



EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 93/2

Multilinear Models for Nonlinear Time Series

CARLO GRILLENZONI

WEP

330

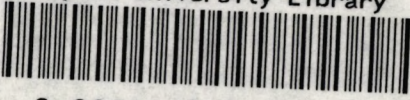
EUR

European University Institute, Florence

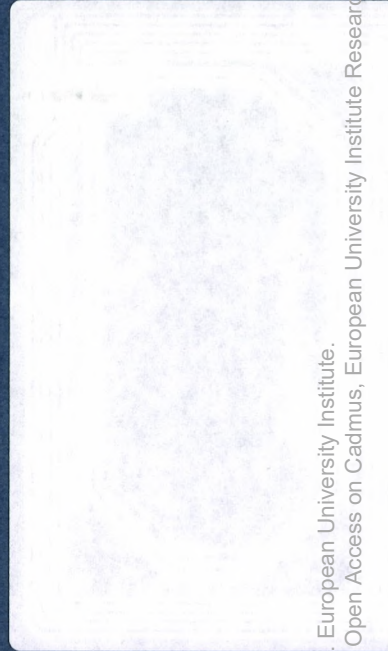
© The Author(s). European University Institute.

Digitised version produced by the EUI Library in 2020. Available Open Access on Cadmus, European University Institute Research Repository.

European University Library



3 0001 0014 9673 8



Please note

As from January 1990 the EUI Working Paper Series is divided into six sub-series, each sub-series is numbered individually (e.g. EUI Working Paper LAW No. 90/1).

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper ECO No. 93/2

Multilinear Models for Nonlinear Time Series

CARLO GRILLENZONI

BADIA FIESOLANA, SAN DOMENICO (FI)

All rights reserved.
No part of this paper may be reproduced in any form
without permission of the author.

© Carlo Grillenzoni
Printed in Italy in February 1993
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

MULTILINEAR MODELS FOR NONLINEAR TIME SERIES

Carlo Grillenzoni

*Dept. of Economics , Univ. of Modena
& European University Institute*

Abstract - A class of multilinear models for nonlinear time series is introduced. It extends the bilinear ARMA representation of Granger-Andersen by including general monomials of lagged input and output. For this class, algorithms of structure identification and parameter estimation are provided, suitable for dealing with subset models and time varying coefficients. An extended application on real economic data illustrates the framework and makes comparisons with other nonlinear models. Contents : 1. Introduction; 2. Testing; 3. Representation; 4. Identification; 5. Estimation; 6. Conclusions, Appendicies.

Key-words - Bilinear processes, Multilinear models, Multicorrelation functions, Recursive pseudolinear regression, Neural networks.

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 & Andersen (1978) and Subba-Rao & 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 & Patterson (1985 a,b) with a class of quadratic innovation models having a nonzero bivariate (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 introduces the case study, focusing on presentation of the data and tests for linearity; an extension of tests presently available is proposed. Section 3 derives a multilinear representation and discusses problems related to its stationarity and stability. In Section 4 techniques of identification mentioned above are developed and applied to the set of data introduced in Section 2. Finally, Section 5 deals with methodological problems related to nonlinear estimation in the case of time-varying parameters; it also concludes the numerical application.

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 statistical literature; their general features can be summarized as follows :

- 1) Tests are carried out utilizing only statistics of residuals generated by linear AR models, which represent the null H_0 . In fact, it seems reasonable to start the model building by applying tests that do not require the direct estimation of the nonlinear alternative H_1 .
- 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, etc.), under the alternative hypothesis.

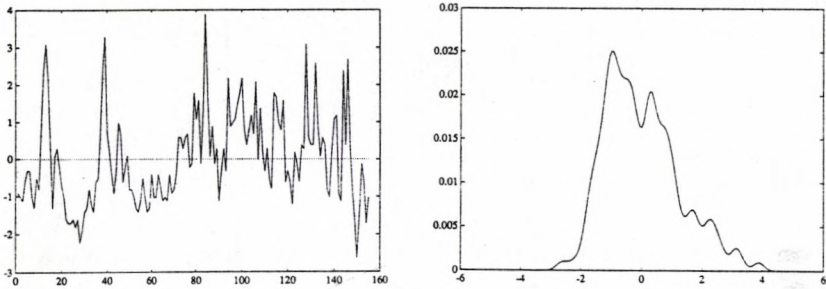
In practical terms, however, many of the tests that do not assume a specific model under H_1 implicitly refer to quadratic systems, i.e. to systems that admit at most a second order Volterra expansion (see Keenan, 1985). On the other hand, tests that are designed for a particular H_1 , may locally have good power in testing for other kinds of alternatives (see Lukkonen *et al.*, 1988). Given this situation of indeterminacy, a reasonable strategy for testing and specifying nonlinear time series consists in comparing several test statistics, with the sole constraint represented by point (1) above. To make these ideas more precise and to motivate further methodological developments, we introduce in advance the numerical application on which the paper focuses.

The data - The application concerns with the index of wholesale prices in the period Jan.1973 - Dec.1985 ($N=156$), which was crucial for price inflation in Europe. The original series $\{Z_t\}$ is not very interesting since it exhibits a marked linear trend representable by a random walk plus drift : $Z_t = \mu + Z_{t-1} + z_t$, where $\{z_t\}$ is a zero mean process.

The transformed series $\{z_t\}$ 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. These studies have shown that $\{z_t\}$ can be significantly explained by the series of exchange rates $\pounds/\$$, while the reverse causality (in the Wiener-Granger sense) is completely absent. This evidence supports the viewpoint of Keynesian economists, as opposed to monetarist theories, about the real determinants of inflation. The implications for economic policy are obvious: exchange rates must be controlled by central banks in order to avoid volatility of prices.

In this paper we refer to the series $\{z_t\}$ for detecting and modeling nonlinearity in the variables. First step is taken by evaluation of descriptive statistics, such as skewness $\mathcal{S} = .679$ and kurtosis $\mathcal{K} = 3.124$; the combination of these in the formula $N[\mathcal{S}^2 + (\mathcal{K}^2 - 3)/4]/6 = 11.95$ is asymptotically distributed as a $\chi^2(2)$ under the hypothesis of gaussianity (see Jarque & Bera, 1980). Detailed evidence of non-normality is also provided by the non-parametric density estimate $\hat{f}(z) = (Nh)^{-1} \sum_{i=1}^N K[(z - z_i)/h]$, obtained with a window width $h = .2$ and the kernel $K = N(0, 1)$ (see Figure 1b).

Figure 1 (a,b) - Plot of series $z_t = (1-B)Z_t - \mu$ and non-parametric density estimate .



Plot of series $\{z_t\}$ in Figure 1a shows a number of "bursts" of variability which are typical of bilinear time series models (Priestley, 1988). Figure 1b confirms the existence of bilinearity in terms of a marked asymmetry of $\hat{f}(z)$; however, there is also a significant bimodality at $z = 0$ which is typical of threshold AR processes (see Tong, 1990 p.157). Tentative nonlinear models for $\{z_t\}$ are then given by

$$\text{BAR} : z_t = \phi z_{t-1} + \beta z_{t-1} a_{t-1} + \sigma a_t, \quad a_t \sim \text{IID}(0, 1)$$

$$\text{TAR} : z_t = \phi_i z_{t-1} + \sigma_i a_t, \quad (\phi_i, \sigma_i) = \begin{cases} (\phi_1, \sigma_1), & z_{t-1} \leq 0 \\ (\phi_2, \sigma_2), & z_{t-1} > 0 \end{cases}$$

In order to refine these guesses we go back to the introduction of the section and apply the linearity tests described in the literature. As stated at point (1), these simply require the estimation of the model specified under H_0 ; in particular, its residuals

$$H_0 : z_t = \underset{(.10)}{.01} + \underset{(.98)}{.62} z_{t-1} + a_t, \quad \sigma_a^2 = .93, R^2 = .38 \quad (2.1)$$

Table 1 - Sample autocorrelations and biconrelations of residuals $\{a_t\}$.

Lag	1	2	3	4	5	6	7	8	9	10	11	12	
$r(a, a)$	-.03	.00	.00	.02	.07	.02	.01	.10	-.06	-.07	-.05	.13	Q(24)=16
$r(a, a^2)$.06	.13	-.04	.12	.04	-.01	-.07	-.08	-.07	.09	-.09	.08	
$r(a^2, a^2)$	-.13	.07	-.14	.01	-.09	-.10	-.01	-.03	-.06	-.01	-.06	.12	Q(24)=21

Some Tests - Table 1 reports autocorrelations and cross-correlations of squared residuals. The utility of these coefficients in detecting bilinearity and heteroscedasticity (of ARCH type) have been discussed by Maravall (1983), Li (1984), Kumar (1986) and Gabr (1988). A portmanteau test of Ljung-Box type was proposed by McLeod & Li (1983) and refers to the statistics $Q(k) \approx N \sum_{j=1}^k r_{a^2, a^2}^2(j)$. It can be shown that asymptotically $Q(k) \approx N R_k^2$ — where R_k^2 is the coefficient of determination from the auxiliary regression $\hat{a}_t^2 = \sigma^2 + \sum_{i=1}^k \alpha_i \hat{a}_{t-i}^2 + e_t$ — therefore its meaning as a test for ARCH is apparent.

Table 2 - Results of typical linearity tests applied to the series $\{z_t\}$.

Type	H_1	Value	Critical 5%	Authors
Portmanteau	Bilin.- ARCH	11.8	$\chi^2(12) = 21.0$	McLeod & Li (1983)
Lagr. Multipliers	Bilinear(2,2)	6.3	$\chi^2(4) = 9.5$	Weiss (1986)
Lagr. Multipliers	Exponent.(2)	5.1	$\chi^2(2) = 6.0$	Lukkonen <i>et al.</i> (1988)
Tukey 1 d.f.	Quadratic	7.8	$F(1, 150) = 3.9$	Keenan (1985)
Tukey - CUSUM	"Threshold"	5.1	$F(1, 140) = 3.9$	Tsay (1989)
Likelihood Ratio	Threshold	7.2	" $\chi^2(2) = 6.0$ "	Tong (1990)

Note: Critical values refer to asymptotic distributions under H_0 .

Lagrange multiplier tests have been discussed by Weiss (1986) and Lukkonen, Saikkonen & Terasvirta (1988), and refer to the quadratic statistic

$$LM = \hat{\sigma}_a^2 \left(\sum_{i=1}^N \hat{z}_{1i} \hat{a}_i \right)' \left(\hat{M}_{11} - \hat{M}_{10} \hat{M}_{00}^{-1} \hat{M}_{01}' \right)^{-1} \left(\sum_{i=1}^N \hat{z}_{1i} \hat{a}_i \right), \quad \hat{M}_{ij} = \left(\sum_{i=1}^N \hat{z}_{1i} \hat{z}_{ji}' \right)_{i,j=0,1}$$

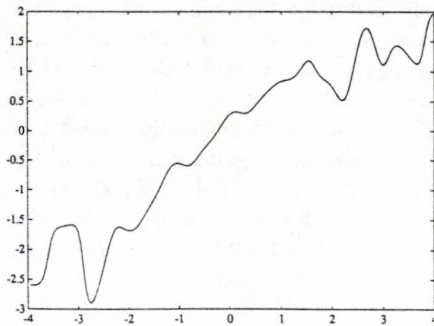
where $\hat{z}_{\alpha} = [1, z_{t-1}, \dots, z_{t-p}]'$ are the regressors of the linear model. The structure of \hat{z}_{1i} depends on the alternative hypothesis; for example, in a bilinear model of order (2,2) it becomes $\hat{z}_{1i} = [z_{t-1} \hat{a}_{1-1}, \dots, z_{t-2} \hat{a}_{1-2}]'$, while for a second order exponential AR we have $\hat{z}_{1i} = [z_{t-1}^3, z_{t-1}^2 z_{t-2}]'$. The choice of the orders under H_1 must be carefully made since they determine the power of the test. As before, it can be shown that the LM statistic is asymptotically equivalent to NR^2 , concerning the regression of \hat{a}_i on \hat{z}_{1i} .

The test proposed by Keenan (1985) assumes under H_1 only a nonlinear model which admits a second order Volterra expansion. The procedure is similar to Tukey's one-degree-of-freedom test for nonadditivity and involves the auxiliary regressions of $\hat{z}_t^2 = (z_t - \hat{a}_t)^2$ on \hat{z}_{α} (with residuals $\hat{\epsilon}_t$) and of \hat{a}_t on $\hat{\epsilon}_t$. Tsay (1989) has developed a method which combines Keenan's test with the CUSUM test by Petrucci & Davies (1986) for threshold autoregressions. The basic steps consist of getting standardized recursive innovations $\hat{a}_{(t)}$ from the linear model with arranged (sorted) observations $z_{(t)}$, and next in regressing $\hat{a}_{(t)}$ on $z_{(0)}$.

The empirical results reported in Tables 1,2 show that the sole statistic significant at 1% is the Keenan test; Tsay's test is significant at 5% and so is the LR test by Tong. This last needs a consideration apart since it requires the direct estimation of the model under H_1 ; moreover, since the likelihood function is not differentiable with respect to the threshold parameter (in the example it was set $z_{t-1} = 0$), standard asymptotic theory does not hold. The value of $-2 \log(LR)$ reported in Table 2 is therefore purely indicative.

By interpreting Tsay's test as a test on the nonstationarity of a nonlinear process — recall that it combines Keenan and CUSUM methods — a tentative representation for the series $\{z_t\}$ is then given by the evolving quadratic model $(z_t - \phi_t z_{t-1}) = \alpha z_{t-1} z_{t-j} + a_t$. This conclusion is supported by the work of Grillenzoni (1990), in which it was shown that ϕ_t follows a cubic polynomial of time, and by the fact that the diagonal cumulant function $\hat{\mu}_3(i+k, j+k) = N^{-1} \sum_{t=1}^N (z_t z_{t-i-k} z_{t-j-k})$ is nonzero for $(i, j) = (1, 3)$ and $k = 0, \dots, 4$. This feature suggests an extension of the test procedures discussed so far.

Figure 2 (a,b) - Non-parametric estimate of the regression function $E[z_t | z_{t-1}]$ with $h = 0.2$.



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 $a_t \sim \text{IID}(0, \sigma^2)$ through a linear filter $\psi(B) = \sum_{k=0}^{\infty} \psi_k B^k$. Now assuming $\mu_3^a = E(a_t^3) \neq 0$, the typical situation is that $z_t = \psi(B)a_t$ have a third order cumulant (bicovariance) function which is uniformly non-zero, namely

$$\mu_3^z(i, j) = E[(z_t - \mu_z)(z_{t-i} - \mu_z)(z_{t-j} - \mu_z)] = \mu_3^a \sum_{k=0}^{\infty} \psi_k \psi_{k+i} \psi_{k+j} \neq 0$$

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 $\hat{a}_t = [z_t - E(z_t | z_{t-1}, z_{t-2}, \dots)]$, by checking if $\mu_3^z(i, j) = 0$ for all $(i, j) \neq 0$.

More generally, a global linearity test may refer to the multicorrelation function

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

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

$$\Delta_{k+1} = \begin{cases} \mu_{k+1}^a(0, \dots, 0) = E(\hat{a}_t^{k+1}) \\ \{E(\hat{a}_t^2)E[A_t - E(A_t)]^2\}^{1/2}, & A_t = \prod_{j=1}^k \hat{a}_{t-i_j} \\ [E(\hat{a}_t^2)]^{(k+1)/2} \end{cases}$$

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

The meaning of the above framework 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 $\hat{A}_t = \prod_{j=1}^k \hat{a}_{t-i_j}$, $\bar{A}_N = N^{-1} \sum_{t=1}^N \hat{A}_t$, and finally

$$r_{k+1}(i_1, \dots, i_k) = \frac{\sum_t \hat{a}_t \hat{A}_t}{[(\sum_t \hat{a}_t^2) \sum_t (\hat{A}_t - \bar{A}_N)^2]^{1/2}} \stackrel{L}{\approx} \frac{1}{h_0} N \left(0, \frac{1}{[N - \max(i_j)]^{1/2}} \right)$$

The approximate distribution of $r_k(\cdot)$ holds under the null $H_0: z_t = a_t \sim \text{IID}$, with $E(a_t^{k+1}) < \infty$, and for N sufficiently large; the argument of the proof being similar to those of standard time series analysis — see McLeod & Li (1983) and Section 4.

Application of the above test to the the residual of model (2.1) has provided several indications of dependence that contradict the results of Table 1. Referring to bi-tri correlation functions, 5% significance level and $k=20$ as the maximum lag allowed, the number of significant coefficients turned out to be greater than 100. Non-trivial examples were $r_3(6, 8) = .25$, $r_3(4, 6) = -.22$, $r_3(3, 3) = -.23$ and $r_4(6, 6, 8) = .31$, $r_4(14, 14, 20) = .23$, $r_4(5, 5, 12) = .25$, $r_4(12, 12, 16) = -.28$, $r_4(1, 3, 5) = .26$, amongst many others. This complexity requires that more general representations for non-linear time series be sought.

3. 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 realized by considering a general nonlinear function of the vector of "regressors" $\underline{x}_t' = [z_{t-1}, \dots, a_{t-q}] = \{x_{it}\}$ (see Priestley, 1988 p.92)

$$\text{NARMA} \quad z_t = f(z_{t-1}, \dots, z_{t-p}, a_{t-1}, \dots, a_{t-q}) + a_t, \quad a_t \sim \text{IID}(0, \sigma^2) \quad (3.1)$$

With this formulation $\{a_t\}$ play the role of innovations of the nonlinear process $\{z_t\}$ and $f(\underline{x}_t)$ that of projection on the past information: $E[z_t | \mathfrak{I}_{t-1} = (z_{t-1}, z_{t-2}, \dots)]$.

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 \underline{x}_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 $\underline{x}_t = \underline{0}$, 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 pseudo-regressors x_{it} , $i = 1, 2, \dots, (p+q) = m$. The explicit constant term $\beta_0 = f(\underline{0})$ 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 :

$$\text{MARMA} \quad z_t = \beta_0 + \sum_{j=1}^n \left[\sum_{k_1 + \dots + k_m = j} \beta_{k_1 \dots k_m} \prod_{i=1}^m (x_{it})^{k_i} \right] + a_t, \quad a_t \sim \text{IID}(0, \sigma^2) \quad (3.2)$$

$$x_{it} = \{z_{t-i}, a_{t-(p+q)}\}_{i=1, 2, \dots, p+q}, \quad \beta_{k_1 \dots k_m} = \left. \frac{\partial^j f(\underline{x}_t)}{\partial^{k_1} x_{1t} \dots \partial^{k_m} x_{mt}} \right|_{\underline{x}_t = \underline{0}}$$

the degree 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 (3.2)

also stems from ordinary series expansions, as the following example shows.

Example 1 - Consider the model $z_t = z_{t-1} \exp(-a_{t-1}) + a_t$; hence from the expansion $\exp(-a) = \sum_{k=0}^{\infty} (-1)^k a^k / k!$ we have, 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 (3.2) can also be viewed as a direct generalization of the Bilinear ARMA system, e.g. taking $n = 2$ we may get the quadratic model

$$z_t = \beta_0 + \left(\sum_{i=1}^P \phi_i z_{t-i} + \sum_{j=1}^Q \theta_j a_{t-j} + \sum_{i=1}^r \sum_{j=1}^s \beta_{ij} z_{t-i} a_{t-j} + \sum_{i=1}^P \sum_{j=1}^Q \alpha_{ij} z_{t-i} z_{t-j} + \sum_{i=1}^R \sum_{j=1}^S \delta_{ij} a_{t-i} a_{t-j} \right) + a_t \quad (3.3)$$

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 & Andersen (1978) on the basis of the fact that they may raise difficult problems of stationarity and invertibility. However, Hinich & 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; Gabr, 1988), this class always generates nonzero third order cumulants and thus it is a candidate to represent more complex processes.

Stability and Stationarity - Stability properties are suitable features for dynamic models, since they determine the reliability of the forecasting and control rules as well as the performance 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 vector Markov chain $\underline{x}_t = f(\underline{x}_{t-1}) + \underline{e}_t$ with $f(\cdot)$ analytic and $\underline{e}_t \sim \text{IID}(\underline{0}, \Sigma < \infty)$. Unfortunately, this feature does not hold for multilinear models, and therefore the Tong's method cannot be applied to equations (3.2)-(3.3).

Even restricting the treatment to bilinear ARMA models (i.e. $P = Q = R = S = 0$ in (3.3)), *compact* parametric conditions of stationarity have been established only for particular orders (p, q, r, s) . For the *superdiagonal* model $z_t = \sum_{i=1}^p \phi_i z_{t-i} + \sum_{j=1}^q \beta_{j1} z_{t-j} a_{t-1} + a_t$ with $E(a_t^2) = \sigma^2 < \infty$, that written with a vector notation becomes $z_t = \underline{\phi}' z_{t-1} + \underline{\beta}_1' z_{t-1} a_{t-1} + a_t$,

Bhaskara-Rao, Subba-Rao & Walker (1983) have obtained the following conditions

$$\text{stationarity : } \lambda_{\max}(\Phi \otimes \Phi + B \otimes B \sigma^2) < 1 \quad (3.4a)$$

$$\text{invertibility : } \beta_1' E(z_{t-1} z_{t-1}') \beta_1 < 1 \quad (3.4b)$$

where $\Phi' = [\phi : I_{p-1}]$, $B' = [\beta_1 : O_{p-1}]$ and $\lambda_{\max}(\cdot)$ denotes the spectral radius. If (3.4a) is satisfied, then the equivalent markovian representation $z_t = \Phi z_{t-1} + B z_{t-1} a_{t-1} + c a_t$, with $c' = (1 : O_{p-1})$, admits the multilinear MA solution $z_t = \sum_{k=1}^{\infty} \prod_{j=1}^k (\Phi + B a_{t-j}) c a_{t-k} + c a_t$ that converges in mean square for all t . Analogously, under (3.4b) there exists a unic function $h : \mathfrak{R}^m \rightarrow \mathfrak{R}$ such that $a_t = h(z_t, z_{t-1}, z_{t-2}, \dots)$ converges with probability one for all t .

Following this approach, Pham (1986) and Liu & Brockwell (1988) have derived stationarity constraints for a general BARMA model, which are very complicated and difficult to apply. Anyway, even though conditions (3.4) seems more transparent, their importance remain of "limited" practical value in view of the following remarks :

- i) It is not clear what they actually mean in terms of system parameters $\{\phi_i, \beta_{ij}\}$. 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 (3.4b) ?
- ii) They are not concerned with cumulants and higher order moments involved by non-linear 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 $z_t = \beta z_{t-k} a_{t-k} + a_t$ with $k > h > 0$; in this case condition (3.4b) 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 (3.4a) holds, i.e. $\beta^2 \sigma^2 < 1$, the asymptotic solution of this equation leads to the invertibility requirement $\beta^2 \cdot \sigma^2 / (1 - \beta^2 \sigma^2) < 1$, i.e. $\beta^2 \sigma^2 < .5$.

Example 3 - Consider the above example with $a_t \sim \text{IN}(0, \sigma^2)$ gaussian; having $E(z_{t-k} a_{t-k}) = 0$ 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 (3.4a).

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 & Subba-Rao, 1988). It is then clear that conditions of stationarity for complex nonlinear 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.

These comments tend to discourage the analysis of the stability properties of multilinear models (3.2)-(3.3). Granger & Andersen (1978) have heuristically shown the virtual nonstationarity and non-invertibility of the schemes $z_t = \alpha z_{t-1}^2 + a_t$, $z_t = \delta a_{t-1}^2 + a_t$. However, unless the contrary is proved, one cannot exclude that suitable properties may hold, even though locally and for particular realizations, for the quadratic model (3.3).

With respect to the nonlinear system (3.1), if $\|f(\underline{x}_t)\| < \|\underline{x}_t\|$, 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.

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

However, these models are nonlinear in the parameters and thus require complicated algorithms of identification, estimation and forecasting. The peculiar feature of representation (3.2) is given by its regression structure which allows for direct application of many recursive procedures of standard time series analysis.

Strong *realistic* considerations that may mitigate the picture of uncertainty outlined so far, are given by the following observations :

- a) Stationarity properties are certainly suitable features, but they are often concerned with abstract asymptotic behaviour of the output of the models. In modeling real data, users typically deal with finite sampling intervals; moreover, many observational time series (mostly in economics) are nearly divergent in nature.
- b) The parameters of the models may be time-varying (deterministically or stochastically). In this case, 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 a number of experiments, we have assessed, for example, that the quadratic process $z_t = \alpha z_{t-1}^2 + a_t$, with $z_0 = 0$ and a_t uniformly distributed in the interval $[-1.5, +1.5]$, tends to overflow within 1000 recursions for $|\alpha| > .30$. However, if $\{\alpha_t\}$ oscillates in the ring $.30 < |\alpha_t| < .40$ (e.g. $\alpha_1 = +.35, \alpha_2 = -.35, \alpha_3 = +.35 \dots$ etc.) the process $\{z_t\}$ may not diverge. In practice, there are infinite trajectories of $\{\alpha_t\}$, 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. This means, in practice, making statistical inference under the next working hypotheses :

Assumptions A - Consider the class of processes (3.2) 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 :

- (A1) $\{z_t\}$ is *asymptotically independent* : $\phi(m) = \sup_t |F_t(z_t | z_{t-m}) - F_t(z_t)| \rightarrow 0$ as $m \rightarrow \infty$;
- (A2) $\{z_t\}$ is *strictly bounded* : $\inf_t P(|z_t| \leq M) = 1, M \ll \infty$.

Condition (A1) is a particular version of the so-called ϕ -mixing property (see Stout, 1974); condition (A2) enables the existence of moments of every order (the so-called moment property), that is (A2) $\Rightarrow \sup_t E |z_t|^k < \infty$ for all $k < \infty$. This result follows im-

mediately by the fact that, given any real random variable Z , the existence of $E|Z|^k, k < \infty$ implies that $z^{k+\delta}P(|Z| \geq z) \rightarrow 0$ as $|z| \rightarrow \infty$ with $\delta \equiv 0$; on the other hand, the converse holds for any value $\delta > 0$ (see Laha & Rohatgi, 1979 p.38). Hence, there must be a precise relationship between extreme values that $\{z_t\}$ may assume and the mass of probability of the corresponding events.

With this in mind we can state that if only moments up to order $K < \infty$ are concerned in the analysis of the process $\{z_t\}$, then (A2) may be weakened as:

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

Instead, a requirement stronger than (A1) is represented by

$$(A1)' \quad \phi(m) = \sup_n \sup_n |F(z_1, \dots, z_n | z_{1-m}, \dots, z_{n-m}) - F(z_1, \dots, z_n)| \rightarrow 0 \quad \text{as } m \rightarrow \infty$$

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:

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

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

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

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 - A useful practical consequence of the analysis of stationarity is that the asymptotic solutions $z_t = g(a, a_{t-1}, a_{t-2}, \dots)$ to the various nonlinear representations, have a *direct* utilization by forecasting algorithms. By contrast, the computation of the optimal multistep predictor from the models in their original form of difference equations, is strongly affected by the presence of nonlinear regressors, such as $(z_{t-h}, a_{t-h}), h < k$. These difficulties clearly increase in the case of multilinear systems, so that suboptimal and pragmatic solutions must be sought.

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 ms-sense. It is well known that

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

but this conditional expectation is always linear only for $\{z_t\}$ gaussian. In the general nonlinear process (3.1) we easily find $\hat{z}_t(1) = f(\underline{x}_t)$, therefore a simplified multistep predictor can be obtained by extrapolating the identified function in the form of a deterministic difference equation, namely

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

The application of this approach to the multilinear model (3.2) 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 ahead 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 $\hat{z}_t(2) = \alpha E[z_{t+1}^2 | \mathfrak{S}_t] + \beta E[z_{t+1} | \mathfrak{S}_t] \cdot z_t$, where $E[z_{t+1}^2 | \mathfrak{S}_t] = E[(\alpha z_t^2 + \beta z_t z_{t-1})^2 + 2(\alpha z_t^2 + \beta z_t z_{t-1}) a_{t+1} + a_{t+1}^2 | \mathfrak{S}_t] = (\alpha z_t^2 + \beta z_t z_{t-1})^2 + \sigma^2$, 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 $\bar{z}_t(2) = \hat{z}_t(2) - \alpha \sigma^2$.

A pragmatic solution must also be adopted for the variances of prediction errors $\hat{\sigma}^2(l) = E[a_{t+l}^2 | \mathfrak{S}_t]$. 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 $\bar{z}_t(l)$, namely

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

4. Identification

Once the class of models for representing a nonlinear time series has been chosen, a crucial phase in the modeling is given by the selection of its orders. With respect to the class (3.2), the task is difficult since it requires the definition of the structure of monomials $\{y_{jt} = \prod_{i=1}^p z_{t-i}^{k_{ij}} \prod_{i=1}^q a_{t-i}^{h_{ij}}\}_{j=1 \dots n}$, i.e. of the powers k_{ij}, h_{ij} . For the sub-class (3.3), 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 $\text{IC} = -2 \log(\text{likelihood}) + f(N) \dim(\text{model})$:

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

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

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

Example 6 - Let $z_t = \delta a_{t-h} a_{t-k} + a_t$, $h < k$; simple calculation show 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 $\mu_3(h, k) = \mu_3(-h, -k) = \mu_3(k-h, -h) = \delta \sigma^4$ plus their permutation symmetries.

Both these approaches are of limited practical value since they rely heavily on the assumption that a true (regular) multilinear system exists. By contrast, data are often gene-

rated 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 (3.2), 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 $y_{jt} = \prod_i z_{t-i} \prod_i a_{t-i}$, namely $\phi(B)z_t = \beta_0 + \sum_{j=m}^n \beta_j y_{jt} + \theta(B)a_t$. Selection of significant regressors may then be developed with "second order" moments $E[z_t y_{jt-k}] = \gamma_{tj}(k)$; in practice we assume that with multiple products in (3.2), the series $\{y_{jt}\}_m^n$ acquire an autonomous nature with respect to the output $\{z_t\}$. Non-parametric techniques based on "cross covariance" can tentatively be used to find coefficients $\{\beta_j\}_m^n$ that have a chance to turn out significant in efficient estimations.

Example 7 - Consider the model of Hinich-Patterson $z_t = \sum_{i=1}^m \sum_{j=1}^n \delta_{ij} a_{t-i} a_{t-j} + a_t$; by defining $y_{t-h,k} = a_{t-h} a_{t-k}$ we clearly have $E[z_t y_{t-h,k}] = \delta_{hk} \sigma^4$. Analogously, for more complex models (3.3) a sufficient condition in order that $\alpha_{ij} \neq 0$, is given by the partial bicovariance $E[z_t y_{t-i,j} | y_{t-i,j-k}; k > 0] \neq 0$ where $y_{t-i,j-k} = z_{t-i-k} z_{t-j-k}$.

The strategy of putting coefficients in correspondence of every significant multicovariance $\gamma_{tj}(k)$, is certainly approximate and leads to overparametrization. However, it drastically reduces the number of terms to be considered in the estimation of information criteria. There are also a number of heuristic facts which make reasonable such a strategy in the earlier phase of identification:

- although $\gamma_{tj}(0) \neq 0$ is not sufficient to enable the existence of $\beta_j \neq 0$, it always represents a necessary condition, that is $\beta_j \neq 0 \Rightarrow E(z_t y_{jt}) \neq 0$;
- even if $\beta_j \neq 0$ implies $\gamma_{tj}(k) \neq 0$ for some $k \neq 0$, under stability conditions $\gamma_{tj}(0)$ provides the greatest value: $|E(z_t y_{jt})| \geq |E(z_t y_{jt-k})|$ for $k > 0$. This means that spurious regressors can be identified before estimation.
- There exists the possibility of strengthening the relationship covariance-coefficient by referring to partial multicovariances $E[z_t y_{jt} | y_{1t} \dots y_{j-1t}]$;
- finally, as shown by Example 7, in case of multilinear MA models (where $p = 0$) the relationship is one-to-one, that is $E[z_t (y_{jt} = \prod_{i=1}^j a_{t-i})] \neq 0 \Rightarrow \beta_j \neq 0$.

Explanation of second remark is given by the following example.

Example 9 - 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 $\gamma_{tj}(k) = \mu_3(i+k, j+k)$, 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 the partial multicorrelations is used to select the most appropriate pseudolinear 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. Hence, by adding and deleting the appropriate coefficients the best model should tentatively be determined.

Partial multicorrelations $\rho_{z_j|1\dots j-1} \propto E(z_j y_j | y_{1t} \dots y_{j-1t})$ can easily be estimated by utilizing the recurrence relationships which connect these coefficients

$$\rho_{z_j|1\dots j-1} = \frac{\rho_{z_j|1\dots j-2} \rho_{z_j-1|1\dots j-2} \rho_{jj-1|1\dots j-2}}{\sqrt{(1-\rho_{z_j-1|1\dots j-2}^2)(1-\rho_{jj-1|1\dots j-2}^2)}}$$

The candidate to enter the model is the variable $\{y_j\}$ which has the highest sample partial multicorrelation $r_{z_j|1\dots j-1}$. Approximate significance of these estimates can be evaluated by adapting the *F*-statistics described by Jenkins & Watts (1968, p.481), namely

$$\hat{F}_j = \frac{r_{z_j|1\dots j-1}^2 (N-p-k)}{1-r_{z_j|1\dots j-1}^2} \stackrel{L}{\approx} F(1, N-p-k)$$

where $j = k+1, \dots, n$ and k is the number of terms already included in the model.

This approach is typical of standard regression models in which the regressors are observable. In our context, the nonlinear system $z_t = \beta_0 + \sum_{i=1}^{j-1} \beta_{ji} y_{it} + a_{jt}$, $j = 2, 3 \dots n$ must be estimated at each step j , both in order to check the significance of the included coefficients as well as to generate the next candidate "regressors" $\hat{y}_{jt} = \left(\prod_{i=1}^{j-1} z_{t-i} \prod_{i=1}^{j-1} \hat{a}_{jt-i} \right)$. A reasonable approach is thus to estimate $r_{z_j|1\dots j-1}$ as simple correlations between \hat{a}_{jt} and \hat{y}_{jt} ; in summary, the two indicator for the selection of the regressors y_{it} are given by $r_{z_j} = \text{Cor}(z_t, \hat{y}_{jt})$ and $r_{z_j|1\dots j-1} = \text{Cor}(\hat{a}_{jt}, \hat{y}_{jt})$ for $j = 2, 3 \dots n$.

To simplify the method further, in particular to reduce the number of intermediate estimations, we suggest a procedure for the system (3.3) which refers to the bi-correlations functions $\rho_{z_y}(i, j) \propto E(z_t y_{t-i, j})$ 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 sample distribution of their estimators.

Proposition 1 - 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 high, we have

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

Proof - A heuristic demonstration of the statement is given in **Appendix A1**. We now present the various steps of the identification algorithm of the system (3.3).

Step 1 - Identify (p, q) (the ARMA 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 (4.2) (i.e. $|r_{z_y}(i, j)| > 2/\sqrt{N - \max(i, j)}$), in which $y_{t-i, j} = (z_{t-i} z_{t-j})$.

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

Step 4 - Fit the partial model $z_t = \sum_i^p \phi_i z_{t-i} + \dots + \sum_i^p \sum_j^q \beta_{ij} z_{t-i} \bar{a}_{t-j} + \hat{a}_t$ 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 $r_{xy}(i,j)$ 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

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

As we have stated previously, the rationale of the algorithm is based on the heuristics that the variables $\{y_{ij}\}$ tend to have a statistical behaviour of their *own* 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 are particularly true 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. There are some substantive points to be discussed in detail:

- If $(p,q) = 0$, i.e. $\phi(B) = \theta(B) = 1$, and $r_{xy}(i+k, j+k) \neq 0$ persistently for $k > 0$ with i, j fixed; then a parsimonious representation is provided by the impulse response function $z_t = [\alpha_{ij0} / \alpha_{ij}(B)] z_{t-i} z_{t-j}$ where $\alpha_{ij}(B)$ is a monic polynomial of, at most, second order. Estimation of this model requires the definition of the auxiliary variable $w_{ij} = \alpha_{ij1} w_{ij,t-1} + \dots + \alpha_{ij0} z_{t-i} z_{t-j}$ which can be generated adaptively; extension of this procedure to Steps 3 and 4 may be obtained by referring to the series $(z_t - w_{ij})$.
- Since the role of monomials $(\prod_j^q z_{t-j}), (\prod_j^p a_{t-j})$ 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.
- The rationale of the residual test at Step 6 is similar to that of the test for nonlinear ARMA models discussed by McLeod & Li (1983). While the reason for considering only *diagonal* coefficients $r(i=j)$ is merely practical, the utilization of the χ^2 -statistic is relatively new. The approximate distribution of (4.3) is a direct consequence of the distribution of multicorrelation coefficients (4.2). As shown in **Appendix A1** optimal statistical properties can be derived under a mild assumption of asymptotic independence.

The application - We now conclude the section, returning to the numerical application commenced in Section 2. Principal goal is to check the model building strategy discussed above, but also to compare the statistical performance of various multilinear models. Step 1 of the identification algorithm of the model (3.3) has been already carried out by (2.1), we therefore proceed directly to the second step.

Step 2 - In order to identify the significant quadratic components $y_{t-i,j} = (z_{t-i}z_{t-j})$ bi-auto correlations $r_{xy}(i,j)$ of Table 3 were considered. Setting coefficients α_{ij} in the same position (i,j) as every significant correlation, we obtained the intermediate model (3.3) 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

$$z_t = .14 + \underset{(1.6)}{.62}z_{t-1} - \underset{(10)}{.21}z_{t-1}z_{t-3} + \hat{a}_t, \quad \hat{\sigma}^2 = .87, R^2 = .43$$

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

Table 3 - 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

Note: The matrix is symmetrical and (i,j) can be interchanged.

Step 3 - Identification of the bilinear terms $(z_{t-i}a_{t-j})$ requires the inspection of partial bi-correlations $r_{xy}(i,j)$, with $\bar{y}_{t-i,j} = (z_{t-i}\hat{a}_{t-j})$, of Table 4. Following the procedure of Step 2 we selected the parameters $(\beta_{1,1}, \beta_{4,11}, \beta_{10,5})$ and the intermediate estimation gave

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

Table 4 - Partial bi-autocorrelations $r[z_t(z_{t-i}\hat{a}_{t-j})]$

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

Note: $\{\hat{a}_t\}$ are the residuals estimated at Step 2.

Steps 4,5 - For the identification of the quadratic terms $(a_{t-1}a_{t-j})$ the partial bicorrelations $r_{xy}(i,j)$ with $y_{t-i,j} = (\hat{a}_{t-i}\hat{a}_{t-j})$, of Table 5 were considered. Tentatively, the sole significant coefficients are $(\delta_{11}, \delta_{22})$; however, the term $\delta_{11}a_{t-1}a_{t-1}$ is clearly 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 impulse response function $z_t = [\delta_{11}/(1-\phi_1 B)]a_{t-1}$. 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 $(\beta_{11}, \delta_{22})$.

Table 5 - Partial bi-autocorrelations $r[z_t(\hat{a}_{t-i}\hat{a}_{t-j})]$

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

Note: $\{\hat{a}_t\}$ are the residuals estimated at Step 3.

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

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

with a value of the portmanteau statistic (4.3) $Q(3 \cdot 12) = 25.3 < 43.8 = \chi_{95}(30)$.

Evaluation based on the indicators $(\hat{\sigma}^2, R^2)$ and the significance of the regression coefficients, does not seem sufficient to check the statistical robustness of the above procedure. It is necessary, in fact, to refer to indicators which take into account the tradeoff between statistical fitting and parametric efficiency. Table 6 summarises the values of three information criteria corresponding to the models of the previous steps; in two cases the model that is indicated as the best is the final one. Results of BIC are partially due to its sensitivity to the number of observations $d = \max(p, P, r, R)$ lost in the estimation.

Table 6 - 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

This empirical example continues in the next section, which is concerned with methodological aspects of nonlinear estimation; in particular in conditions of time-varying parameters. In that context we will evaluate the forecasting performance of the model identified at Step 6 and other classes of nonlinear models will be considered.

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 the input process $\{a_t\}$ is known a-priori a natural candidate is the maximum likelihood method; however this condition contradicts the general formulation (3.1) where $a_t \sim \text{IID}$. There is also another feature that makes this approach unsuitable: explicitation of the ML-estimator is not always possible and optimization of the likelihood function must proceed by strictly numerical methods. For the MARMA model this situation involves problems in finding the unconditional algorithm and in analysing its statistical behaviour. In this section we adopt and extend the least-squares approach followed by Subba-Rao & Gabr (1984) for bilinear models.

In order to simplify the treatment, we rewrite model (3.2) in regression form and assume a *simplified* structure for its monomials, such as

$$\text{MARMA}^* \quad z_t = \beta_0 + \sum_{i=1}^n \beta_i y_{it} + a_t, \quad y_{it} = \left(\prod_{j=1}^{p_i} z_{t-j} \cdot \prod_{j=1}^{q_i} a_{t-j} \right) \quad (5.1)$$

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

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

improves the situation with respect to the ML-approach; but analytical expression of the gradient of the Gauss-Newton algorithm still remains difficult to implement

$$\text{NLS} \quad \hat{\beta}_N(k+1) = \hat{\beta}_N(k) + \left[\sum_{t=1}^N \xi_{\beta}(k) \xi_{\beta}'(k)' \right]^{-1} \sum_{t=1}^N \xi_{\beta}(k) a_t(k) \quad (5.2a)$$

$$\xi_{\beta}(\beta) = -\frac{\partial a_t}{\partial \beta} = y_t - \sum_{j=1}^q \left[\beta' \frac{\partial y_t}{\partial a_{t-j}} \right] \xi_{z_{t-j}} \quad (5.2b)$$

where $y_t(\beta)' = [1, y_{1t}, \dots, y_{nt}]$ and $q = \max(q_i)$ (see **Appendix A2** for the derivation). From (5.2b) we note that calculation of the gradient consists of a filtering on the pseudolinear regressors y_t , but with a filter that depends on random variables. This makes explicit the dependence of the algorithm (5.2a) on higher order moments of $\{z_t\}$.

In the linear ARMA context we simply have $\xi_{\beta} = x_t / \theta(\beta)$, i.e. $\xi_{\beta} = x_t - \sum_{j=1}^q \theta_j \xi_{z_{t-j}}$. By assuming $x_t' = [z_{t-1}, \dots, a_{t-q}]$ stationary gaussian, we also have $\{\xi_{\beta}\}$ covariance stationary and therefore $\{\xi_{\beta} \xi_{\beta}'\}$ is stationary in mean. Since $\hat{\beta}_N(k)$ is a minimizer, its consistency can be proved by applying the ergodic theorem to $t^{-1} \sum_{t=1}^t \xi(\tau) a(\tau) \rightarrow 0$ where $t = (k=N)$. In (5.2b) $\{\xi_{\beta}\}$ is not a linear transformation 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 (5.2a) may then follow simply because the process (5.2b) 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 $\delta_i \prod_{j=1}^m a_{i-j}^{k_j}$, 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 $a_t = h_t(z_t, z_{t-1}, z_{t-2}, \dots) = z_t - f_t(z_t)$.

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 $\bar{\beta} = \bar{E}(\beta_t)$. 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 $\left(\sum_{j=1}^{p_i} k_{ij}\right) \leq \left(\sum_{j=1}^{p_h} k_{hj}\right)$ for $i < h$ and $k_{ij} \geq 0$

$$\text{MAR}_t \quad z_t = \underline{y}'_t \beta_t + a_t, \quad (a_t | \mathfrak{S}_{t-1}) \sim \text{IID}(0, \sigma_t^2) \quad (5.3)$$

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

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

$$\hat{\beta}_N = \left(\frac{1}{N} \sum_{t=1}^N \underline{y}_t \underline{y}'_t \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^N \underline{y}_t z_t \right) = \arg \min \bar{E}(a_t) \quad (5.4)$$

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 coefficient function $\{\beta_t\}$ and the variance function of $\{z_t\}$ is needed, namely

$$(C1) \quad \bar{E}[E(\underline{y}_t \underline{y}'_t) \beta_t] = \bar{E}(\underline{y}_t \underline{y}'_t) \cdot \bar{E}(\beta_t)$$

$$(C2) \quad \bar{E}(a_t^2) < \infty, \quad \bar{E}(\underline{y}_t a_t) = 0, \quad \bar{E}(\underline{y}_t \underline{y}'_t) > 0$$

Indeed, multiplying (5.3) by \underline{y}_t and taking expectation we have $E(\underline{y}_t z_t) = E(\underline{y}_t \underline{y}'_t) \beta_t + E(\underline{y}_t a_t)$, next applying the operator $\bar{E}[\cdot]$, under the above conditions we may get

$$\bar{\beta} = \bar{E}(\underline{y}_t \underline{y}'_t)^{-1} \bar{E}(\underline{y}_t a_t) \quad (5.5)$$

Notice that, in general, given two deterministic bounded functions $f(t), g(t)$, the mean value of their product $h(t)$ 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. However, the admissibility of the orthogonality assumption (C1) stems from the fact that (i) $\{\beta_t\}$ is deterministic and has a stabilizing effect on $\{z_t\}$ (therefore it tends to move rapidly); (ii) $\bar{E}(\cdot)$ is an asymptotic operator, so that for finite intervals $t \in [1, N]$ we may really assume

$$(C1)' \quad \bar{E}_N(\underline{y}_t, \underline{y}'_t, \underline{\beta}_t) = \frac{1}{N} \sum_{i=1}^N E(y_{it}, y'_{it}, \underline{\beta}_t) = \bar{E}(\underline{y}_t, \underline{y}'_t) \bar{\beta} + O\left(\frac{1}{N}\right)$$

Clearly, (C1)-(C1)' are always satisfied if $\{\underline{\beta}_t\}$ is constant, periodic or monotonic.

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

Proposition 2 - Consider the time-varying multilinear AR model (5.3) in which :

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

(A1)' with $\phi(m) = O(1/m^h)$, $h > r/(2r - 1)$, $r \geq 1$ and

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

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

iii) conditions (C1),(C2) are satisfied, or alternatively (C1)' ;

iv) the sequence $\bar{E}_N(\underline{y}_t, \underline{y}'_t) = \{N^{-1} \sum_{i=1}^N E(y_{it}, y'_{it})\}$ is uniformly positive definite.

Then for N sufficiently large the OLS estimator (5.4) exists with probability one and is consistent for the average trajectory $\bar{\beta} = \bar{E}(\underline{\beta}_t)$.

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 **Appendix A3**. In simulation experiments we have checked that OLS is an accurate estimator for $\bar{\beta}$ if the sequence $\{\underline{\beta}_t\}$ 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 (5.1), 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 (5.2b) as $\xi_{\beta}(\underline{\beta}) \approx \underline{y}_t(\underline{\beta})$ and inserting the corresponding iterative expression $\hat{\underline{y}}_t(k)$, together with $\hat{a}_t(k) = z_t - \hat{\underline{y}}_t(k)' \hat{\underline{\beta}}(k)$, in (5.2a) ; the final algorithm has the same structure as the OLS (5.4), but is iterative. In the context of nonlinear models, this approach significantly reduce 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 & 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 $\{\underline{\beta}_t\}$ 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 (5.2a) and with $\xi_{\beta}(k)$ replaced by $\hat{\underline{y}}_t(k)$, the Recursive PLR estimator of (5.1) becomes

$$\text{R-PLR} \quad \bar{a}(t) = z_t - \hat{\underline{\beta}}(t-1)' \hat{\underline{y}}(t) \tag{5.6a}$$

$$R(t) = \lambda \cdot R(t-1) + \hat{\underline{y}}(t) \hat{\underline{y}}(t)' \tag{5.6b}$$

$$\hat{\underline{\beta}}(t) = \hat{\underline{\beta}}(t-1) + R(t)^{-1} \hat{\underline{y}}(t) \bar{a}(t) \tag{5.6c}$$

$$\hat{a}(t) = z_t - \hat{\underline{\beta}}(t)' \hat{\underline{y}}(t) \tag{5.6d}$$

$$J(t) = \lambda \cdot J(t-1) + \hat{d}(t)^2 \quad (5.6e)$$

$$\hat{y}(t+1) = \left\{ \prod_{j=1}^{t+1} z_{t+1-j} \prod_{j=1}^t \hat{d}(t+1-j) \right\} \quad (5.6f)$$

The terms \hat{a}_t, \hat{d}_t are respectively the prediction error and the recursive residual; the factor $0 < \lambda < 1$ by preventing $R(t)$ from vanishing, enables parameter changes ($\underline{\beta}_t - \underline{\beta}_{t-1}$) 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^2 .

Returning to the efficient NLS estimation, the exact recursive expression of (5.2) can be obtained from (5.6) by replacing $\hat{y}(t)$ with $\underline{\xi}(t)$ in the equations of $R(t), \underline{\beta}(t)$ and inserting the filter

$$\underline{\xi}(t) = \hat{y}(t) - \sum_{j=1}^q \left[\underline{\beta}(t)' \frac{\partial \hat{y}(t)}{\partial \hat{d}(t-j)} \right] \underline{\xi}(t-j) \quad (5.6g)$$

Under the assumption of constant parameters the resulting algorithm tends to minimize the weighted functional $J_t = \sum_{i=1}^t \lambda^{t-i} a_i^2(\underline{\beta})$; however, in the context of evolving systems, it is not clear what may be the improvement in terms of the MSE $E \|\underline{\beta}(t) - \underline{\beta}_t\|^2$. Given the complexity and multilinearity of the filtering (5.6g), a worsening of the tracking capability with respect to (5.6c) cannot be ruled out.

The parameters of (5.1) may vary with time depending on the goodness with which the multilinear model (3.2) approximates the "true" nonlinear function (3.1). 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 approximated by a linear ARMA, 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, used in state dependent systems (see Priestley, 1988 p.99), algorithm (5.6) is much more easy to implement since it only requires as *priors* $0 < \lambda < 1, R(0)^{-1} = \rho \cdot I_n$ (and usually $.95 < \lambda < .99, .01 < \rho < 2.0$); moreover, it only assumes that parameters do not change suddenly, in a deterministic or stochastic fashion. To be more specific, Kalman Filter implementation requires that parameters follow a linear, or a linearizable, process, whereas recursive algorithms, by making estimates smooth functions of the past observations, implicitly assume $\underline{\beta}_t = f(\mathcal{S}_{t-1})$. The weighting sequence $\{\lambda(t, \tau)\}$ should then be designed according to the path of conditional probabilities $P_t(\underline{\beta}_t | z_{t-\tau})$ or to that of cross correlations $\text{Cor}(\underline{\beta}_t, z_{t-\tau})$. These informations, however, are not generally available a priori and other, more pragmatic, criteria must be followed.

A way of avoiding altogether the problem of priors in the recursive estimation, consists of reducing algorithm (5.6) to a stochastic approximation scheme. This may be approached by setting $\lambda \equiv 1$ and replacing $R(t)$ with $\bar{R}(t) = R(t)/t$, so as to retain the

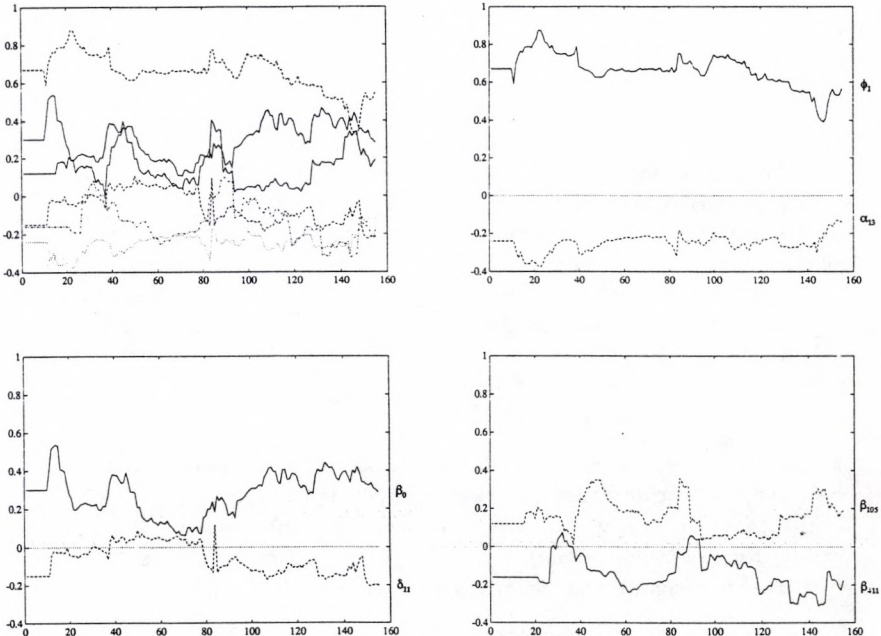
tracking capability . In this case the updating rule equivalent to (5.6b) becomes

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

and k -iterating the recursions one may initialize $\bar{R}_{k+1}(0) = \bar{R}_k(N)$. A stochastic approximation type solution is achieved if $\bar{R}(t)$ converge to a matrix $0 < \bar{R} < \infty$ as $t \rightarrow \infty$; however, for problems explained above this may be guaranteed only if $\{q_i\}_1^n = 0$.

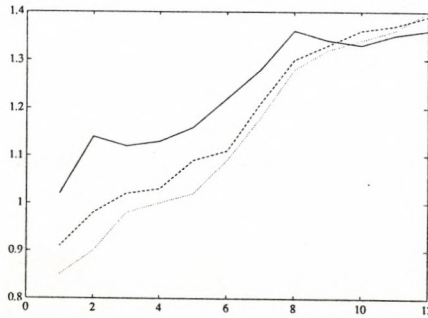
The Application - The parameters of the model identified at Step 6 in the previous section, were successively estimated with the recursive algorithm (5.6) in order to check their stability over time. Initial values were set $\hat{\beta}(0) = \hat{\beta}_N$, $R(0) = \bar{R}_N$ obtained from the iterative estimation. The optimal forgetting factor $\hat{\lambda}_w = .974$ was derived with a search procedure by minimizing the global objective function $J_N(\lambda) = \sum_{t=d}^N [\hat{d}(t)^2 + \|\hat{\beta}(t) - \hat{\beta}_N\|]$ which establishes a tradeoff between tracking and accuracy. Trajectories of $\hat{\beta}_i(t)$, $i=1..6$ are shown in Figure 3 ; notice that in spite of the mild value taken by λ a significant variability of parameters occurs. This is a clear indication that the process $\{z_t\}$ is nonstationary 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 in stabilizing the behaviour of the output. Finally, the gain in statistical fitting allowed by recursive estimation is pointed out by the value of the residual variance : $(N-d)^{-1} \sum_{t=d}^N \hat{d}(t)^2 = .63$.

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



Results of Table 6 and Figure 3 then confirm the effectiveness of the modeling procedure discussed so far; however, they are essentially in-sample behaviour. Another necessary test concerns with the operative, out-of-sample performance; specifically, does the multilinear modeling enable an improvement over the forecasting ability of standard linear models? To check this point, we took 24 post-sample observations and we evaluated the mean absolute prediction error $\text{MApE}_t(l|h) = h^{-1} \sum_{i=1}^h |z_{t+\tau+i} - \hat{z}_{t+\tau+i}(l)|$. In this statistic $h = 12$ is the sample size of the mean, $t + \tau = 1985.12 + \tau$ are the forecast origins and $l = 1, 2 \dots 12$ are prediction steps. Results corresponding to models at Step 1 (2.1) and Step 6 (with parameters estimated off-line and on-line) are displayed in Figure 4. It can be appreciated the better forecasting ability of nonlinear models, in particular with time varying parameter. This outcome is particular important in view of the fact that the forecasting algorithm utilized in practice was the sub-optimal one described in Section 3, and in the case of time varying model, extrapolation of the parameter function was not made, i.e. $\hat{\beta}_t(l) = \hat{\beta}(t)$.

Figure 4 - MApE statistics for models at Steps 1 (—) and 6 (---, ...)



To complete this section we now evaluate the statistical performance of neural network and exponential autoregression models on the same set of data. The framework of neural networks (see White, 1989) is based on a sequence of hidden units having common input variables $\{\underline{x}_t\}$ and the same structure $\psi(\cdot)$. The responses of these units interact at an intermediate layer, before reaching the output $\{y_t\}$. Accordingly, the shape of $\psi(\cdot)$ is represented by a threshold rule or a sigmoid function (e.g. a probability distribution function)

$$y_t = \Psi \left[\sum_{j=1}^q \beta_j \psi(\underline{x}_t' \alpha_j) \right] + a_t$$

Example 10 - Let $y = \Psi(\underline{x}'\beta)$ and $\Psi(\cdot)$ be the normal distribution; y is then the conditional expectation of a Bernoulli random variable generated by a probit model.

This general framework has important connections with the *projection pursuit regression* designed for non-parametric curve fittings (see Friedman & 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) are also evident; to appreciate this, assume that $\psi(\cdot)$ is the logistic function $[1 + \exp(\cdot)^{-1}]^{-1}$, $\Psi \equiv 1$ and let $y_t = z_t$, $x_t = z_{t-1}$, then

$$\text{N-N: } z_t = \beta_0 + \sum_{j=1}^q \beta_j [1 + \exp(-\alpha_j z_{t-j})]^{-1} + a_t \tag{5.7a}$$

$$\text{EAR: } z_t = \beta_0 + \sum_{j=1}^q [\phi_j + \beta_j \exp(-\alpha_j z_{t-j}^2)] z_{t-j} + a_t \tag{5.7b}$$

Neural networks like (5.7a) were successfully applied by White (1989) for approximating the Hénon map $y_t = 1 - 1.4 y_{t-1}^2 + .3 y_{t-2}$ whose series, given values for (y_0, y_{-1}) , is highly erratic. The number of parameters required, however, was rather large (16-21).

The application of (5.7a) to the data of Figure 1 provided disappointing results. Given the sensitivity of the behaviour of the model to initial values, the identification of the order q has been developed stepwise. For $q > 2$ no significant improvement of $\hat{\sigma}_a^2$ was achieved and the corresponding estimate was

$$z_t = \underset{(1.3)}{.27} - \underset{(3.2)}{2.1} \left[1 + \exp\left(\underset{(1.1)}{.41} + \underset{(2.1)}{1.9} z_{t-1}\right) \right]^{-1} + \underset{(1.7)}{.87} \left[1 + \exp\left(\underset{(-3.4)}{-43.6} - \underset{(-2.8)}{29.1} z_{t-1}\right) \right]^{-1}, \sigma^2 = .88, R^2 = .44$$

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*) forecasting algorithms, for practical use, are very complicated.

Data of Figure 1 were also fitted with the exponential model (5.7b), with $q = 2$, providing a residual variance equal to that of the AR(1) model at Step 1. The reasons for this behaviour are similar to those of neural networks since this class of models has been designed for particular physical systems (see Ozaki, 1985).

6. Conclusions

If time series data are dominated by non-gaussian characteristics it is often necessary to use a nonlinear representation and this immediately poses the question as to what class of models to use. The multilinear ARMA system provides a natural and compact representation for covering, in a parsimonious manner, the higher order moment information useful in prediction.

Since multilinear models retains the regression structure, many recursive procedures of identification, filtering and forecasting of standard time series analysis can be generalized without difficulties. In this paper a combined partial- multicorrelation and pseudolinear- regression algorithm has been discussed and successfully applied to real economic data.

Appendices

A1 - Proof of Proposition 1

Assume $\{z_t\}$ to be zero mean and consider the investigation of the sample bicovariance function $c_3(i, j) = r_{z_3}(i, j) \cdot \hat{\sigma}_z \hat{\sigma}_y$, with $y_{t-l, j} = (z_t z_{t-l} z_{t-j})$. For $l = \max(h, k, i, j) > 0$ small we may set $(N-l) = N$, so that

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

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

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

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

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

and under the null hypothesis $H_0: z_t \equiv a_t$ we may easily obtain

$$\text{Var}[c_3(i, j)] = \frac{1}{N} \mu_6(0, i, j, i, j) = \frac{1}{N} E[z_i^2 z_{i-i}^2 z_{i-i}^2] = \frac{1}{N} \sigma^6$$

indeed, recall that by stationarity a 0-lag term in $\mu_6(\cdot)$ has been omitted.

The distribution of (4.2) easily follows by noting that $\sigma_x \sigma_y = \sigma^3$ and that $\{z_t, z_{t-1}, z_{t-2}, \dots\}$ is an asymptotically independent sequence under the same condition for $\{z_t\}$ (see Lemma 1). The central limit theorem for the sums of asymptotically independent random variables (see White, 1984) can be applied to show that

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

As a corollary we easily find the distribution of the portmanteau statistic (4.3).

A2 - Derivation of Estimator (5.2)

Algorithm (5.2a) is obtained by expanding $a_t(\beta)$ in Taylor series about an initial estimate $\hat{\beta}$ and applying iteratively the OLS method to the resulting equation : $\hat{a}_t = (\hat{\beta} - \beta)' \xi_{2t} + a_t$. To derive the gradient (5.2b), consider the model (5.1) in regression form $z_t = \beta' y_t(\beta) + a_t$, where $y_t' = [1, y_{1t}, \dots, y_{nt}]$; hence, by definition

$$\begin{aligned} \xi_{2t}(\beta) &= -\frac{\partial a_t(\beta)}{\partial \beta} = +\frac{\partial \beta' y_t(\beta)}{\partial \beta} \\ &= y_t - \beta' \frac{\partial y_t}{\partial \beta} = y_t - \sum_{i=1}^n \beta_i \frac{\partial y_{it}}{\partial \beta} \\ &= y_t - \sum_{i=1}^n \beta_i \left[\sum_{j=1}^q \left(\prod_{k=1}^{p_i} z_{t-k} \prod_{k \neq j}^{q_i} a_{t-k} \right) \right] \\ &= y_t - \sum_{j=1}^q \left[\sum_{i=1}^n \beta_i \left(\prod_{k=1}^{p_i} z_{t-k} \prod_{k \neq j}^{q_i} a_{t-k} \right) \right] \xi_{2t-j} \\ &= y_t - \sum_{j=1}^q \left[\beta' \frac{\partial y_t}{\partial a_{t-j}} \right] \xi_{2t-j} \end{aligned}$$

where $q = \max(q_i)$ and by assuming stationarity we have used $\xi_{2jt} \equiv \xi_{2t-j}$.

Example 11 - Consider the model (3.3) with $(p = q = r = s = P = Q = R = S) = 1$; the vector of parameters is given by $\beta' = [\beta_0, \phi_1, \theta_1, \beta_1, \alpha_1, \delta_1]$ and that of "regressors" $y_t' = [1, z_{t-1}, a_{t-1}, z_{t-1}a_{t-1}, z_{t-1}^2, a_{t-1}^2]$. Hence $\beta' \frac{\partial y_t}{\partial a_{t-1}} = (\theta_1 + \beta_1 z_{t-1} + 2\delta_1 a_{t-1})$ is a linear process.

A3 - Proof of Proposition 2

To prove the proposition we need to recall two formal results concerning the law of large number for nonstationary processes and the properties of transformations of mixing sequences.

Theorem 1 (McLeish, 1975) - Let $\{z_t\}$ be a heterogeneous mixing sequence with $\phi_r(m) = O(m^{-h})$ for $h > r/(2r - 1)$, $r \geq 1$ and finite means $\mu_r = E(z_t) = O(1)$. If for some $\delta, 0 < \delta \leq r$, we have $\sum_{t=1}^N (E|z_t - \mu_r|^{r+\delta}/t^{r+\delta})^{1/r} < \infty$, then $N^{-1} \sum_{t=1}^N (z_t - \mu_r) = (\bar{z}_N - \bar{\mu}_N) \rightarrow 0$ with probability one.

Remark 1 - Notice that the third condition of the theorem may be replaced by $E|z_t|^{r+\delta} < \infty, \forall t$. Given in fact $\mu_r = O(1)$, it follows that $\sup_t E|z_t - \mu_r|^{r+\delta} = M < \infty$; therefore $\sum_{t=1}^N (\cdot)^{1/r} < M^{1/r} \sum_{t=1}^N (1/t^{r+\delta})^{1/r} < \infty$. With respect to the standard ergodic theorem, which requires $E|z_t| < \infty, \forall t$, the presence of nonstationarity then increases the restrictions on the moments, proportionally to the amount of memory in the process.

Lemma 1 (White & Domowitz, 1984) - Let $\{z_t\}$ be a mixing sequence with $\phi_r(m) = O(m^{-h})$ for $h > 0$ and $g(\cdot)$ a measurable function from \mathfrak{R}^{n+1} onto \mathfrak{R} . Then the process $y_t = g(z_t, z_{t-1}, \dots, z_{t-n})$ with $n < \infty$ is mixing of the same order as $\{z_t\}$; i.e. $\phi_r(m) = O(m^{-h}), h > 0$.

With respect to stationary sequences, it is interesting to recall that ergodicity (a form of asymptotic independence on average), is invariant to transformations that involve an infinite number of past values of the process (see Stout, 1974, p.192).

Now concentrating the attention on the component $(N^{-1} \sum_{t=1}^N y_t y_t')$ of the OLS estimator, we remark the following facts:

1) given that variances \geq autocovariances, by assumption (i)-(A2)' we obtain

$$E|y_{it}^2|^{r+\delta} = E \left| \prod_{j=1}^{p_i} z_{t-j}^{k_{ij}} \right|^{2(r+\delta)} \leq E|z_t^{\kappa_i}|^{2(r+\delta)} \leq E|z_t^{\kappa_i}| < M < \infty \quad \forall i, t$$

where $\kappa_i = \sum_{j=1}^{p_i} k_{ij}$. As a consequence, by the Cauchy-Schwarz inequality it follows that

$$E|y_{it} y_{jt}|^{r+\delta} \leq (E|y_{it}^2|^{r+\delta})^{1/2} \cdot (E|y_{jt}^2|^{r+\delta})^{1/2} < (M^{1/2} M^{1/2}) \quad \forall i, j, t$$

therefore the matrix process $\{(y_t, y_t')\}$ is quasi stationary of order $K = 2(r + \delta)\kappa$, $\kappa = \max_i(\kappa_i)$, by condition (ii)-(B2).

2) Given the assumption (i)-(A1)', and $p = \max_i(p_i) < \infty$, the univariate processes $\{y_{it}^2\}, \{y_{it} y_{jt}\}$ are mixing of the same order as $\{z_t\}$. Therefore the matrix process $\{(y_t, y_t')\}$ is mixing of order $h > r/(2r - 1), r \geq 1$.

3) Combining these results we may apply Theorem 1 to each sequence $w_{ijt} = \{y_{it} y_{jt}\}$, to show that

$$\left(\frac{1}{N} \sum_{t=1}^N y_t y_t' \right) = \left\{ \frac{1}{N} \sum_{t=1}^N w_{ijt} \right\} \xrightarrow{P} \bar{E}(y_t y_t') = \{\bar{E}(w_{ijt})\} < \infty \quad (*)$$

4) Utilizing a generalization of Slutsky theorem we have $\{\text{Det}(N^{-1} \sum_{t=1}^N y_t y_t') - \text{Det}[\bar{E}_N(y_t y_t')]\} \rightarrow 0$ in probability. Since by hypothesis (iv) $\text{Det}[\bar{E}_N(y_t y_t')] > \delta$, for N sufficiently large we also have $\text{Det}(N^{-1} \sum_{t=1}^N y_t y_t') > \delta/2$; therefore the OLS estimator (5.4)

exists nearly uniformly. The estimator may also be unbiased since $\bar{E}_N(\underline{y}_i \underline{y}_i')$ is uniformly bounded; indeed, its entries satisfy

$$\left| \frac{1}{N} \sum_{i=1}^N E(y_{it} y_{jt}) \right| \leq \frac{1}{N} \sum_{i=1}^N |E(y_{it} y_{jt})| \leq \frac{1}{N} \sum_{i=1}^N |y_{it} y_{jt}|$$

$$\leq \frac{1}{N} \sum_{i=1}^N (E |y_{it}^2|)^{1/2} (E |y_{jt}^2|)^{1/2} < \frac{1}{N} \sum_{i=1}^N M = M$$

by Jensen and Cauchy-Schwarz inequalities.

5) With a similar reasoning it can be shown that the statistic $(N^{-1} \sum_{i=1}^N \underline{y}_i z_i)$ is bounded in probability and is a consistent estimator for $\bar{E}(\underline{y}_i z_i)$. Combining this result with (*), by means of Slutsky theorem, we finally obtain that $\hat{\beta}_N \xrightarrow{p} \bar{E}(\underline{y}_i \underline{y}_i')^{-1} \bar{E}(\underline{y}_i z_i) \equiv \bar{\beta}$ (5.5), as a consequence of assumption (ii) - (C1)'.

Remark - Note that in obtaining the consistency of $\hat{\beta}_N$, no restrictions have been placed on $\text{Var}(a_i)$ and $\text{Cov}(\underline{y}_i a_i)$. However, mild conditions such as $\bar{E}_N(\underline{y}_i a_i) = 0$, $\bar{E}_N(a_i^2) < \infty$, uniformly in N , are suitable ones for the predictive ability of the model and approximate properties of unbiasedness and efficiency of the OLS estimator.

REFERENCES

- BHASKARA-RAO M., SUBBA-RAO T., WALKER A.M. (1983), On the existence of some bilinear time series models. *Jour. of Time Series Analysis* 4, 93-110.
- FRIEDMAN J.H., STUETZLE W. (1981), Projection pursuit regression. *Jour. of American Statistical Association* 76, 817-824.
- GABR M.M. (1988), On the third order moment structure of some bilinear time series. *Jour. of Time Series Analysis* 9, 11-20.
- GRANGER C.W.J., ANDERSEN A.P. (1978), *An introduction to bilinear time series models*. Göttingen: Vandenhoeck & Ruprecht.
- GRILLENZONI C. (1990), Modeling Time-Varying Dynamical Systems. *Jour. of American Statistical Association* 85, 499-507.
- GRILLENZONI C. (1991), Simultaneous transfer functions versus vector ARMA models. *Journal of Forecasting* 10, 477-499.
- 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,a), Evidence of nonlinearity in daily stock returns. *Jour. of Business & Economic Statistics* 3, 69-77.
- HINICH M.J., PATTERSON D.M. (1985,b), Identification of the coefficients in a non-linear time series of the quadratic type. *Jour. of Econometrics* 30, 269-288.
- JARQUE C.M., BERA A.K. (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters* 6, 225-259.
- JENKINS G.M., WATTS D.G. (1968), *Spectral analysis and its applications*. San Francisco: Holden Day.

- LAHA M., ROHATGI R. (1979), *An introduction to probability theory*. New York: Wiley.
- LI W.K. (1984), On the autocorrelation structure and identification of some bilinear time series. *Jour.of Time Series Analysis* 5, 173-181.
- LIU J., BROCKWELL P.J. (1988), On the general bilinear time series model. *Jour.of Applied Probability* 25, 553-564.
- 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 time series nonlinearity. *Biometrika* 72, 39-44.
- KUMAR K. (1986), On the identification of some bilinear time series models. *Jour.of Time Series Analysis* 7, 117-122.
- MARAVALL A. (1983), An application of nonlinear time series forecasting. *Jour.of Business & Economic Statistics* 1, 66-74.
- McLEISH D.L. (1975), A maximal inequality and dependent strong laws. *Annals of Probability* 3, 826-836.
- 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.
- OZAKI T. (1985), Nonlinear time series models and dynamical systems. In *Handbook of Statistics* Vol.5 (eds. Hannan E.J., Krishnaiah P.R.). Amsterdam : North-Holland.
- PETRUCCELLI J.D., DAVIES N. (1986), A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* 73, 687-694.
- PHAM D. T. (1985), Bilinear markovian representation and bilinear models. *Stochastic Processes and Applications* 20, 295-306.
- PRIESLEY M.B. (1988), *Nonlinear and nonstationary time series analysis*. New York : Academic Press.
- RUGH W.J. (1981), *Nonlinear system theory*. Baltimore : J.Hopkins University Press.
- SESAY S.A.O., SUBBA-RAO T. (1988), Yule-Walker type difference equations for higher order moments for bilinear models. *Jour.of Time Series Analysis* 9, 385-401.
- SOLO V. (1978), *Time series recursions and stochastic approximation* (Ph.D. dissertation). Canberra: A.N.U. Dept.of Statistics.
- STOUT H. (1974), *Almost sure convergence*. New York : Academic Press.
- 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: Clarendon Press.
- TSAY R.S. (1989), Testing and modeling threshold autoregressive processes. *Jour.of American Statistical Association* 84, 231-240.
- WEISS A.A. (1986), ARCH and bilinear time series models : comparison and combination. *Jour.of Business and Economic Statistics* 4, 59-70.
- WHITE H. (1984), *Asymptotic theory for econometricians*. Orlando(FL): Academic Press.
- WHITE H., DOMOWITZ I. (1984), Nonlinear regression with dependent observations. *Econometrica* 52, 143-161.
- WHITE H. (1989), Some asymptotic results for learning in single hidden-layer feedforward network models. *Jour.of American Statistical Association* 84, 1003-1013.



EUI WORKING PAPERS

EUI Working Papers are published and distributed by the
European University Institute, Florence

Copies can be obtained free of charge
– depending on the availability of stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

Publications of the European University Institute

To The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

From Name

 Address

- Please send me a complete list of EUI Working Papers
- Please send me a complete list of EUI book publications
- Please send me the EUI brochure Academic Year 1993/94
- Please send me the EUI Research Report

Please send me the following EUI Working Paper(s):

No, Author

Title:

No, Author

Title:

No, Author

Title:

No, Author

Title:

Date

Signature



**Working Papers of the Department of Economics
Published since 1990**

ECO No. 90/1

Tamer BASAR and Mark SALMON
Credibility and the Value of Information
Transmission in a Model of Monetary
Policy and Inflation

ECO No. 90/2

Horst UNGERER
The EMS – The First Ten Years
Policies – Developments – Evolution

ECO No. 90/3

Peter J. HAMMOND
Interpersonal Comparisons of Utility:
Why and how they are and should be
made

ECO No. 90/4

Peter J. HAMMOND
A Revelation Principle for (Boundedly)
Bayesian Rationalizable Strategies

ECO No. 90/5

Peter J. HAMMOND
Independence of Irrelevant Interpersonal
Comparisons

ECO No. 90/6

Hal R. VARIAN
A Solution to the Problem of
Externalities and Public Goods when
Agents are Well-Informed

ECO No. 90/7

Hal R. VARIAN
Sequential Provision of Public Goods

ECO No. 90/8

T. BRIANZA, L. PHLIPS and J.F.
RICHARD
Futures Markets, Speculation and
Monopoly Pricing

ECO No. 90/9

Anthony B. ATKINSON/ John
MICKLEWRIGHT
Unemployment Compensation and
Labour Market Transition: A Critical
Review

ECO No. 90/10

Peter J. HAMMOND
The Role of Information in Economics

ECO No. 90/11

Nicos M. CHRISTODOULAKIS
Debt Dynamics in a Small Open
Economy

ECO No. 90/12

Stephen C. SMITH
On the Economic Rationale for
Codetermination Law

ECO No. 90/13

Elettra AGLIARDI
Learning by Doing and Market Structures

ECO No. 90/14

Peter J. HAMMOND
Intertemporal Objectives

ECO No. 90/15

Andrew EVANS/Stephen MARTIN
Socially Acceptable Distortion of
Competition: EC Policy on State Aid

ECO No. 90/16

Stephen MARTIN
Fringe Size and Cartel Stability

ECO No. 90/17

John MICKLEWRIGHT
Why Do Less Than a Quarter of the
Unemployed in Britain Receive
Unemployment Insurance?

ECO No. 90/18

Mrudula A. PATEL
Optimal Life Cycle Saving With
Borrowing Constraints:
A Graphical Solution

ECO No. 90/19

Peter J. HAMMOND
Money Metric Measures of Individual
and Social Welfare Allowing for
Environmental Externalities

ECO No. 90/20

Louis PHLIPS/
Ronald M. HARSTAD
Oligopolistic Manipulation of Spot
Markets and the Timing of Futures
Market Speculation

ECO No. 90/21

Christian DUSTMANN
Earnings Adjustment of Temporary
Migrants

ECO No. 90/22

John MICKLEWRIGHT
The Reform of Unemployment
Compensation:
Choices for East and West

ECO No. 90/23

Joerg MAYER
U. S. Dollar and Deutschmark as
Reserve Assets

ECO No. 90/24

Sheila MARNIE
Labour Market Reform in the USSR:
Fact or Fiction?

ECO No. 90/25

Peter JENSEN/
Niels WESTERGÅRD-NIELSEN
Temporary Layoffs and the Duration of
Unemployment: An Empirical Analysis

ECO No. 90/26

Stephan L. KALB
Market-Led Approaches to European
Monetary Union in the Light of a Legal
Restrictions Theory of Money

ECO No. 90/27

Robert J. WALDMANN
Implausible Results or Implausible Data?
Anomalies in the Construction of Value
Added Data and Implications for Esti-
mates of Price-Cost Markups

ECO No. 90/28

Stephen MARTIN
Periodic Model Changes in Oligopoly

ECO No. 90/29

Nicos CHRISTODOULAKIS/
Martin WEALE
Imperfect Competition in an Open
Economy

ECO No. 91/30

Steve ALPERN/Dennis J. SNOWER
Unemployment Through 'Learning From
Experience'

ECO No. 91/31

David M. PRESCOTT/Thanasis
STENGOS
Testing for Forecastable Nonlinear
Dependence in Weekly Gold Rates of
Return

ECO No. 91/32

Peter J. HAMMOND
Harsanyi's Utilitarian Theorem:
A Simpler Proof and Some Ethical
Connotations

ECO No. 91/33

Anthony B. ATKINSON/
John MICKLEWRIGHT
Economic Transformation in Eastern
Europe and the Distribution of Income*

ECO No. 91/34

Svend ALBAEK
On Nash and Stackelberg Equilibria
when Costs are Private Information

ECO No. 91/35

Stephen MARTIN
Private and Social Incentives
to Form R & D Joint Ventures

ECO No. 91/36

Louis PHILIPS
Manipulation of Crude Oil Futures

ECO No. 91/37

Xavier CALSAMIGLIA/Alan KIRMAN
A Unique Informationally Efficient and
Decentralized Mechanism With Fair
Outcomes

ECO No. 91/38

George S. ALOGOSKOUFIS/
Thanasis STENGOS
Testing for Nonlinear Dynamics in
Historical Unemployment Series

ECO No. 91/39

Peter J. HAMMOND
The Moral Status of Profits and Other
Rewards:
A Perspective From Modern Welfare
Economics

* Working Paper out of print

ECO No. 91/40

Vincent BROUSSEAU/Alan KIRMAN
The Dynamics of Learning in Mis-Specified Models

ECO No. 91/41

Robert James WALDMANN
Assessing the Relative Sizes of Industry- and Nation Specific Shocks to Output

ECO No. 91/42

Thorsten HENS/Alan KIRMAN/Louis PHILIPS
Exchange Rates and Oligopoly

ECO No. 91/43

Peter J. HAMMOND
Consequentialist Decision Theory and Utilitarian Ethics

ECO No. 91/44

Stephen MARTIN
Endogenous Firm Efficiency in a Cournot Principal-Agent Model

ECO No. 91/45

Svend ALBAEK
Upstream or Downstream Information Sharing?

ECO No. 91/46

Thomas H. McCURDY/
Thanasis STENGOS
A Comparison of Risk-Premium Forecasts Implied by Parametric Versus Nonparametric Conditional Mean Estimators

ECO No. 91/47

Christian DUSTMANN
Temporary Migration and the Investment into Human Capital

ECO No. 91/48

Jean-Daniel GUIGOU
Should Bankruptcy Proceedings be Initiated by a Mixed Creditor/Shareholder?

ECO No. 91/49

Nick VRIEND
Market-Making and Decentralized Trade

ECO No. 91/50

Jeffrey L. COLES/Peter J. HAMMOND
Walrasian Equilibrium without Survival: Existence, Efficiency, and Remedial Policy

ECO No. 91/51

Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
Preferred Point Geometry and Statistical Manifolds

ECO No. 91/52

Costanza TORRICELLI
The Influence of Futures on Spot Price Volatility in a Model for a Storable Commodity

ECO No. 91/53

Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
Preferred Point Geometry and the Local Differential Geometry of the Kullback-Leibler Divergence

ECO No. 91/54

Peter MØLLGAARD/
Louis PHILIPS
Oil Futures and Strategic Stocks at Sea

ECO No. 91/55

Christian DUSTMANN/
John MICKLEWRIGHT
Benefits, Incentives and Uncertainty

ECO No. 91/56

John MICKLEWRIGHT/
Gianna GIANNELLI
Why do Women Married to Unemployed Men have Low Participation Rates?

ECO No. 91/57

John MICKLEWRIGHT
Income Support for the Unemployed in Hungary

ECO No. 91/58

Fabio CANOVA
Detrending and Business Cycle Facts

ECO No. 91/59

Fabio CANOVA/
Jane MARRINAN
Reconciling the Term Structure of Interest Rates with the Consumption Based ICAP Model

ECO No. 91/60

John FINGLETON
Inventory Holdings by a Monopolist Middleman

ECO No. 92/61

Sara CONNOLLY/John
MICKLEWRIGHT/Stephen NICKELL
The Occupational Success of Young Men
Who Left School at Sixteen

ECO No. 92/62

Pier Luigi SACCO
Noise Traders Permanence in Stock
Markets: A Tâtonnement Approach.
I: Informational Dynamics for the Two-
Dimensional Case

ECO No. 92/63

Robert J. WALDMANN
Asymmetric Oligopolies

ECO No. 92/64

Robert J. WALDMANN /Stephen
C. SMITH
A Partial Solution to the Financial Risk
and Perverse Response Problems of
Labour-Managed Firms: Industry-
Average Performance Bonds

ECO No. 92/65

Agustín MARAVALL/Víctor GÓMEZ
Signal Extraction in ARIMA Time Series
Program SEATS

ECO No. 92/66

Luigi BRIGHI
A Note on the Demand Theory of the
Weak Axioms

ECO No. 92/67

Nikolaos GEORGANTZIS
The Effect of Mergers on Potential
Competition under Economies or
Diseconomies of Joint Production

ECO No. 92/68

Robert J. WALDMANN/
J. Bradford DE LONG
Interpreting Procyclical Productivity:
Evidence from a Cross-Nation Cross-
Industry Panel

ECO No. 92/69

Christian DUSTMANN/John
MICKLEWRIGHT
Means-Tested Unemployment Benefit
and Family Labour Supply: A Dynamic
Analysis

ECO No. 92/70

Fabio CANOVA/Bruce E. HANSEN
Are Seasonal Patterns Constant Over
Time? A Test for Seasonal Stability

ECO No. 92/71

Alessandra PELLONI
Long-Run Consequences of Finite
Exchange Rate Bubbles

ECO No. 92/72

Jane MARRINAN
The Effects of Government Spending on
Saving and Investment in an Open
Economy

ECO No. 92/73

Fabio CANOVA and Jane MARRINAN
Profits, Risk and Uncertainty in Foreign
Exchange Markets

ECO No. 92/74

Louis PHILIPS
Basing Point Pricing, Competition and
Market Integration

ECO No. 92/75

Stephen MARTIN
Economic Efficiency and Concentration:
Are Mergers a Fitting Response?

ECO No. 92/76

Luisa ZANCHI
The Inter-Industry Wage Structure:
Empirical Evidence for Germany and a
Comparison With the U.S. and Sweden

ECO NO. 92/77

Agustín MARAVALL
Stochastic Linear Trends: Models and
Estimators

ECO No. 92/78

Fabio CANOVA
Three Tests for the Existence of Cycles
in Time Series

ECO No. 92/79

Peter J. HAMMOND/Jaime SEMPERE
Limits to the Potential Gains from Market
Integration and Other Supply-Side
Policies

ECO No. 92/80

Víctor GÓMEZ and Agustín MARAVALL
Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman Filter

ECO No. 92/81

Víctor GÓMEZ and Agustín MARAVALL
Time Series Regression with ARIMA Noise and Missing Observations
Program TRAM

ECO No. 92/82

J. Bradford DE LONG/ Marco BECHT
"Excess Volatility" and the German Stock Market, 1876-1990

ECO No. 92/83

Alan KIRMAN/Louis PHILIPS
Exchange Rate Pass-Through and Market Structure

ECO No. 92/84

Christian DUSTMANN
Migration, Savings and Uncertainty

ECO No. 92/85

J. Bradford DE LONG
Productivity Growth and Machinery Investment: A Long-Run Look, 1870-1980

ECO NO. 92/86

Robert B. BARSKY and J. Bradford DE LONG
Why Does the Stock Market Fluctuate?

ECO No. 92/87

Anthony B. ATKINSON/John MICKLEWRIGHT
The Distribution of Income in Eastern Europe

ECO No.92/88

Agustín MARAVALL/Alexandre MATHIS
Encompassing Univariate Models in Multivariate Time Series: A Case Study

ECO No. 92/89

Peter J. HAMMOND
Aspects of Rationalizable Behaviour

ECO 92/90

Alan P. KIRMAN/Robert J. WALDMANN
I Quit

ECO No. 92/91

Tilman EHRBECK
Rejecting Rational Expectations in Panel Data: Some New Evidence

ECO No. 92/92

Djordje Suvakovic OLGIN
Simulating Cotermination in a Cooperative Economy

ECO No. 92/93

Djordje Suvakovic OLGIN
On Rational Wage Maximisers

ECO No. 92/94

Christian DUSTMANN
Do We Stay or Not? Return Intentions of Temporary Migrants

ECO No. 92/95

Djordje Suvakovic OLGIN
A Case for a Well-Defined Negative Marxian Exploitation

ECO No. 92/96

Sarah J. JARVIS/John MICKLEWRIGHT
The Targeting of Family Allowance in Hungary

ECO No. 92/97

Agustín MARAVALL/Daniel PEÑA
Missing Observations and Additive Outliers in Time Series Models

ECO No. 92/98

Marco BECHT
Theory and Estimation of Individual and Social Welfare Measures: A Critical Survey

ECO No. 92/99

Louis PHILIPS and Ireneo Miguel MORAS
The AKZO Decision: A Case of Predatory Pricing?

ECO No. 92/100

Stephen MARTIN
Oligopoly Limit Pricing With Firm-Specific Cost Uncertainty

ECO No. 92/101
Fabio CANOVA/Eric GHYSELS
Changes in Seasonal Patterns: Are They
Cyclical?

ECO No. 92/102
Fabio CANOVA
Price Smoothing Policies: A Welfare
Analysis

ECO No. 93/1
Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary
Time Series

ECO No. 93/2
Carlo GRILLENZONI
Multilinear Models for Nonlinear Time
Series



© The Author(s). European University Institute.

Digitised version produced by the EUI Library in 2020. Available Open Access on Cadmus, European University Institute Research Repository.