the general formulation (2.1) where $a_t \sim \text{IID}$. In this section we adopt and extend the least-squares approach followed by Subba-Rao and Gabr (1984) for bilinear models.

In order to simplify the treatment, we rewrite model (2.2) in regression form and assume a *simplified* structure for its monomials, such as

$$\text{MARMA}^* \quad z_t = \beta_0 + \sum_{i=1}^n \beta_i y_{it} + a_t, \quad y_{it} = \left(\prod_{j=1}^{p_i} z_{t-j} \cdot \prod_{j=1}^{q_i} a_{t-j} \right) \quad (3.1)$$

with the general constraint $(p_i + q_i) \leq (p_k + q_k)$ for $i < k$. Resorting to nonlinear least squares (NLS) method, *i.e.* letting $\beta' = [\beta_0, \beta_1, \dots, \beta_n]$ and defining

$$\hat{\beta}_N = \arg \min \left[J(\beta) = \sum_{t=1}^N a_t^2(\beta) \right], \quad a_t \sim \text{IID}(0, \sigma^2)$$

improves the situation with respect to the ML-approach. However, the analytical expression of the gradient of the Gauss-Newton algorithm still remains complicate

$$\text{NLS} \quad \hat{\beta}_N(k+1) = \hat{\beta}_N(k) + \left[\sum_{t=1}^N \hat{\xi}_t(k) \hat{\xi}_t(k)' \right]^{-1} \sum_{t=1}^N \hat{\xi}_t(k) \hat{a}_t(k) \quad (3.2a)$$

$$\xi_t(\beta) = -\frac{\partial a_t}{\partial \beta} = y_t - \sum_{j=1}^q \left[\beta' \frac{\partial y_t}{\partial a_{t-j}} \right] \xi_{t-j} \quad (3.2b)$$

where $y_t(\beta)' = [1, y_{1t}, \dots, y_{nt}]$ and $q = \max(q_i)$ (see Grillenzoni, 1993 for the deri-

vation). From (3.2b) we note that calculation of the gradient consists of a filtering on the vector of the pseudo-regressors y_t , but with a filter that depends on random variables. This makes explicit the dependence of the algorithm (3.2a) on higher order moments of $\{z_t\}$.

In the linear ARMA context we simply have $\xi_t = x_t/\theta(B)$, i.e. $\xi_t = x_t - \sum_{j=1}^q \theta_j \xi_{t-j}$.

By assuming $x'_t = [z_{t-1}, \dots, a_{t-q}]$ stationary gaussian, we also have $\{\xi_t\}$ covariance stationary and therefore $\{\xi_t \xi'_t\}$ is stationary in mean. Since $\hat{\beta}_N(k)$ is a minimizer, its consistency can be proved by applying the ergodic theorem to

$t^{-1} \sum_{\tau=1}^t \hat{\xi}(\tau) \hat{a}(\tau) \rightarrow 0$ where $t = (k = N)$. In (3.2b) $\{\xi_t\}$ is not a linear transforma-

tion of $\{y_t\}$ and this in turn could not be stationary in covariance - for $\{z_t\}$ should be $2p$ -th order stationary, with $p = \max(p_i)$. The divergence of the estimator (3.2a) may then follow simply because the process (3.2b) has not second order moments. For a multilinear model the problem of invertibility is even more urgent than that of stationarity, since it enables the iterative estimates to be computed. A general observation is that, unless the range of $\{z_t\}$ is restricted

within certain bounds, it is impossible to identify terms like $\delta_i \prod_{j=1}^m a_{t-j}^{k_j}$, using

algorithms that involve the calculation of residuals. Also in this case, however, we may resort to the stabilizing effect induced by time-varying coefficients, and assume the existence of trajectories of $\{\beta_t\}$ which allow the stability of the filtering $a_t = h_t(z_t, z_{t-1}, z_{t-2}, \dots) = z_t - f_t(x_t)$.

Off-line Inference. Referring to time-varying parameter models, we now address the important problem of finding a convergent off-line estimator for the mean value $\bar{\beta} = E(\beta_t)$. This is, indeed, the sole question that can be posed, in the off-line inference, with evolving models. As we shall show, even restricting the analysis to multilinear AR models, consistent estimators may only exist under Assumptions (A)-(B) and other conditions concerning the behaviour of the parameter function. The system of reference for the analysis is given below, in

which $\left(\sum_{j=1}^{p_i} k_{ij} \right) \leq \left(\sum_{j=1}^{p_b} k_{bj} \right)$ for $i < b$ and $k_{ij} \geq 0$

$$\text{MAR}_t \quad z_t = y'_t \beta_t + a_t, \quad (a_t | \mathcal{S}_{t-1}) \sim \text{IID}(0, \sigma_t^2) \quad (3.3)$$

$$y_{it} = \left(\sum_{j=1}^{p_i} z_{t-j}^{k_{ij}} \right), \quad \{|\beta_{it}|\} < \infty, \quad i = 1, \dots, n.$$

Since it is linear in the parameters, we may consider, as an estimator for $\bar{\beta}$, the ordinary least squares method

$$\hat{\beta}_N = \left(\frac{1}{N} \sum_{t=1}^N y_t y'_t \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^N y_t z_t \right) = \arg \min \bar{E}(a_t). \quad (3.4)$$

Assumptions C. In order to derive this estimator and to establish its consistency with respect to the parameter $\bar{\beta}$, some *form of orthogonality* between the parameter function $\{\beta_t\}$ and the variance function of $\{y_t\}$ is needed, namely

$$\overline{E(y_t y_t') \beta_t} = \overline{E(y_t y_t')} \cdot \overline{E(\beta_t)} \quad (C1)$$

$$\overline{E(a_t^2)} < \infty, \quad \overline{E(y_t a_t)} = 0, \quad \overline{E(y_t y_t')} > 0. \quad (C2)$$

Indeed, multiplying model (3.3) by y_t and taking expectation we get $E(y_t z_t) = E(y_t y_t') \beta_t + E(y_t a_t)$, next applying the operator $\overline{E[\cdot]}$, under the above conditions we may get

$$\bar{\beta} = \overline{E(y_t y_t')}^{-1} \overline{E(y_t z_t)}. \quad (3.5)$$

Notice that, in general, given two deterministic bounded functions, the mean value of their product does not coincide with the product of their means. An immediate example is given by taking $f(t) = \sin(t)$, $g(t) = -\sin(t)$, in which $\bar{f} = \bar{g} = 0$, whereas $h(t) = f(t) \cdot g(t)$ is negative nearly everywhere. Hence, assumption (C1) may be very difficult to check and satisfy in practice; despite of this, its admissibility stems from the fact that $\{\beta_t\}$ is deterministic and has a stabilizing effect on $\{z_t\}$ (therefore it tends to move rapidly). Otherwise, (C1) always holds if $\{\beta_t\}$ is constant, periodic or monotonic.

On the basis of assumptions stated so far we have the following formal result:

Proposition 1. Consider the time-varying multilinear AR model (3.3) in which:

I) assumptions (A1), (A2) are satisfied, in particular:

$$(A1) \quad \text{with } \phi(m) = O(1/m^b), \quad b > r/(2r-1), \quad r \geq 1 \text{ and}$$

$$(A2) \quad \text{with } K = (2\kappa)(r + \delta), \quad \kappa = \max_i \left(\sum_{j=1}^{p_i} k_{ij} \right), \quad 0, d \leq r;$$

II) assumptions (B1), (B2) hold for every order $k \leq K$;

III) conditions (C1), (C2) are satisfied;

IV) the sequence $\overline{E}_N(y_t y_t') = \left\{ N^{-1} \sum_{t=1}^N E(y_{it} y_{it}') \right\}$ is uniformly positive definite.

Then for N sufficiently large the OLS estimator (3.4) exists with probability one and is consistent for the average trajectory $\bar{\beta} = \overline{E(\beta_t)}$.

Proof. The proof is not short and requires some auxiliary results concerning the law of large number and the transformations of mixing sequences; it is given in Grillenzoni (1993). In simulation experiments we have checked that OLS is an accurate estimator for $\bar{\beta}$ if the sequence $\{\beta_t\}$ is not near the border of the (extended) stability region. Anyway, in the case of time-varying parameters off-line estimators are not the proper ones.

On-line Inference. Returning to the system (3.1), we note that its regression

structure also enables the application of *pseudo-linear regression* (PLR) methods in the estimation. These methods simply come from approximating the gradient (3.2b) as $\xi_t(\beta) \approx y_t(\beta)$ and inserting the corresponding iterative expression $\hat{y}_t(k)$, together with $\hat{a}_t(k) = z_t - \hat{y}_t(k)' \hat{\beta}(k)$, in (3.2a); the final algorithm has the same structure as the OLS (3.4), but is iterative. In the context 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 case of linear models (see Hannan and McDougall, 1988 and Grillenzoni, 1990), the approximation of the gradient makes the resulting estimators not always consistent and generally inefficient. Utilization of the PLR approach should then be limited to recursive (on-line) methods, applied for tracking the sequence of parameters $\{\beta_t\}$ in time-varying models. In this context, questions of stationarity and convergence do not matter (by definition) and the adaptive properties of PLR, allowed by the greater computational speed, are preferable to those of accuracy of NLS.

Proceeding as in Solo (1978) or Grillenzoni (1990), by equating ($k = N$) = t in (3.2a) and with $\hat{\xi}_t(k)$ replaced by $\hat{y}_t(k)$, the Recursive PLR estimator of (3.1) becomes

$$\bar{a}(t) = z_t - \hat{\beta}(t-1)' \hat{y}(t) \tag{3.6a}$$

$$R(t) = \lambda \cdot R(t-1) + \hat{y}(t) \hat{y}(t)' \tag{3.6b}$$

$$\text{R-PLR} \quad \hat{\beta}(t) = \hat{\beta}(t-1) + R(t)^{-1} \hat{y}(t) \bar{a}(t) \tag{3.6c}$$

$$\hat{a}(t) = z_t - \hat{\beta}(t)' \hat{y}(t) \tag{3.6d}$$

$$J(t) = \lambda \cdot J(t-1) + \hat{a}(t)^2 \tag{3.6e}$$

$$\hat{y}(t+1) = \left\{ \prod_{j=1}^{p_t} z_{t+1-j} \prod_{j=1}^{q_t} \hat{a}(t+1-j) \right\}. \tag{3.6f}$$

The terms \bar{a}_t , \hat{a}_t are respectively the prediction error and the recursive residual; the factor $0 < \lambda < 1$ by preventing $R(t)$ from vanishing, enables parameter changes ($\beta_t - \beta_{t-1}$) to be tracked. Since $R(t)$ is an approximation of the Hessian matrix, λ should be designed to provide a suitable compromise between unbiasedness (fast tracking) and estimation accuracy. Finally, $\hat{\sigma}(t)^2 = (1 - \lambda)J(t)$ provides an on-line estimator for σ_t^2 .

Returning to the efficient NLS estimation, the exact recursive expression of (3.2) can be obtained from (3.6) by replacing $\hat{y}(t)$ with $\hat{\xi}(t)$ in the equations of $R(t)$, $\hat{\beta}(t)$ and inserting the filter

$$\hat{\xi}(t) = \hat{y}(t) - \sum_{j=1}^q \left[\hat{\beta}(t)' \frac{\partial \hat{y}(t)}{\partial \hat{a}(t-j)} \right] \hat{\xi}(t-j). \tag{3.6g}$$

Under the assumption of constant parameters the resulting algorithm tends to minimize the weighted functional $J_t = \sum_{\tau=1}^t \lambda^{t-\tau} a_\tau^2(\beta)$; however, in the context of

evolving systems, it is not clear what may be the improvement in terms of the MSE $E\|\hat{\beta}(t) - \beta_t\|^2$. Given the complexity and multilinearity of the filtering (3.6g), a worsening of the tracking capability with respect to (3.6c) cannot be ruled out.

The parameters of (3.1) may vary with time depending on the goodness with which the multilinear model (2.2) approximates the "true" nonlinear function (2.1). This situation can be illustrated with a simple example.

Example 5. Consider the bilinear system $z_t = \beta z_{t-1} a_{t-2} + z_{t-1} a_{t-1} + a_t$; this can easily be decomposed into a time-varying AR(1) model $z_t = \phi_{t-1} z_{t-1} + a_t$ whose parameter behaves like an MA(1) process $\phi_t = \beta a_{t-1} + a_t$ with the same input. Hence, whenever a nonlinear model is treated as linear, stochastic variability of parameters occurs.

In certain circumstances the lack of nonlinear representation may then be rectified by admitting that the model is time-varying and by estimating its coefficients on-line. With respect to the Kalman Filter approach, algorithm (3.6) is much more easy to implement since it only requires as *priors* $0 < \lambda < 1$, $\mathbf{R}(0)^{-1} = \rho \cdot \mathbf{I}_n$ (and usually $.95 < \lambda < .99$, $\rho = 1$); moreover, it only assumes that parameters do not change suddenly as jump functions. To be more specific, while Kalman Filter implementation requires that parameters follow a linear (or a linearizable) process, recursive algorithms, by making estimates smooth functions of past observations, implicitly assume $\beta_t = f(\mathfrak{S}_{t-1})$. The weighting sequence $\{\lambda(t, \tau)\}$ should then be designed according to the path of conditional probabilities $P_t(\beta_t | z_{t-\tau})$ or to that of cross correlations $\text{Cor}(\beta_t, z_{t-\tau})$. These informations, however, are not available *a priori* and other, more pragmatic, criteria must be used.

A way of avoiding altogether the problem of priors in the recursive estimation, consists of reducing algorithm (3.6) to a stochastic approximation scheme. This may be approached by setting $\lambda \equiv 1$ and replacing $\mathbf{R}(t)$ by $\bar{\mathbf{R}}(t) = \mathbf{R}(t)/t$, so as to retain the tracking capability. In this case the updating rule equivalent to (3.6b) becomes

$$\bar{\mathbf{R}}(t) = \bar{\mathbf{R}}(t-1) + \frac{1}{t} [\bar{\mathbf{R}}(t-1) - \hat{y}(t)\hat{y}(t)']$$

and k -iterating the recursions one may initialize $\bar{\mathbf{R}}_{k+1}(0) = \bar{\mathbf{R}}_k(N)$. A stochastic approximation type solution is achieved if $\bar{\mathbf{R}}(t)$ converges to a matrix $0 < \bar{\mathbf{R}} < \infty$ as $t \rightarrow \infty$; however, for problems explained above this may be guaranteed only for MAR models (3.3).

4. IDENTIFICATION

Once the dynamic representation and the estimation methods have been defined, a crucial phase in the modeling is given by the identification of the *orders*. With respect to the class (2.2), the task is difficult since requires the

definition of the structure of monomials $\left\{ y_{jt} = \prod_{i=1}^p z_{t-i}^{k_{ij}} \prod_{i=1}^q a_{t-i}^{b_{ij}} \right\} i = 1, \dots, n$, i.e.

of the powers k_{ij} , b_{ij} . For the sub-class (2.3), which is more regular, two techniques developed for linear and bilinear models may be used.

1) *Parametric*. By assuming $a_t \sim \text{IN}(0, \sigma^2)$, independent normal, the orders are selected by minimizing some information criterion $\text{IC} = -2 \log(\text{likelihood}) + f(N) \text{dim}(\text{model})$:

$$\hat{n} = \arg \min \left[\log(\hat{\sigma}^2) + n \cdot \frac{f(N-d)}{(N-d)} \right], \quad n = (p+q+r \cdot s + P \cdot Q + R \cdot S) \quad (4.1)$$

where $(N-d)$, $d = \max(p, r, P, R)$ is the effective number of observations used for calculating the maximum of the log-likelihood $-(N-d) \log \sigma^2/2$, and normalizing the IC. The function $f(N-d)$ is what characterizes the kind of IC used in practice; Akaike, Schwarz, Hannan and Quinn have suggested, respectively, $f(N) = 2, \log(N), \log(\log(N))$.

2) *Nonparametric*. This approach simply assumes $a_t \sim \text{IID}(0, \sigma^2)$, and it selects models by comparing the sample behaviour of some higher order moments with those theoretically generated by a class of low order models.

Example 6. Let $z_t = \delta a_{t-b} a_{t-k} + a_t$, $b < k$; simple calculation shows that $\{z_t\}$ is white noise. However, third order moments $\mu_3 = (i, j) = E[z_t z_{t+i} z_{t+j}]$ have six nonzero values, namely $\mu_3(b, k) = \mu_3(-b, -k) = \mu_3(k-b, -b) = \delta \sigma^4$ plus their permutation symmetries.

Both these approaches are of limited practical utility since they rely heavily on the assumption that a *true* (regular) multilinear system exists. By contrast, data are often generated by irregular (subset) models, having sparse coefficients at strange lags. The main consequences are that the estimation of information criteria may fail owing to the presence of many insignificant and collinear terms, which make the Hessian matrix associated with the nonlinear estimator, ill-conditioned. Secondly, analysis of the theoretical multicovariance functions, related to all the subset alternatives of (2.2), is practically impossible and some patterns are shared by different model structures.

The identification procedure that we now propose stems from viewing the multilinear system as an ARMAX model whose inputs are given by the monomials

$$y_{jt} = \prod_i z_{t-i} \prod_i a_{t-i}, \text{ namely } \phi(B)z_t = \beta_0 + \sum_{j=m}^n \beta_j y_{jt} + \theta(B)a_t.$$

Assuming that with multiple products the series $\{y_{jt}\}_m^n$ acquire an *autonomous* nature with respect to the output $\{z_t\}$, the selection of significant regressors may be developed on the inspection of the "cross covariances" $E[z_t y_{jt-k}] = \gamma_{zj}(k)$, as discussed in Grillenzoni (1991). The *strategy* of putting coefficients β_{jk} in correspondence of every significant multicovariance $\gamma_{zj}(k)$ yields exact identification only in the case of multilinear MA models (see Example 7 below). In the other cases, it is approximate and leads to overparametrization; however, it drastically reduces the number of terms to be considered in the estimation of information criteria.

Example 7. Consider the model of Hinich-Patterson: $z_t = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{ij} a_{t-i} a_{t-j} + a_t$ with $a_t \sim \text{IID}$; by defining $y_{t-b,k} = a_{t-b} a_{t-k}$ we clearly have $E[z_t y_{t-b,k}] = \delta_{bk} \sigma^4$.

Similarly, for a general MA system we have that $E\left[z_t \left(y_{jt} = \prod_{i=1}^{q_j} a_{t-i}\right)\right] \neq 0$ implies $\beta_j \neq 0$ for all j .

If $\{a\}$ were be an observable input, there would be no problems to implement the above identification strategy by estimating the cross multicorrelations $\rho_{z_j}(k) = \gamma_{z_j}(k) / \sigma_z \sigma_j$ (see Priestley, 1988, p. 38). In our (univariate) context, however, a procedure must be sought to generate $\{a_t\}$ before the identification process. A natural solution, exploited in many tests for nonlinearity (see Keenan, 1985), is provided by the innovations $\tilde{a}_t = \theta(B)^{-1} \phi(B) z_t$ of the linear part of the model. The degree of approximation depends on how much $\{\tilde{a}\}$ is an independent sequence; here, it should be recalled that innovations processes are also asymptotically independent (see Spanos, 1986, p. 148).

Substantial problems arise in extending the previous strategy to models of type (2.2). The requirement $\gamma_{z_j}(0) \neq 0$ would be, in fact, only a necessary condition for the existence of $\beta_j \neq 0$ since this implies $\gamma_{z_j}(k) \neq 0$ for some $k \neq 0$. There is, however, the possibility of identifying "spurious regressors" by noting that, under stability conditions, $\gamma_{z_j}(0)$ usually provides the greatest value; that is $|E(z_t y_{jt})| \geq |E(z_t y_{jt-k})|$ for $k > 0$. Moreover, the relationship covariance-coefficient may be strengthened by referring to *partial* multicovariances. In system (2.3), for example, the selection of $\alpha_{ij} \neq 0$, should refer to the partial bicovariances $E[z_t y_{t-i,j} | y_{t-i,j+k}, k > 0]$ where $y_{t-i,j+k} = z_{t-i+k} z_{t-j+k}$. Explanation of these remarks are given by the following example.

Example 8. Consider the model $(1 - \phi B)z_t = \alpha_{ij} z_{t-i} z_{t-j} + a_t$ with $|\phi| < 1$ and stationary up to moments of 4-th order. Now expanding $z_t = \sum_{k=0}^{\infty} v_{ijk} z_{t-i-k} z_{t-j-k} + n_t$, with $v_{ijk} = \alpha_{ij} \phi^k$ and $n_t = a_t / \phi(B)$, it is clear that the diagonal cumulant function $\gamma_{z_j}(k) = \mu_3(i+k, j+k)$, where $y_t = z_{t-i} z_{t-j}$, is decreasing and has a maximum at $k = 0$.

An efficient identification method is that of stepwise regression in which the intermediate information provided by partial multicorrelations $\rho_{z_j|1, \dots, j-1} \propto E(z_t y_{jt} | y_{1t}, \dots, y_{j-1t})$ is used to select the most appropriate regressors to be included in the model. If a coefficient that was significant at an earlier stage, later becomes insignificant (after some other inclusions) then the corresponding pseudolinear regressor is deleted. In the context of nonlinear ARMA systems calculation of $\rho_{z_j|1, \dots, j-1}$ requires the estimation of the model $z_t = \beta_0 + \sum_{i=1}^{j-1} \beta_{ji} y_{it} + a_{jt}$ at each step $j = 2, 3, \dots, n$, both in order to check the significance of the included coefficients as well as to generate the next candidate "regressors" $\hat{y}_{jt} = \left(\prod_{i=1}^{p_j} z_{t-i} \prod_{i=1}^{q_j} \hat{a}_{jt-i} \right)$.

A reasonable approach is thus to estimate $\rho_{z_j|1, \dots, j-1}$ as simple correlations between \hat{a}_{jt} and \hat{y}_{jt} . In summary, the two indicators for the selection of the regressors y_{it} are given by $\rho_{z_j} = \text{Cor}(z_t \hat{y}_{jt})$ and $\rho_{z_j|1, \dots, j-1} = \text{Cor}(\hat{a}_{jt} \hat{y}_{jt})$ for $j = 2, 3, \dots, n$; by adding and deleting appropriate terms the best model should be determined.

To simplify the method, in particular to reduce the number of intermediate estimations, we suggest a procedure for the system (2.3) which refers to the bi-correlations functions $\rho_{zy}(i, j) \propto E(z_t y_{t-i, j})$ where $y_{t-i, j} = (z_{t-i} z_{t-j}), (z_{t-i} a_{t-j}), (a_{t-i} a_{t-j})$. First of all it is necessary to derive the sampling distribution of their estimators.

Proposition 2. Let $\{z_t\}$ be a non-gaussian process, asymptotically independent, stationary up to moment of order 6, and $\{y_t\}$ defined as above. Then, under the null hypothesis $H_0 : z_t \equiv a_t \sim \text{IID}(0, \sigma^2)$ and for N sufficiently large, we have

$$r_{zy}(i, j) = \frac{\sum_{t=\max(i, j)+1}^N (z_t - \bar{z})(y_{t-i, j} - \bar{y})}{\hat{\sigma}_z \hat{\sigma}_y [N - \max(i, j)]} \underset{H_0}{\overset{L}{\approx}} \text{IN}\left(0, \frac{1}{N - \max(i, j)}\right). \quad (4.2)$$

Proof. A similar result is stated in Hinich and Patterson (1985) and its heuristic demonstration is given in Grillenzoni (1993). We now present the various steps of the identification algorithm of the quadratic model (2.3).

Step 1. Identify (p, q) (the linear part of the model) with standard methods such as analysis of sample autocorrelations $r_{zz}(k)$ and partial autocorrelations.

Step 2. Identify (P, Q) by setting coefficients α_{ij} in the same position (i, j) as every significant correlation (3.2), i.e. $|r_{zy}(i, j)| > 2/\sqrt{N - \max(i, j)}$, in which $y_{t-i, j} = (z_{t-i} z_{t-j})$.

Step 3. Fit the partial model $z_t = \sum_i^p \phi_i z_{t-i} + \dots + \sum_i^P \sum_j^Q \alpha_{ij} z_{t-i} z_{t-j} + \tilde{a}_t$ and generate the corresponding residuals $\{\tilde{a}_t\}$; then identify (r, s) (i.e. the significant β_{ij} coefficients) as in Step 2 by setting $\tilde{y}_{t-i, j} = (z_{t-i} \tilde{a}_{t-j})$.

Step 4. Fit the partial model $z_t = \sum_i^p \phi_i z_{t-i} + \dots + \sum_i^r \sum_j^s \beta_{ij} z_{t-i} \tilde{a}_{t-j} + \hat{a}_t$ and generate the residuals $\{\hat{a}_t\}$; then identify (R, S) as in Steps 2, 3 by setting $\hat{y}_{t-i, j} = (\hat{a}_{t-i} \hat{a}_{t-j})$.

Step 5. Fit the global model using as initial values the estimates of Step 3 and for δ_{ij} the correlations $r_{zy}(i, j)$ of Step 4. Then drop all the insignificant coefficients.

Step 6. Estimate the final model and check its adequacy with residual correlations $r_{aa}(k), r_{aa^2}(k), r_{a^2a^2}(k)$. The corresponding *portmanteau* test is given by

$$Q(3K) = \sum_{k=1}^K (n-k) [r_{aa}(k) + r_{aa^2}(k) + r_{a^2a^2}(k)] \underset{H_0}{\overset{L}{\approx}} \chi^2(3K - n). \quad (4.3)$$

As we have stated previously, the rationale of the algorithm is based on the heuristics that the variables $\{y_{jt}\}$ tend to have an autonomous statistical behaviour with respect to $\{z_t\}$, and that a necessary condition for $(\alpha_{ij}, \beta_{ij}, \delta_{ij}) \neq 0$ is given by $E[z_t y_{t-i, j}] \neq 0$. These features may be realistic, in particular for models which have an irregular (subset) structure. Under these assumptions, Steps 1-4 probably lead to a moderate overparameterization; however, in Step 5 all the unnecessary coefficients are identified and deleted. Since the role of monomials

$\left(\prod_j^p z_{t-j} \right), \left(\prod_j^q a_{t-j} \right)$ may be competitive, the selection of coefficients α_{ij} before δ_{ij} implicitly accords priority to the autoregressive terms $(z_{t-i} z_{t-j})$. The identification algorithm then leads to a saving of nonlinear terms $(a_{t-i} a_{t-j})$ which complicate the estimation problem and have a limited forecasting horizon. Finally, the rationale of the residual test at Step 6 is similar to that of the test for nonlinear ARMA models discussed by McLeod and Li (1983). While the reason for considering only *diagonal* coefficients $r(i=j)$ is merely practical, the utilization of composite statistic (4.3) is relatively new. Its approximate distribution is a direct consequence of Proposition 2.

5. APPLICATION

We now conclude the paper with a numerical application that deals with a real data-set. Principal goal is to check the model building strategy discussed in sections 3 and 4, but also to compare the statistical performance of various nonlinear models.

The application concerns the ISTAT index of wholesale prices of industrial goods in the period Jan. 1973-Dec. 1985 ($N = 156$), which was crucial for price inflation in Italy. The original series $\{Z_t\}$ is not very interesting since it exhibits a marked linear trend representable by a random walk plus drift: $Z_t = \mu + Z_{t-1} + z_t$. The transformed series $\{z_t\}$ (reported in figure 1a), has been already modelled by Grillenzoni (1990) in terms of linear models with time-varying parameters, and by Grillenzoni (1991) in the context of multivariable transfer functions. In this section we refer to $\{z_t\}$ for detecting and modeling non-linearity in the variables. The first diagnostic step is taken by the non-parametric estimation of the density $f(z)$, $-\infty < z < +\infty$ and of the regression function $E[z_t | z_{t-1}]$. These are

obtained from smoothers of the type $\hat{f}(z) = (Nb)^{-1} \sum_{t=1}^N K[(z - z_t)/b]$, choosing the window width $b = .2$ and the kernel $K = N(0, 1)$ (see figures 2a and 2b).

Detailed evidence of non-gaussianity and non-linearity follows from the "bursts" of variability of the series $\{z_t\}$ (which are typical of bilinear time-series), the bimodality and asymmetry of $\hat{f}(z)$ (which are typical of threshold AR processes, see Tong, 1990, p. 157) and the fluctuation of $E[z_t | z_{t-1}]$ in correspondence of large values of the series. In order to refine these guesses we now apply the model building strategy discussed in the previous section.

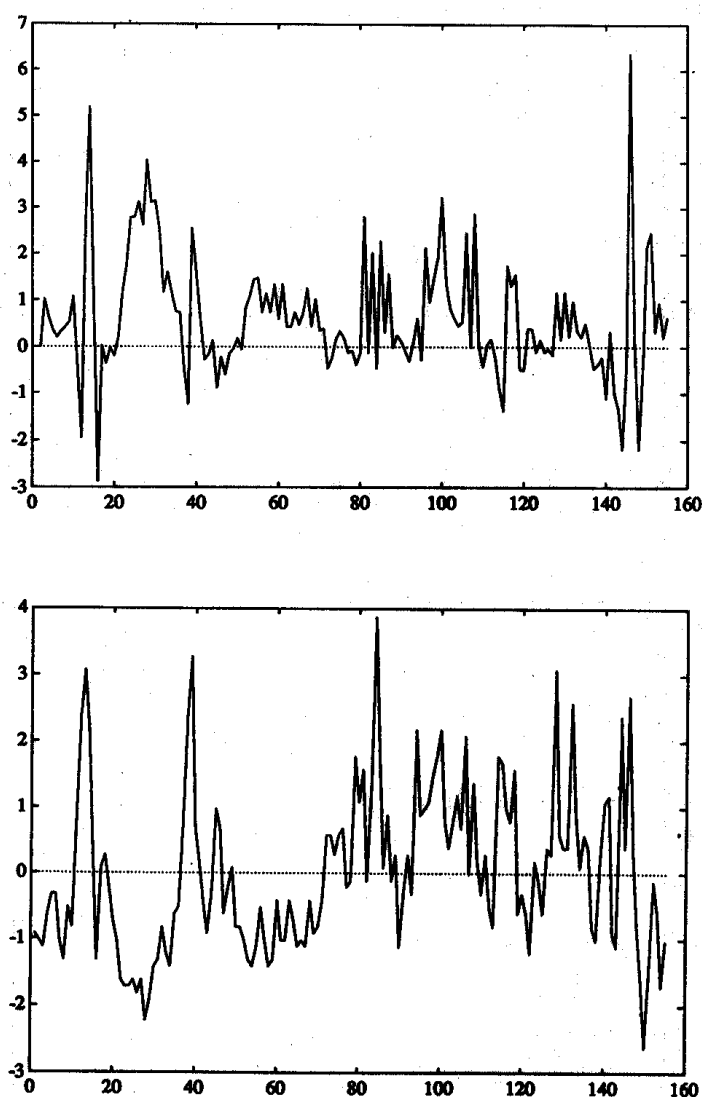


Figure 1(a,b) - Plot of original series $z_t = (Z_t - Z_{t-1}) - \bar{z}$ (a), and artificial series $y_{t-1,3} = z_{t-1} z_{t-3}$ (b).

Step 1. The first step is taken by the identification of the linear ARMA part of model (2.3); from the inspection of the sample autocorrelation functions, we got

$$z_t = .01 + .62 z_{t-1} + a_t, \quad \sigma_a^2 = .93, \quad R^2 = .38$$

(.10) (.9.8)

where the constant was included for subsequent work and t -statistics are in parentheses.

Step 2. In order to identify the significant quadratic components $y_{t-i,j} = (z_{t-i} z_{t-j})$ bi-auto correlations $r_{zy}(i, j)$ of table 1 were considered. Setting coefficients α_{ij} in the same position (i, j) as every significant correlation, we

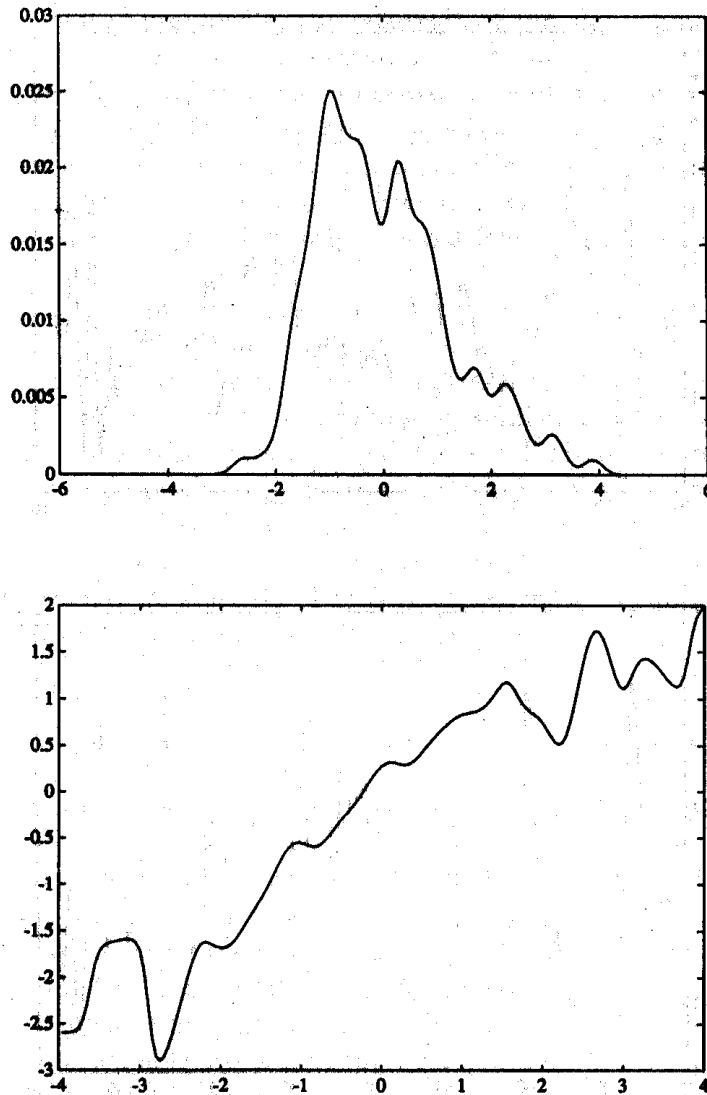


Figure 2(a,b) - Graphs of non-parametric estimates of the density $f(z)$ (a), and of the regression function $E[z_t | z_{t-1}]$ (b).

obtained the intermediate model (2.3) with $(\beta_0, \phi_1, \alpha_{13}, \alpha_{23}, \alpha_{14}, \alpha_{17}, \alpha_{18})$. However, in the subsequent OLS estimation many parameters turned out insignificant so that the resulting partial model was

$$z_t = .14 + .62 z_{t-1} - .21 z_{t-1} z_{t-3} + \tilde{a}_t, \quad \tilde{\sigma}^2 = .87, \quad R^2 = .43$$

(1.6) (10.) (-3.6)

Notice that the above includes the transfer function $z_t = [-\alpha_{13}/(1 - \phi_1 B)] z_{t-1} z_{t-3}$ which covers the sequence of correlations $r_{z_t}(1+k, 3+k)$, $k = 0, 1, \dots, 4$ of table 1.

TABLE 1
Sample bi-autocorrelations $r[z_t, (z_{t-i}, z_{t-j})]$

| $j \downarrow, i \rightarrow$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-------------------------------|------|------|------|------|------|------|------|------|------|------|------|-----|
| 1 | .13 | | | | | | | | | | | |
| 2 | -.04 | .00 | | | | | | | | | | |
| 3 | -.22 | -.21 | -.08 | | | | | | | | | |
| 4 | -.16 | -.24 | -.16 | -.12 | | | | | | | | |
| 5 | -.13 | -.17 | -.22 | -.20 | -.12 | | | | | | | |
| 6 | -.15 | -.10 | -.14 | -.19 | -.13 | -.11 | | | | | | |
| 7 | -.18 | -.14 | -.08 | -.10 | -.11 | -.07 | -.11 | | | | | |
| 8 | -.23 | -.17 | -.01 | .01 | .02 | .02 | -.10 | -.09 | | | | |
| 9 | -.10 | -.12 | -.04 | .00 | -.01 | -.05 | -.01 | -.10 | -.11 | | | |
| 10 | -.07 | -.03 | -.05 | -.03 | .03 | -.10 | .00 | -.02 | -.07 | -.01 | | |
| 11 | -.04 | .00 | .05 | -.04 | .00 | -.03 | -.01 | .11 | .05 | .01 | -.05 | |
| 12 | -.06 | .02 | .15 | .13 | .06 | -.06 | -.04 | .12 | .16 | .18 | .03 | .03 |

Step 3. Identification of bilinear terms $(z_{t-i} a_{t-j})$ requires the inspection of partial bi-correlations $r_{z\bar{y}}(i, j)$, with $\bar{y}_{t-i, j} = (z_{t-i} \hat{a}_{t-j})$. Since the significant coefficients were in correspondence of the lags $(i, j) = (1, 1), (4, 11), (10, 5)$, following the procedure of Step 2 we selected the parameters $(\beta_{1,1}, \beta_{4,11}, \beta_{10,5})$. The intermediate estimation gave

$$z_t = \begin{matrix} .25 \\ (3.2) \end{matrix} + \begin{matrix} .70 \\ (10.7) \end{matrix} z_{t-1} - \begin{matrix} .22 \\ (-4.7) \end{matrix} z_{t-1} z_{t-3} - \begin{matrix} .12 \\ (-2.5) \end{matrix} z_{t-1} \hat{a}_{t-1} - \begin{matrix} .14 \\ (-2.0) \end{matrix} z_{t-4} \hat{a}_{t-11} + \begin{matrix} .10 \\ (1.9) \end{matrix} z_{t-10} \hat{a}_{t-5} + \hat{a}_t, \quad \hat{\sigma}^2 = .82, \quad R^2 = .47.$$

Steps 4, 5. For the identification of the quadratic terms $(a_{t-i} a_{t-j})$ the partial bicorrelations $r_{z\bar{y}}(i, j)$, with $\bar{y}_{t-i, j} = (\hat{a}_{t-i} \hat{a}_{t-j})$, were considered. The sole significant coefficients were in correspondence of $(i, j) = (1, 1), (2, 2)$ and this leads to the selection of $(\delta_{11}, \delta_{22})$. However, the term $\delta_{11} a_{t-1} a_{t-1}$ is competitive with $\beta_{11} z_{t-1} a_{t-1}$, included at Step 3, and $\delta_{22} a_{t-2}^2$ is implied by the transfer function $z_t = [\delta_{11}/(1 - \phi_1 B)] a_{t-1}^2$. Estimation of the global model with $(\beta_0, \phi_1, \alpha_{13}, \beta_{11}, \beta_{4,11}, \beta_{10,5}, \delta_{11}, \delta_{22})$ confirmed the insignificance of $(\beta_{11}, \delta_{22})$.

Step 6. The nonlinear estimation of the final multilinear model provided

$$z_t = \begin{matrix} .30 \\ (3.7) \end{matrix} + \begin{matrix} .67 \\ (13.4) \end{matrix} z_{t-1} - \begin{matrix} .24 \\ (-5.4) \end{matrix} z_{t-1} z_{t-3} - \begin{matrix} .16 \\ (-2.4) \end{matrix} z_{t-4} \hat{a}_{t-11} + \begin{matrix} .12 \\ (2.2) \end{matrix} z_{t-10} \hat{a}_{t-5} - \begin{matrix} .15 \\ (-2.6) \end{matrix} \hat{a}_{t-1}^2 \hat{a}_t, \quad \hat{\sigma}^2 = .80, \quad R^2 = .49$$

with a value of the portmanteau statistic (4.3) $Q(3 \cdot 12) = 25.3 < 43.8 = \chi_{.95}(30)$. Evaluation based on the statistics $(\hat{\sigma}^2, R^2)$ and the significance of the coefficients, may not be sufficient to check the robustness of the modeling procedure. It is necessary, in fact, to use indicators which take into account the tradeoff between statistical fitting and parametric efficiency. Table 2 reports

the values of three information criteria calculated at previous steps; in two cases the model that is indicated as the best is the final one.

TABLE 2
Information Criteria evaluated at Steps 1-6

| Step, (N - d) | 1, (154) | 2, (152) | 3, (145) | 6, (145) |
|------------------------------|----------|----------|----------|----------|
| AIC : $f(N) = 2$ | -.047 | -.100 | -.116 | -.140 |
| BIC : $f(N) = \log(N)$ | -.007 | -.040 | +.007 | -.017 |
| HIC : $f(N) = \log(\log(N))$ | -.031 | -.076 | -.066 | -.091 |

The parameters of the model identified at Step 6 were successively estimated with the recursive algorithm (3.6) in order to check their stability over time. The algorithm was initialized with $\hat{\beta}(0) = \hat{\beta}_N$, $\mathbf{R}(0) = \mathbf{R}_N$ obtained from the iterative estimation. The optimal forgetting factor $\hat{\lambda}_N = .974$ was obtained with a search procedure by minimizing the global objective function $S_N(\lambda) = \sum_{t=d}^N [\hat{a}(t)^2 + \|\hat{\beta}(t) - \hat{\beta}_N\|^2]$, which establishes a tradeoff between tracking and accuracy. Trajectories of estimates $\hat{\beta}_i(t)$, $i = 1, \dots, 6$ are shown in figure 3.

Notice that despite the mild value taken by λ a significant variability of parameters occurs. This is a clear indication that the process $\{z_t\}$ is nonstationarity in higher order moments. Other important features regard the fact the on-line estimates move around their off-line value $\hat{\beta}_N$ and their fluctuations are asymmetric, *i.e.* compensate each other. This confirms, in a certain sense, the role played by the variability of the parameters to stabilize the behaviour of the output. Finally, the gain in statistical fitting allowed by recursive estimation is summarised by the residual variance: $(N-d)^{-1} \sum_{t=d}^N \hat{a}(t)^2 = .63$.

To complete the exercise we now evaluate the performance of neural network (NN) and exponential autoregressive (EAR) models on the same data-set. The framework of neural networks (see White, 1989) is based on a sequence of hidden units having common input variables $\{x_t\}$ and the same structure $\psi(\cdot)$. The responses of the units interact at an intermediate layer, before reaching the

output $y_t = \Psi \left[\sum_{j=1}^q \beta_j \psi(x_t' \alpha_j) \right] + a_t$. Accordingly, the shape of $\psi_j(\cdot)$ is represented

by a threshold rule or a sigmoid function (*e.g.* a probability distribution). This framework has important connections with the *projection pursuit regression* designed for non-parametric curve fittings (see Friedman and Stuetzle, 1981), in which $\psi_j(\cdot)$ are smooth functions with different structure and $(\Psi = \beta_j) \equiv 1$. Similarities with exponential time series models of Ozaki (1985) can be established by taking $\psi(\cdot)$ as the logistic function $[1 + \exp(\cdot)^{-1}]^{-1}$, $\Psi \equiv 1$ and letting $y_t = z_t$, $x_t = z_{t-1}$:

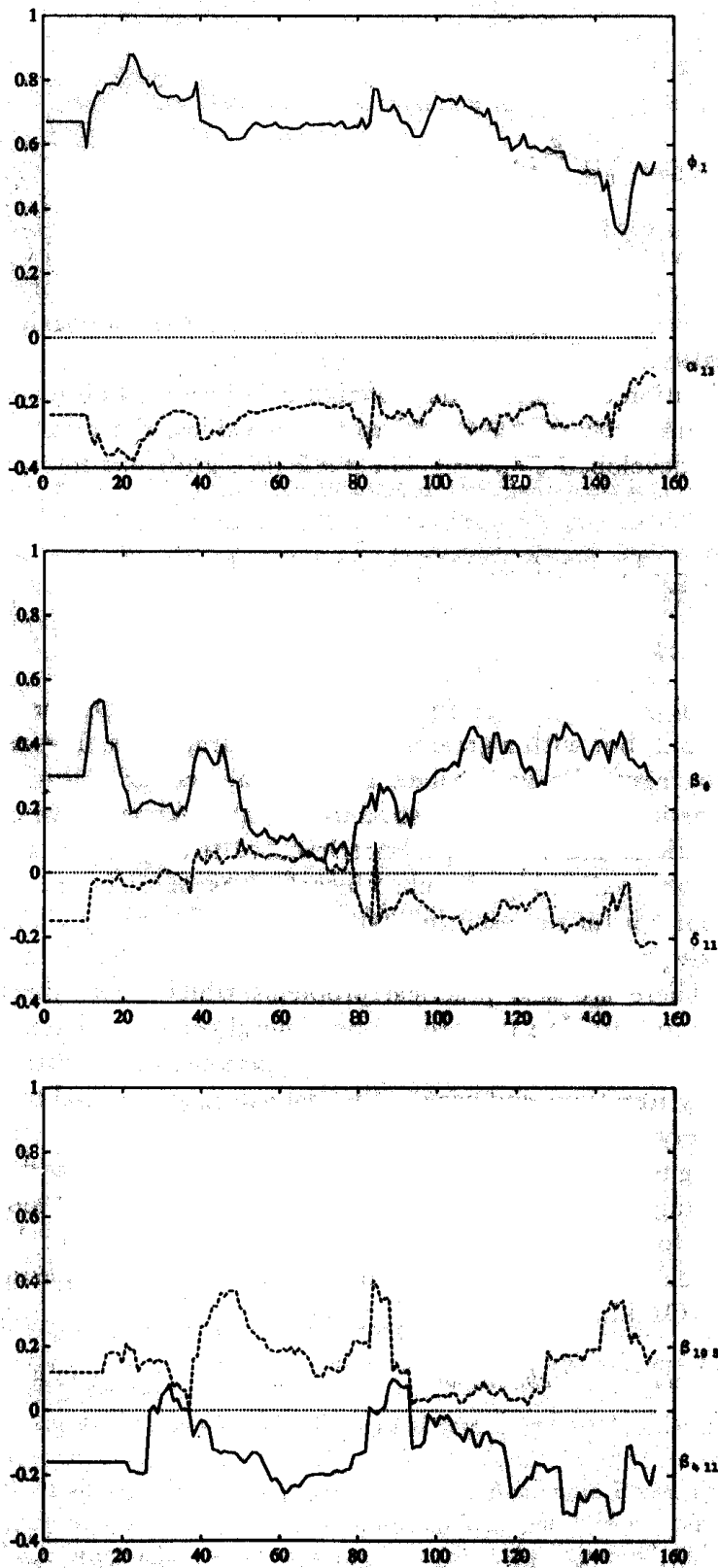


Figure 3(a,b,c) - Recursive estimates of the parameters of the model at Step 6.

$$\text{N-N: } z_t = \beta_0 + \sum_{j=1}^q \beta_j [1 + \exp(-\alpha_j z_{t-1})]^{-1} + a_t \quad (5.1a)$$

$$\text{EAR: } z_t = \beta_0 + \sum_{j=1}^q [\phi_j + \beta_j \exp(-\alpha_j z_{t-1}^2)] z_{t-j} + a_t. \quad (5.1b)$$

Network (5.1a) was applied by White (1989) to fit the Hénon map $y_t = 1 - 1.4 y_{t-1}^2 + .3 y_{t-2}$ with $\mathbf{x}_t' = [1, y_{t-1}^2, y_{t-2}]$ and $q = 5$. The number of parameters involved (21) was excessive in view of the fact a better approximation is given by a quadratic AR(2) model.

The application of (5.1a) to data of figure 1a provided disappointing results. Given the sensitivity of the model to initial values, the identification of the order q was developed stepwise. For $q > 2$ no significant improvement of $\hat{\sigma}_a^2$ was achieved

$$z_t = \underset{(1.3)}{.27} - \underset{(3.2)}{2.1} \left[1 + \exp \left(\underset{(1.1)}{.41} + \underset{(2.1)}{1.9} z_{t-1} \right) \right]^{-1} + \underset{(1.7)}{.87} \left[1 + \exp \left(\underset{(-3.4)}{-43.6} - \underset{(-2.8)}{29.1} z_{t-1} \right) \right]^{-1},$$

$$\sigma^2 = .88, \quad R^2 = .44$$

The statistical performance is equivalent to that of the intermediate nonlinear model at Step 2, but with a greater number of parameters. The reasons for this disappointing result lie in the fact that neural networks, such as projection pursuit regression, have been designed for the approximation of complex deterministic functions and for nonparametric curve fitting. They are effective, in general, for approximating the chaotic solutions of difference - differential equations, but not for representing data generated by stochastic processes. This is particularly true for time series that exhibit nonlinear and nonstationary behaviour. There are also practical problems which make the utilization of these models unsuitable: I) there are no general rules for choosing the shape of $\Psi(\cdot)$, $\psi(\cdot)$ and representations are not parsimonious; II) the estimation process requires many iterations and accurate initial values; III) optimal forecasting algorithms are very complicate.

Data of figure 1a were also fitted with the exponential model (5.1b), with $q = 2$, providing a residual variance equal to that of the AR(1) model at Step 1. The reasons for this disappointing behaviour are similar to those of neural networks since this class of models has been designed for particular physical systems (see Ozaki, 1985).

Istituto Universitario di Architettura
Venezia

CARLO GRILLENZONI

REFERENCES

- M. BHASKARA-RAO, T. SUBBA-RAO, A.M. WALKER (1983), *On the existence of some bilinear time series models*, "Journal of Time Series Analysis", 4, pp. 93-110.

- J.M. FRIEDMAN, W. STUETZLE (1981), *Projection pursuit regression*, "Journal of the American Statistical Association", 76, pp. 817-824.
- C.W.J. GRANGER, A.P. ANDERSEN (1978), *An introduction to bilinear time series models*, Vandenhoeck and Ruprecht, Göttingen.
- C. GRILLENZONI (1990), *Modeling time-varying dynamical systems*, "Journal of the American Statistical Association", 85, pp. 499-507.
- C. GRILLENZONI (1991), *Simultaneous transfer functions versus vector ARMA models*, "Journal of Forecasting", 10, pp. 477-499.
- C. GRILLENZONI (1993), *Multilinear models for nonlinear time series*, "E.U.I. Working Papers in Economics", 93/2, European University, Florence.
- E.J. HANNAN, A.J. MCDUGALL (1988), *Regression procedures for ARMA estimation*, "Journal of the American Statistical Association", 83, pp. 490-498.
- M.J. HINICH, D.M. PATTERSON (1985a), *Evidence of nonlinearity in daily stock returns*, "Journal of Business and Economic Statistics", 3, pp. 69-77.
- M.J. HINICH, D.M. PATTERSON (1985b), *Identification of the coefficients in a non-linear time series of the quadratic type*, "Journal of Econometrics", 30, pp. 269-288.
- M. LAHA, R. ROHATGI (1979), *An introduction to probability theory*, Wiley, New York.
- J. LIU, P.J. BROCKWELL (1988), *On the general bilinear time series model*, "Journal of Applied Probability", 25, pp. 553-564.
- D.M. KEENAN (1985), *A Tukey non-additivity-type test for nonlinearity*, "Biometrika", 72, pp. 39-44.
- K. KUMAR (1986), *On the identification of some bilinear time series models*, "Journal of Time Series Analysis", 7, pp. 117-122.
- A.I. MCLEOD, W.K. LI (1983), *Diagnostic checking ARMA time series models using squared-residuals autocorrelations*, "Journal of Time Series Analysis", 4, pp. 269-273.
- T. OZAKI (1985), *Nonlinear time series models and dynamical systems*, in E.J. Hannan and P.R. Krishnaiah (eds.), "Handbook of Statistics", vol. 5, North-Holland, Amsterdam.
- T.D. PHAM (1985), *Bilinear markovian representation and bilinear models*, "Stochastic Processes and Applications", 20, pp. 295-306.
- M.B. PRIESTLEY (1988), *Nonlinear and nonstationary time series analysis*, Academic Press, London.
- M.M. RAO (1961), *Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations*, "Annals of Mathematical Statistics", 32, pp. 195-218.
- W.J. RUGH (1981), *Nonlinear system theory*, J. Hopkins University Press, Baltimore.
- S.A.O. SESAY, T. SUBBA-RAO (1988), *Yule-Walker type difference equations for higher order moments for bilinear models*, "Journal of Time Series Analysis", 9, pp. 385-401.
- V. SOLO (1978), *Time series recursions and stochastic approximation*, "Ph.D. dissertation", A.N.U. Dept. of Statistics, Canberra.
- A. SPANOS (1986), *Statistical foundations of econometric modeling*, University Press, Cambridge.
- H. STOUT (1974), *Almost sure convergence*, Academic Press, New York.
- T. SUBBA-RAO, M. GABR (1984), *An introduction to bispectral analysis and bilinear time series models*, Springer-Verlag, Berlin.
- H. TONG (1990), *Nonlinear time series: a dynamical system approach*, Clarendon Press, Oxford.
- H. WHITE (1984), *Asymptotic theory for econometricians*, Academic Press, Orlando.
- H. WHITE (1989), *Some asymptotic results for learning in single hidden-layer feedforward network models*, "Journal of the American Statistical Association", 84, pp. 1003-1013.