Journal of Forecasting, Vol. 10, 477–499 (1991)

Simultaneous Transfer Functions versus Vector ARMA Models

CARLO GRILLENZONI
University of Modena and University of Padova, Italy

ABSTRACT

This work compares two classes of multiple time series models which have been developed in past decades and are usually believed to be equivalent: the vector ARMA model and the system of simultaneous transfer functions (STF). The first part analyzes the mathematical structure of the two schemes; their properties of stability, structural identification and realization. In the second, algorithms of order identification and parameter estimation are derived, following the approach of stochastic approximation. The proposed solutions are easily implementable on standard statistical software and in an extended empirical example their performance is checked. The superiority of the STF model will be well established.

KEY WORDS Vector ARMA model Simultaneous transfer functions Iterative pseudolinear regression Box–Jenkins identification techniques

INTRODUCTION

Box and Jenkins (1970) introduced the class of transfer function models (TF) as an open-loop multivariate generalization of the univariate autoregressive moving average (ARMA) scheme. The extension followed a natural approach, by noting that an ARMA process $\{y_t\}$ sets up a dynamic system having a non-observable white noise input $\{e_t\}$ and a rational impulse response function:

$$\text{ARMA} \quad y_t = \psi(B)e_t, \qquad \psi(B) = \frac{\theta(B)}{\phi(B)}, \qquad e_t \sim IN(0, \sigma_e^2)$$

with $\theta(B)$, $\phi(B)$ finite linear polynomials and B the back-shift operator.

Thus, supposing we have an observable input $\{x_t\}$, correlated with the output $\{y_t\}$, the extension has been obtained by adding a rational term as

$$\text{TF} \quad y_t = v(B)x_{t-b} + \psi(B)a_t, \qquad v(B) = \frac{\omega(B)}{\delta(B)}, \qquad a_t \sim \text{IN}(0, \sigma_a^2)$$

where $\omega(B)$, $\delta(B)$ are finite linear polynomials and b the delay factor. As for the univariate

0277–6693/91/050477–23$11.50

Received January 1989
Revised March 1990

models, the identification of the impulse response function $v(B)$ has referred to the second-order properties of the system expressed by the parametric cross-covariance function:

$$\text{CCVF} \quad \gamma_{xy}(B) = v(B)B^b\gamma_{xx}(B), \qquad \gamma_{xx}(B) = \psi_x(B)\psi_x(B^{-1})\sigma_\alpha^2$$

This raises the well-known question of prewhitening, i.e. $v(B)B^b \propto \gamma_{xy}(B)$ only if $\{x_t\}$ is whitened.

Following the work of Box–Jenkins, the study of open-loop transfer function models was continued by Pierce (1972), analyzing the properties of nonlinear estimators and by Box and MacGregor (1974), investigating the consequences of feedback on identification. Solo (1978) and Young and Jakeman (1979) developed a powerful dynamic extension of the framework by means of recursive estimators able to track change of parameters and improving the statistical fitting. More recently, Liu and Hanssens (1982) and Lii (1985) have defined efficient parametric techniques of identification, based on the 'corner method', suitable in case of multiple inputs, and Priestley (1983) has analysed the properties of linear estimators in the presence of autocorrelated errors and feedback. The TF model has found wide application in empirical analysis of causality in economics (see e.g. Maloney and Ireland, 1980; Grillenzoni, 1983b). This experience has provided a positive check on the performance of the method.

As in econometrics, a *natural* extension of the TF model to closed-loop multivariate (m) process $\{\mathbf{z}_t\}$ is yielded by a system of simultaneous transfer functions

$$\text{STF} \quad \mathbf{z}_t = V(B)\mathbf{z}_{t-b} + \ddot{\Psi}(B)\mathbf{a}_t, \qquad \mathbf{a}_t \sim IN_m(\mathbf{0}, \ddot{\Sigma})$$

with $V(B) = \{v_{ij}(B)\}$, $\ddot{\Psi}(B) = \text{Diag}[\psi_i(B)]$ matrices of *rational* polynomials. In this context, however, the identification approach based on the analysis of the second-order moments has encountered a serious limitation in the extreme complexity of the covariance function matrix of the system:

$$\text{CVFM} \quad \Gamma(B) = [I - V(B)]^{-1}\ddot{\Psi}(B)\ddot{\Sigma}\ddot{\Psi}(B^{-1})[I - V(B^{-1})']^{-1}$$

For bivariate processes, however, Haugh and Box (1977), Granger and Newbold (1977) and Jenkins (1979) have implemented efficient 'two-stage' techniques of identification based on a bilateral prewhitening of the series and a joint estimation of the impulse response coefficients form the sample cross-correlation functions.

The difficulty of extending these techniques to general vector processes, their complexity and tendency to overparametrize the models and the absence of packages for the joint estimation of transfer functions equation have halted the study of SFT models, the sole exception being the applied works of Hanssens and Liu (1983) and Liu (1987). The multivariate approach which has prevailed in the 1980s is the vector ARMA (see Tiao and Box, 1981; Jenkins and Alavi, 1981; Tiao and Tsay, 1983). This provides a formal matrix extension of the corresponding univariate scheme, taken in linear form:

$$\text{ARMA}_m \quad \begin{aligned} &\Phi(B)\mathbf{z}_t = \Theta(B)\mathbf{e}_t, \qquad \mathbf{e}_t \sim \text{IN}_m(\mathbf{0}, \Sigma) \\ &\Gamma(B) = \Phi(B)^{-1}\Theta(B)\Sigma\Theta(B^{-1})'\Phi(B^{-1})'^{-1} \end{aligned}$$

where $\Phi(B) = \{\phi_{ij}(B)\}$, $\Theta(B) = \{\theta_{ij}(B)\}$ are matrices of *linear* polynomials. The advantages of the approach lie in the non-specialized character of the vector ARMA_m structure, so that its identification may be developed, in an *aggregate* manner, with scalar information criteria (FPE, AIC, BIC, CAT, MDL, etc.).

Apart from the works of Zellner and Palm (1974) and Liu and Hudak (1985), in the literature very little has been done to compare the structures of STF-ARMA_m models, and many authors have simply stated that the two classes are equivalent. Now, although it is

formally possible to linearize the SFT model by row (obtaining a constrained $ARMA_m$ with $\Theta(B)$ diagonal), and although we can put the $ARMA_m$ in rational autoregressive form (obtaining a constrained STF with common denominator $\mathrm{Det}\Theta(B)$). In substantial terms the STF scheme seems preferable to the $ARMA_m$ for the following reasons:

(1) The structure with *rational* polynomials is more powerful and parsimonious than the linear one;
(2) It has an *unconstrained* covariance function matrix $\Gamma(B)$, having polynomials of different order;
(3) It possesses a *complete* realization framework from the multivariate spectral factorization theorem;
(4) It respects the *different* nature of auto and cross-dynamic regressions, treating each with two specific filters $\ddot{\Psi}(B)$, $V(B)$;
(5) The matrix $V(B)$ is directly *interpretable* in terms of causal structure of the process $\{z_t\}$;
(6) The univariate and orthogonal ($\ddot{\Sigma}$ = Diag) residuals enables a *simplified* estimation by row,
(7) Potentially, it may be identified in a *disaggregate* manner by means of the Box–Jenkins techniques.

From these descriptive notes, it is apparent that the STF class is more general and flexible than the $ARMA_m$ one, and it tends to respect substantial rules of multivariate statistical analysis. In what follows these considerations will be analytically and empirically developed.

MULTIVARIABLE SYSTEMS

In this section we analyze the structure and the properties of the system of simultaneous transfer functions, comparisons with the vector $ARMA_m$ model will be given. We consider a vector stochastic process $\mathbf{z}_t' = [z_{1t}, z_{2t}, \dots z_{mt}]$, zero mean Gaussian, stationary in covariance and mean square summable (ergodic):

$$\mathbf{z}_t \sim N_m[\mathbf{0}, \Gamma_k], \qquad \Gamma_k = E[\mathbf{z}_t\mathbf{z}_{t-k}'] = \{\gamma_{ij}(k)\}, \qquad \sum_k \| \Gamma_k \| < \infty$$

By the linearity of the regression (conditional mean) under gaussianity, these distributive hypotheses are clearly equivalent to the SFT representation

$$\mathbf{z}_t = \sum_{i=0}^{\infty} V_i\mathbf{z}_{t-b-i} + \sum_{j=1}^{\infty} \ddot{\Psi}_j\mathbf{a}_{t-j} + \mathbf{a}_t, \qquad \mathbf{a}_t \sim \mathrm{IN}_m[\mathbf{0}, \ddot{\Sigma} = \mathrm{Diag}(\sigma_i^2)]$$

$$[I - V(B)]\mathbf{z}_t = \ddot{\Psi}(B)\mathbf{a}_t; \qquad V(B) = \left\{\frac{\omega_{ij}(B)B^{b_{ij}}}{\delta_{ij}(B)}\right\}, \qquad \ddot{\Psi}(B) = \mathrm{Diag}\left[\frac{\theta_i(B)}{\phi_i(B)}\right]$$

where $(\delta_{ij}, \omega_{ij}, \phi_i, \theta_i)$ are linear polynomials with degree, respectively $(r_{ij}, s_{ij}, p_i, q_i)$, finite; b_{ij} is the delay factor of $(z_{jt} \to z_{it})$, $b = \min(b_{ij})$. Moreover $V(B)$ has zeros on the principal diagonal and $\ddot{\Psi}(B)$, $\ddot{\Sigma}$ are diagonal matrices. In the following subsections we shall investigate some analytical properties of the above representation.

Invertibility and stationarity

The conditions of stationarity relate to the possibility of representing the system in infinite MA_m form (multivariate Wold decomposition). Having

$$\mathbf{z}_t = [I - V(B)]^{-1}\ddot{\Psi}(B)\mathbf{a}_t = \Psi(B)\mathbf{a}_t, \qquad \Psi(B) = \{\psi_{ij}(B)\}$$

from the definition of determinant as a sum of products and taking the least common denominator $\delta^*(B)$ we may set $\mathrm{Det}[I - V(B)] = \omega^*(B)/\delta^*(B)$. Hence

$$\psi_{ij}(B) = \tilde{\theta}_{ij}(B)\delta^*(B)/\phi_i(B)\omega^*(B) \prod_{h \neq i}^{m} \prod_{l \neq j}^{m} \delta_{hl}(B)$$

The terms $\tilde{\theta}_{ij}(B)$, which result by taking the common denominator in every cell of the adjoint matrix $[I - V(B)]^*$, have quite complicated expressions, and practically they are functions of all ω_{ij}, δ_{ij}.

The matrix $\Psi(B)$ can then be expanded in power series, only if the polynomials ϕ_i, ω^*, $\delta_{ij} \forall ij$ have roots outside the unit circle (*stability*). The sequence $\{\Psi_k\}$ can also be obtained in an aggregate manner by noting that $[I - V(B)]\Psi(B) = \ddot{\Psi}(B)$ and equating products of matrices corresponding to the same powers:

$$\Psi_k = (I - V_0)^{-1}\left[\sum_{j=1}^{k} V_j \Psi_{k-j} + \ddot{\Psi}_k\right]$$

Here, it is clear that a *necessary* condition for the convergence of $\{\Psi_k\}$ is the convergence of $\{V_k\}$, i.e. the stability of each $\delta_{ij}(B)$.

The conditions of invertibility deal with the possibility of expressing the system in infinite AR_m form. Setting $\pi_{ij}(B) = \phi_i(B)\omega_{ij}(B)/\theta_i(B)\delta_{ij}(B)$, we have

$$\ddot{\Psi}(B)^{-1}[I - V(B)]\mathbf{z}_t = \Pi(B)\mathbf{z}_t = \mathbf{a}_t, \qquad \Pi(B) = \{\pi_{ij}(B)\}$$

Thus $\Pi(B)$ can be expanded in power series only if the polynomials θ_i, δ_{ij} and $\omega_{ij} \forall ij$ are, respectively, stable and bounded on $|B| = 1$, the algorithm of calculation being

$$\Pi_k = \sum_{i=1}^{k} \ddot{\Psi}_i \Pi_{k-i} - V_k$$

In summary, the *general* conditions of stability of the STF model are given by

$$\textit{Invertibility} \quad [\delta_{ij}(B), \theta_i(B)] \neq 0, \ |\omega_{ij}(B)| < \infty, \qquad |B| \leqslant 1 \tag{1a}$$

$$\textit{Stationarity} \quad [\delta_{ij}(B), \phi_i(B), \omega^*(B)] \neq 0, \qquad |B| < 1 \tag{1b}$$

Unlike the $ARMA_m$ model, it is interesting to note that the two properties are *interdependent* since they require the same condition on the polynomials $\delta_{ij}(B)$. Moreover, although $\delta^*(B)$ is 'automatically' stable because $\delta^*(B) = \Pi_i^m \Pi_j^m \delta_{ij}(B)$ (assuming the $\delta_{ij}(B)$ are *relatively prime*, i.e. without common linear factors), in general, it is not possible to establish conditions on the $\omega_{ij}(B)$ to raise the stability of $\omega^*(B)$. This situation is due to the fact that $\omega^*(B)$ is a complex sum of products of polynomials, and that a sum of stable polynomials is, in general, still stable only if its elements are, at most, of second order (see Appendix 1).

The definition of representations MA_m, AR_m infinite and of the related properties of stationarity and invertibility tries not only to satisfy formal requirements. As we shall see, the first representation has a role in forecasting since the covariance of the l-step-ahead prediction error utilizes the Ψ_k weights; the second, by expressing $\mathbf{a}_t$ (non-observable) as a combination of $\mathbf{z}_t$ is crucial in estimation since the gradient is a linear function of $\mathbf{a}_t$.

Identification and realization

Using the moving-average representation we can obtain the expression of the spectral density of $\{\mathbf{z}_t\}$ in terms of the STF parameters. Let $\lambda \in [-\pi, +\pi]$. Then

$$\Gamma(e^{-i\lambda}) = \Psi(e^{-i\lambda})\ddot{\Sigma}\Psi(e^{+i\lambda})', \qquad \Psi(e^{-i\lambda}) = [I - V(e^{-i\lambda})]^{-1}\ddot{\Psi}(e^{-i\lambda}) \tag{2}$$

Now, since the estimators of the system are functions of the second-order moments of the process it follows that the STF scheme is *identified* only if the factorization (2) is unique. Denoting the complex variable $z = e^{-i\lambda}$, in Appendix 2 we show that conditions to this end are provided by:

(1) The polynomials $[\delta_{ij}(z), \omega_{ij}(z)]$, $[\phi_i(z), \theta_i(z)]$ have no common factor;
(2) The properties of invertibility and stationarity (1) hold;
(3) $[\delta_{ij}(0) = \phi_i(0) = \theta_i(0) = 1]$, $[\omega_{ij}(0) = \omega_{ji}(0)]$, i.e. $V_0 = V_0'$, $\ddot{\Sigma} = \text{Diag}$, $\ddot{\Psi}_0 = I$.

The question of *realization* of system structures arises as an inverse problem to that of representing the spectrum. Given the parametric expression of the spectral density $\Gamma(z)$, a factorization must be defined which yields a transfer function $\Psi(z)$ (unique and minimal), to be used for representing the process $\{\mathbf{z}_t\}$ in the time domain. A fairly general starting point is provided by the spectral factorization theorem of Rozanov (1967, p. 47), extended by Hannan (1979, p. 85).

THEOREM
Let $\Gamma(e^{-i\lambda})$, $\lambda \in [-\pi, +\pi]$ *a square matrix, hermitian, non-negative definite, rational and integrable. Then a square matrix* $\Psi(z)$, *rational, non-singular, analytic in* $|z| \leqslant 1$, *exists such that:* $\Gamma(e^{-i\lambda}) = \Psi(e^{-i\lambda})\Psi(e^{+i\lambda})'$. *The factorization is unique if* $\Psi(z)^{-1}$ *is analytic in* $|z| \leqslant 1$ *and* $\Psi(0) = \Psi(0)'$.

The proof of the theorem is achieved by construction and so defines an algorithm which has practical relevance. The way of realizing the *further* factorization of $\Psi(z)$ is what specifically characterizes the type of system representation to be reached. This problem has received a systematical treatment only in the *state-space* context, with the algorithms of Kalman, Ho-Silverman and Rissanen (see Faurre *et al.*, 1979, for a good survey).

In the *input–output* context the realization of STF filters is immediate; setting $\Psi(z)^{-1} = [\ddot{A}(z) + C(z)]$, where $\ddot{A}(z)$ is the principal diagonal and $\ddot{A}(0) = I$, $C(0) = C(0)'$, we obtain two solutions:

$$[\ddot{\Psi}(z) = \ddot{A}(z)^{-1}, V(z) = -\ddot{A}(z)^{-1}C(z)], \qquad [\ddot{\Pi}(z) = \ddot{A}(z), W(z) = -C(z)\ddot{A}(z)^{-1}]$$

From the above we may then conceive the existence of two *equivalent* STF representations, yielding two kinds of uncorrelated processes: $\{\mathbf{n}_t\}$. $\{\mathbf{u}_t\}$:

$$\text{STF(1)}\begin{cases}[I - V(B)]\,\mathbf{z}_t = \mathbf{n}_t \\ \mathbf{n}_t = \ddot{\Psi}(B)\mathbf{a}_t\end{cases} \qquad \text{STF(2)}\begin{cases}\ddot{\Pi}(B)\mathbf{z}_t = \mathbf{u}_t \\ [I - W(B)]\,\mathbf{u}_t = \mathbf{a}_t\end{cases}$$

where cross-covariance $\text{CCV}(\mathbf{n}_t) = 0$ and auto-covariance $\text{ACV}(\mathbf{u}_t) = 0$. The latter is the version with prewhitening, in which $\ddot{\Pi}(z)$ is the diagonal matrix of univariate ARMA filters.

For the vector ARMA model the matter is not so simple; the sole possible further factorization of $\Psi(z)$ is very constrained. Indeed, if we set $\Psi(z) = \Phi(z)^{-1}\Theta(z)$, $\psi_{ij}(z) = \alpha_{ij}(z)/\beta_{ij}(z)$ we may only get

$$\ddot{\Phi}(z) = \text{Diag}\left[\prod_{j=1}^{m} \beta_{1j}(z), \ldots, \prod_{j=1}^{m} \beta_{mj}(z)\right], \qquad \tilde{\Theta}(z) = \ddot{\Phi}(z)^{-1}\Psi(z)$$

which corresponds to a linearization by row of the rational MA representation $\mathbf{z}_t = \Psi(B)\mathbf{a}_t$, assuming $[\beta_{i1}(z), \ldots, \beta_{im}(z)]$ relatively prime $\forall i$. Alternatively, if we aim to realize a general unconstrained, $\Phi \neq \text{Diag}$, ARMA_m structure, we must have as necessary one of the conditions: (1) $\beta_{ij}(z) = \beta(z)$, $\beta(z) = \text{Det}\Phi(z) \forall ij$, (2) $[\alpha_{ij}(z), \beta_{ij}(z)]$ *having common factors*. That is, either we must assume a stochastic process with constrained spectral density matrix (having functions

$\gamma_{ij}(z)$ with common denominator $|\beta(z)|^2$), or go on to define (with (2)) a non-identified (and non-minimal) spectral factorization.

In any case, the fact that only constrained matrices are realizable has the advantage of simplifying the conditions of structural identification. Indeed, since models with one of $\Phi(z)$, $\Theta(z)$ diagonal represent *canonical* forms, from the results of Appendix 2 it is easy to assess that additional restrictions, such as $\text{Rank}[\Phi_p : \Theta_q] = m$ (see Hannan, 1969), are no longer necessary. This seems particularly important in the case of specialized ARMA_m models with matrices of orders $P = \{p_{ij}\}$, $Q = \{q_{ij}\}$ having unique maxima, because the rank condition could never be fulfilled for $m > 2$.

In summary, we may conclude that the STF model, since unconstrained in rational form and completely realizable, includes as a particular case the ARMA_m one, and is a candidate to represent a more general class of processes.

APPROXIMATE ALGORITHMS

In this section we derive algorithms of identification and estimation that refer in general to principles of *stochastic approximation*. Given the complexity of the STF model, the non-rigorous approach of the approximation seems unavoidable. The proposed solutions, however, have the advantages of simplicity and flexibility. They recall well-known procedures and are easily implementable on standard software.

Disaggregate identification

Here, we provide elements that heuristically justify the use of Box–Jenkins techniques in the identification of multivariable transfer functions. Consider the ith row of the system $[I - V(B)]\mathbf{z}_t = \mathbf{n}_t$:

$$z_{i_t} = \sum_{j \neq i}^{m} v_{ij}(B) z_{j_{t-b_{ij}}} + n_{i_t} \qquad i = 1, 2, \ldots, m$$

By multiplying the above expression with $z_{j_{t-k}} \forall j, k$, taking expectation and considering the generating function, we obtain the system of deterministic equations:

$$\begin{bmatrix} \gamma_{i1}(B) \\ \gamma_{i2}(B) \\ \vdots \\ \gamma_{im}(B) \end{bmatrix} = \begin{bmatrix} \gamma_{11}(B) & \gamma_{12}(B) & \cdots & \gamma_{1m}(B) \\ \gamma_{21}(B) & \cdot & \cdots & \cdot \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{m1}(B) & \cdot & \cdots & \gamma_{mm}(B) \end{bmatrix} \begin{bmatrix} v_{i1}(B) \\ v_{i2}(B) \\ \vdots \\ v_{im}(B) \end{bmatrix}$$

i.e.

$$\boldsymbol{\gamma}_i(B) = \Gamma_{ii}(B)\mathbf{v}_i(B), \qquad i = 1, 2, \ldots, m$$

Now, solving for $\mathbf{v}_i(B)$ by extending the arguments of stochastic approximation to algorithms of linear estimation (see Tsypkin, 1971, p. 65), that is, by approximating $\Gamma_{ii}(B)$ to a diagonal matrix, we have $v_{ij}(B) = \gamma_{ij}(B)/\gamma_{ii}(B)$, $\forall ij$. Moreover, assuming that $\text{ACV}(\mathbf{n}_t) \approx \text{ACV}(\mathbf{z}_t)$ it follows $\gamma_{ii}(B) \approx \psi_i(B)\psi_i(B^{-1})\sigma^2$, $\forall i$. Summarizing these results in matrix terms, we may have the factorization

$$\Gamma_z(B) \approx [I + V(B)]\Gamma_n(B), \qquad \Gamma_n(B) \approx \ddot{\Psi}(B)\ddot{\Sigma}\ddot{\Psi}(B^{-1})$$

This, although approximate, enables a *disaggregate* identification of the STF model by straightforward application of the univariate-unidirectional Box–Jenkins schemes. Note, in

particular, that by using the STF(2) form, i.e. with prewhitening, we *roughly* get

$$\Gamma_u(B) \propto [I + V(B)] \tag{3}$$

Hence the orders (r_{ij}, s_{ij}, b_{ij}) of each impulse response function $v_{ij}(B)$ can be derived from the corresponding sample cross-correlation function $r_{u_i u_j}(k)$.

As an interpretation, the validity of the proposed procedure requires not so much that the process $\{\mathbf{z}_t\}$ be weakly correlated, but more loosely that the CCVF$_s$ $\gamma_{ij}(B)$ be small with respect to ACVF$_s$ $\gamma_{ii}(B)$, and in general that the indirect interactions between the univariate processes $\{z_{i_t}, z_{j_t}\}$, considered by pairs (i.e. between the functions $\gamma_{ij}(B)$), be low. This interpretation clearly relates to concepts like principal components and partial correlation. Here, we recall that for stochastic processes, the second has been well investigated only for the autocorrelation. In open-loop systems the *partial cross-correlation function* may simply be defined as *proportional* to the marginal regression coefficients of linear systems:

$$\text{PCCRF} \quad \{v_{kk}\} \in y_t = \sum_{j=0}^{k} v_{jk} x_{t-k} + n_t$$

This, however, has never been required in the identification of $v(B)$. Moreover, if $\{x_t\}$ is white noise, it is easy to assess that $v_{kk} \propto \rho_{xy}(k)$, i.e. cross-correlation and partial cross-correlation coincide. The approximate expression (3) seems then a *natural* multivariate extension of this equivalence.

In time series the problem of spurious correlation (indirect causality) may sensibly be raised only at the level of simultaneous correlations $\rho_{ij}(0)$, but not for the lagged ones, i.e. for the whole CCRF $\rho_{ij}(k)$. On the other hand, about the approximation $\text{ACV}(\mathbf{n}_t) \approx \text{ACV}(\mathbf{z}_t)$, we note that the dumping action on the CCR of a prewhitening of the ACR is sure because $\text{Var}[r_{ij}(k)] = f[\rho_{ij}(h)\rho_{ii}(l)]$, but the effects of the reverse filtering are uncertain since $\text{Var}[r_{ii}(k)] = f[\rho_{ii}(h)\rho_{ii}(l)]$ (for details see Box and Jenkins, 1970). As a consequence, it seems that for stationary processes, summable in covariance (i.e. with restricted memory), the hypothesis of low indirect interactions of the pairs $\{z_{i_t}, z_{j_t}\}$, and thus of validity of the proposed method, be acceptable. Finally, establishing approximations in the specification phase does not create a serious problem in its own right. In fact, in this early step, only a *general* idea of the dynamic inside the system is required. It will be the subsequent phases of estimation and checking that must refine the first guess. The procedure of identification we have defined, however, has the advantages of simplicity and disaggregation.

As regards the identification of the vector $\text{ARMA}_m(p, q)$ model, we recall that the multivariate extension it provides seems to be an exercise of matrix algebra. This means that autocorrelation and cross-correlation are treated in the same way and therefore, by the *unspecialized* structure of the model, the second becomes a trivial projection of the first. Clearly, it is true that autorelationships are more powerful and significant than cross ones. However, cross-correlation has a *different nature* ($\rho_{ij}(0) \neq 1$, $\rho_{ij}(k) \neq \rho_{ij}(-k)$, $b > 1$, CCR $\approx$ PCCR), for which Box–Jenkins have set up autonomous apparatus of identification and modeling (impulse response functions). Following these considerations a coherent strategy of identification would seem to be

$$p = \max_i(p_i), \qquad q = \max_i(q_i)$$

where (p_i, q_i) are the orders of the univariate ARMA models of the series z_{i_t}.

Pseudolinear estimation

One of the reasons the applied research on the STF model has been interrupted was the absence of packages for the simultaneous estimation of rational equations. In this section we derive

an algorithm for this purpose, which extends to the *iterative-multivariate* level the *recursive-univariate* techniques called pseudo-linear regression (PLR; see Solo, 1978 and Ljung and Söderström, 1983, for the ARMAX case). The method of estimation is fundamentally that of non-linear least squares (NLS), and is characterized by approximating the gradient with the input–output quantities of the system. It also has important connections with the instrumental variables method (IVM) of Young and Jakeman (1979) which can highlight the meaning and the properties of the algorithm.

Given the univariate-orthogonal structure of residuals, the STF model may be *initially* estimated, without loss of efficiency and consistency, by rows with NLS techniques. For the ith equation, assuming common orders (r, s, b, p, q) to simplify the notation, we have

$$\text{NLS} \quad \hat{\boldsymbol{\beta}}_i(k+1) = \hat{\boldsymbol{\beta}}_i(k) + \left[\sum_{t=1}^{N} \hat{\boldsymbol{\xi}}_{i_t}(k)\hat{\boldsymbol{\xi}}_{i_t}'(k)\right]^{-1} \sum_{t=1}^{N} \hat{\boldsymbol{\xi}}_{i_t}(k)\hat{a}_{i_t}(k)$$

$$\boldsymbol{\xi}_{i_t}(\boldsymbol{\beta}_i) = \left\{-\frac{\partial a_{i_t}(\boldsymbol{\beta}_i)}{\partial \boldsymbol{\beta}_i}\right\}, \qquad a_{i_t}(\boldsymbol{\beta}) = \pi_i(B)\left[z_{i_t} - \sum_{j \neq i}^{m} v_{ij}(B) z_{j_{t-b}}\right] \tag{4}$$

$$\boldsymbol{\beta}_i' = [\delta_{i_1}^1, \ldots, \delta_{i_1}^r, \ \omega_{i_1}^0, \ldots, \omega_{i_1}^s, \ldots, \ldots, \ldots, \omega_{i_m}^s, \phi_i^1, \ldots, \phi_i^p, \theta_i^1, \ldots, \theta_i^q]$$

with $\pi_i(B) = 1/\psi_i(B)$. To derive a useful expression of the gradient we define the auxiliary variables $w_{ij_t} = v_{ij}(B)z_{j_{t-b}}$, $n_{i_t} = \psi_i(B)a_{i_t}$. Hence standard calculus shows that

$$\boldsymbol{\xi}_{i_t}(\boldsymbol{\beta}_i) \begin{cases} -\partial a_{i_t}(\boldsymbol{\beta}_i)/\partial \delta_{ij}^h = [\pi_i(B)/\delta_{ij}(B)]\, w_{ij_{t-h}} \\ -\partial a_{i_t}(\boldsymbol{\beta}_i)/\partial \omega_{ij}^h = [\pi_i(B)/\delta_{ij}(B)]\, z_{j_{t-b-h}} \\ -\partial a_{i_t}(\boldsymbol{\beta}_i)/\partial \phi_i^h = [1/\theta_i(B)]\, n_{i_{t-h}} \\ -\partial a_{i_t}(\boldsymbol{\beta}_i)/\partial \theta_i^h = [1/\theta_i(B)]\, a_{i_{t-h}} \end{cases} \qquad h = 0, 1, 2 \ldots r, s, p, q$$

The computation of the gradient thus consists in a filtering operation on *observable* (z), *auxiliary* (w, n) and *non-observable* (a) quantities, by means of the same filters of the system. The calculation of the ith residual is carried out in three steps

$$\text{(i)} \quad w_{ij_t} = \sum_{h=1}^{r} \delta_{ij}^h w_{ij_{t-h}} + \sum_{h=0}^{s} \omega_{ij}^h z_{j_{t-b-h}}$$

$$\text{(ii)} \quad n_{i_t} = z_{i_t} - \sum_{j \neq i}^{m} w_{ij_t}$$

$$\text{(iii)} \quad a_{i_t} = n_{i_t} - \sum_{h=1}^{p} \phi_i^h n_{i_{t-h}} - \sum_{h=1}^{q} \theta_i^h a_{i_{t-h}}$$

Recomposing the three steps, we may re-write the model in pseudolinear form as

$$z_{i_t} = \sum_{j \neq i}^{m} (\boldsymbol{\delta}_{ij}' \mathbf{w}_{ij_{t-1}} + \boldsymbol{\omega}_{ij}' \mathbf{z}_{j_{t-b}}) + (\boldsymbol{\phi}_i' \mathbf{n}_{i_{t-1}} + \boldsymbol{\theta}_i' \mathbf{a}_{i_{t-1}}) + a_{i_t}$$

where

$$\mathbf{x}_{i_t}(\boldsymbol{\beta}_i) \begin{cases} \mathbf{w}_{ij_{t-1}} = [w_{ij_{t-1}}, \ldots, w_{ij_{t-r}}]' & [\delta_{ij}^1, \ldots, \delta_{ij}^r]' = \boldsymbol{\delta}_{ij} \\ \mathbf{z}_{j_{t-b}} = [z_{j_{t-b}}, \ldots, z_{j_{t-b-s}}]' & [\omega_{ij}^0, \ldots, \omega_{ij}^s]' = \boldsymbol{\omega}_{ij} \\ \mathbf{n}_{i_{t-1}} = [n_{i_{t-1}}, \ldots, n_{i_{t-p}}]' & [\phi_i^1, \ldots, \phi_i^p]' = \boldsymbol{\phi}_i \\ \mathbf{a}_{i_{t-1}} = [a_{i_{t-1}}, \ldots, a_{i_{t-q}}]' & [\theta_i^1, \ldots, \theta_i^q]' = \boldsymbol{\theta}_i \end{cases} \boldsymbol{\beta}_i$$

Hence

$$z_{i_t} = \boldsymbol{\beta}_i' \mathbf{x}_{i_t}(\boldsymbol{\beta}_i) + a_{i_t} \tag{5}$$

where $\mathbf{x}_{i_t}(\boldsymbol{\beta}_i)$ is the vector of pseudolinear regressors, depending on $\boldsymbol{\beta}_i$.

Now, the pseudolinear estimator may formally arise from the NLS one by approximating $\hat{\boldsymbol{\xi}}_{i_t}(k) \approx \hat{\mathbf{x}}_{i_t}(k)$ (i.e. by avoiding the filtering with $\pi_i(B)/\delta_{ij}(B)$ and $1/\theta_i(B)$). Indeed, since $\hat{a}_{i_t}(k) = z_{i_t} - \hat{\mathbf{x}}_{i_t}(k)'\hat{\boldsymbol{\beta}}_i(k)$, substituting these quantities in equation (4) the PLR algorithm takes on the compact form:

$$\text{PLR} \quad \hat{\boldsymbol{\beta}}_i(k+1) = \left[\sum_{t=1}^{N} \hat{\mathbf{x}}_{i_t}(k)\hat{\mathbf{x}}_{i_t}'(k)\right]^{-1} \sum_{t=1}^{N} \hat{\mathbf{x}}_{i_t}(k) z_{i_t} \tag{6}$$

This approach of derivation has important connections with the method of instrumental variables developed by Young and Jakeman (1979) for TF models. To show this in detail we assume, for simplicity, a STF system with $m = 2$ and we consider the estimation of the first equation:

$$z_{1_t} = \frac{(\omega_{12}^0 + \omega_{12}^1 B + \cdots + \omega_{12}^s B^s)}{(1 - \delta_{12}^1 B - \cdots - \delta_{12}^r B^r)} z_{2_{t-b}} + \frac{(1 + \theta_1^1 B + \cdots + \theta_1^q B^q)}{(1 - \phi_1^1 B - \cdots - \phi_1^p B^p)} a_{1_t}$$

As stated before, the assumption $E[a_{1_t} a_{2_{t \pm k}}] = 0 \forall k$ guarantees no loss of optimality. Now, the question originally posed by Young and Jakeman, in recursive terms, was that of deriving simplified, but consistent, estimators of the parameters $\mathbf{v}_{12}' = [\delta_{12}^1 \ldots \delta_{12}^r, \omega_{12}^0 \ldots \omega_{12}^s]$, $\boldsymbol{\psi}_1' = [\phi_1^1 \ldots \phi_1^p, \theta_1^1 \ldots \theta_1^q]$. First, the analytical result of Pierce (1972) on the asymptotic independence of the NLS estimates $\hat{\mathbf{v}}_{12}$, $\hat{\boldsymbol{\psi}}_1$ enables us to set up the solution in sequential form; i.e. estimation of $\mathbf{v}_{12}$ from the system $z_{1_t} = v_{12}(B) z_{2_{t-b}} + n_{1_t}$ and then estimation of $\boldsymbol{\psi}_1$ from the realization $\hat{n}_{1_t} = z_{1_t} - \hat{v}_{12}(B) z_{2_{t-b}}$.

About the first step, Young and Jakeman (1979) argue that the OLS estimator of $\mathbf{v}_{12}$, implemented on the linearized equation $\delta_{12}(B) z_{1_t} = \omega_{12}(B) z_{2_{t-b}} + \eta_{1_t}$, with $\eta_{1_t} = \delta_{12}(B) n_{1_t}$, is not consistent due to the high autocorrelation of $\{\eta_{1_t}\}$. However, this drawback can be avoided by 'flanking' a set of instrumental variables to the vector of regressors $\mathbf{z}_{12_t}' = [z_{1_{t-1}} \ldots z_{1_{t-r}}, z_{2_{t-b}} \ldots z_{2_{t-b-s}}]$. Now, the most convenient set of instruments is certainly represented by the auxiliary sequence $w_{12_t} = v_{12}(B) z_{2_{t-b}}$, which can be generated in an adaptive manner through the equation $\delta_{12}(B) w_{12_t} = \omega_{12}(B) z_{2_{t-b}}$. Note that $\{w_{12_t}\}$ were already pseudolinear regressors in equation (6) but also are instrumental variables since they are correlated with $\{z_{2_t}\}$ and independent of $\{a_{1_t}\}$. Thus, by defining $\mathbf{x}_{12_t}' = [w_{12_{t-1}} \ldots w_{12_{t-r}}, z_{2_{t-b}} \ldots z_{2_{t-b-s}}]$ the IV estimator becomes

$$\text{IVM} \quad \hat{\mathbf{v}}_{12}(k+1) = \left[\sum_{t=1}^{N} \hat{\mathbf{x}}_{12_t}(k) \mathbf{z}_{12_t}'\right]^{-1} \sum_{t=1}^{N} \hat{\mathbf{x}}_{12_t}(k) z_{1_t} \tag{7}$$

The resemblance with equation (6) is straightforward. Moreover, about the estimation of $\boldsymbol{\psi}_1$, Young and Jakeman (1979) propose the method of approximate maximum likelihood (AML), which, in practice, is a PLR scheme with the vector of 'regressors' $\mathbf{x}_{1_t}' = [n_{1_{t-1}} \ldots n_{1_{t-p}}, a_{1_{t-1}} \ldots a_{1_{t-q}}]$ and the 'dependent' variable n_{1_t}.

The IV method can be improved by means of refined instrumental variables (RIV). To introduce them we re-write the previous TF model as

$$a_{1_t} = \frac{\pi_1(B)}{\delta_{12}(B)} [\delta_{12}(B) z_{1_t} - \omega_{12}(B) z_{2_{t-b}}], \qquad \pi_1(B) = \frac{\phi_1(B)}{\theta_1(B)}$$

Now, setting $z_{1_t}^* = (\pi_1/\delta_{12}) z_{1_t}$, $z_{2_t}^* = (\pi_1/\delta_{12}) z_{2_t}$, the RIV estimator of $\mathbf{v}_{12}$ may again be obtained

by applying the IV method to the linear equation $\delta_{12}(B)z^*_{1,t} = \omega_{12}(B)z^*_{2,t-b} + a_{1,t}$. In practice, the improved algorithm substitutes in equation (7) the refined instruments $\mathbf{x}^*_{12,t} = (\pi_1/\delta_{12})\mathbf{x}_{12,t}$ and the filtered regressors $\mathbf{z}^*_{12,t} = (\pi_1/\delta_{12})\mathbf{z}_{12,t}$, obviously in iterative form. In this case, the estimates $\hat{\mathbf{v}}_{12}(k)$, $\hat{\psi}_1(k)$ are no more computationally independent, but the gain of statistical efficiency obtained in passing from IV to RIV is almost the same as that lost in approximating NLS with PLR. Note. in fact, that the instrument $\mathbf{x}^*_{12,t} = \boldsymbol{\xi}_{12,t} = -\partial a_{1,t}/\partial \mathbf{v}_{12}$ coincides with the gradient of the estimator (4).

Although this example is difficult to extend to closed-loop STF with $m > 2$, it provides important elements of interpretation and legitimation for the PLR method. Some words of care are, however, in order. The pseudolinear regressors $\{w_{ij,t-k}\}\,\forall ijk$ do coincide with instrumental variables, but it seems that only in algorithm (7) do they play this role. In practice, the IV method allows for the separation of estimates $\hat{\mathbf{v}}_{12}$, $\hat{\psi}_1$ without affecting consistency, while in the PLR context this holds asymptotically and under certain 'passivity' conditions of the system (see below). On the other hand, the instrumental nature of $w_{12,t}$ is guaranteed only if $z_{2,t}$ is well correlated with $z_{1,t}$; a feature that in real data may be difficult to have (see the next section).

To complete the comparison of PLR-NLS and IV-RIV methods we briefly consider their role in identification strategies of *parametric* type, i.e. based on the optimization of some criterion function related to the models. The first technique, concerned with IV-RIV and developed by Young *et al.* (1980) and Young (1989), combines two negatively related test statistics: the coefficient of determination $R^2_{zn} = 1 - \hat{\sigma}^2_n/\hat{\sigma}^2_z$ and the error variance norm (EVN):

$$\mathrm{YIC}(r, s \mid b) = \ln(1 - R^2_{zn}) + \ln\left(\frac{\mathrm{EVN}}{\hat{\mathbf{v}}'\hat{\mathbf{v}}}\right), \qquad \mathrm{EVN} = \frac{1}{r+s+1}\,\mathrm{tr}\left[\sum_{t=b}^{N} \hat{\mathbf{x}}_t \mathbf{z}_t'\right]^{-1} \hat{\sigma}^2_n$$

The last provides the average variance of parameters and appears to be a sensitive indicator of overparameterization. Indeed, if a model has too many parameters, then the instrumental product matrix $[\Sigma_t \hat{\mathbf{x}}_t \mathbf{z}_t']$ tends to singularity and so EVN increases sharply. Now, under stability conditions, for a given $b \geqslant 0$, we have $R^2 \to 1$ as $r, s \to N$, but this, in turn, implies $\mathrm{EVN} \to \infty$. By taking logarithm the Young information criterion (YIC) then provides (in its minimum) a good compromise between model fit and parametric efficiency.

The second method, recently proposed by Poskitt (1989), is concerned with NLS (Gauss–Newton) and minimizes the well-known criterion

$$\mathrm{BIC}(r, s, p, q \mid b) = \ln\,\hat{\sigma}^2_a + (r + s + 1 + p + q)\,\frac{\ln(N-b)}{(N-b)}$$

with $(r, s, p, q) < \ln(N-b)$. With respect to YIC, BIC has a well-defined global minimum and consistency properties, but it is computationally expensive. In any case, both these methods heavily rely on a hypothesis that a *true* TF system exists. In contrast, real data are often generated by irregular operators $\omega(B)^* = (\omega_0 + \omega_1 B^j + \omega_2 B^{h+k} \ldots)$, periodic filters $\phi(B^k) = (1 - \phi_1 B^k - \phi_2 2^k - \cdots)$ and, in general, by multivariate systems with sparse coefficients at 'strange' lags. In these cases, *non-parametric* techniques of identification based on the inspection of sample correlation functions (as that suggested in the previous subsection) are preferable.

Heuristically, algorithm (6) may also be derived by applying iteratively OLS to the pseudolinear model (5). The crucial step, however, is made by the approximation $\boldsymbol{\xi}_{i,t} \approx \mathbf{x}_{i,t}$ in equation (4). It is the goodness of this approximation that actually influences the statistical properties of the PLR method. Now, since the NLS estimator enjoys optimal properties (convergence in mean square), and since the filtering to calculate the gradient is characterized

by the rational polynomials $1/\theta_i(z)$, $\pi_i(z)/\delta_{ij}(z)$, the statistical behaviour of equation (6) must relate to mathematical properties of the two filters. Thus, generalizing the analysis developed by Ljung and Soderstrom (1983, Chapter 4), for the *recursive* PLR estimation of ARMAX models, we may state that the *iterative* version we have provided is strongly consistent if the monic polynomials of the system behave like *passive* filters:

$$\mathrm{Re}\left[\frac{\phi_i(e^{i\lambda})}{\theta_i(e^{i\lambda})\delta_{ij}(e^{i\lambda})}-\frac{1}{2}\right]>0, \qquad \mathrm{Re}\left[\frac{1}{\theta_i(e^{i\lambda})}-\frac{1}{2}\right]>0 \qquad \forall\lambda\forall_{ij}$$

As Hannan and McDougall (1988) have proved, this extension is admissible since recursive estimators arise from simple algebraic transformations of the corresponding iterative ones (see Ljung, 1985). Hence, for stationary, time-invariable models the two versions are asymptotically equivalent.

Finally, in presence of simultaneous correlation and assuming $V_0 = \mathrm{O}$, i.e. $\omega_{ij}(0) = 0 \forall ij$, we may define an improved system estimator through the *seemingly unrelated* structure

$$\begin{aligned}
&\beta' = [\beta_1', \beta_2', \ldots, \beta_m'] \\
&X = \mathrm{Diag}[X_1, X_2, \ldots, X_m], \qquad [\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_N}] = X_i' \\
&\mathbf{z}' = [\mathbf{z}_1', \mathbf{z}_2', \ldots, \mathbf{z}_m'], \qquad [z_{i_1}, z_{i_2}, \ldots, z_{i_N}] = \mathbf{z}_i' \\
&\hat{\beta}(k+1) = \{\hat{X}(k)'[\hat{\Sigma}(k) \otimes I_N]^{-1}\hat{X}(k)\}^{-1}\hat{X}(k)'[\hat{\Sigma}(k) \otimes I_N]^{-1}\mathbf{z}
\end{aligned}$$

This is the *final* expression of the iterative pseudolinear regression algorithm for the whole STF system. As we may see, it can be easily implemented on standard statistical software.

The pseudolinear estimation of vector ARMA_m models has already been considered by Spliid (1983). His derivation, however, is heuristic and so the analysis of the statistical properties of the algorithm are not correct (see Hannan and McDougall, 1988). In what follows we briefly reconsider the matter.

Since an $\mathrm{ARMA}_m(p, q)$ can be recast in a vector $\mathrm{ARMA}_M(1, 1)$ form with $M = m \cdot \max(p, q)$, we shall consider, without loss of generality, the estimation of the model $\mathbf{z}_t = \Phi\mathbf{z}_{t-1} + \Theta\mathbf{e}_{t-1} + \mathbf{e}_t$. Now let $\beta = \mathrm{Vec}[\Phi : \Theta]'$, and define the multivariate expansion

$$\mathbf{e}_t(\beta) \approx \mathbf{e}_t(\hat{\beta}) - \Xi_t(\hat{\beta})'(\beta - \hat{\beta}), \qquad \Xi_t(\beta) = -\left[\frac{\partial \mathbf{e}_t(\beta)}{\partial \beta} = \left\{\frac{\partial e_{i,}(\beta)}{\partial \beta_j}\right\}\right]_{m \cdot 2m^2}$$

the corresponding iterative NLS estimator then becomes

$$\mathrm{NLS} \quad \hat{\beta}(k+1) = \hat{\beta}(k) + \left[\sum_{t=1}^{N} \hat{\Xi}_t(k)\hat{\Xi}_t'(k)\right]^{-1} \sum_{t=1}^{N} \hat{\Xi}_t(k)\hat{\mathbf{e}}_t(k)$$

Now, using rules of matrix differentiation, typical rows of Ξ_t are given by

$$\xi_{h,}(\beta)\begin{cases} -\partial \mathbf{e}_t(\beta)/\partial\phi_{ij} = \Theta(B)^{-1}J_{ij}\mathbf{z}_{t-1} \\ -\partial \mathbf{e}_t(\beta)/\partial\theta_{ij} = \Theta(B)^{-1}J_{ij}\mathbf{e}_{t-1} \end{cases} \qquad h = 1, 2, \ldots, 2m$$

where J_{ij} is a matrix with 1 in position ij and 0 elsewhere. Proceeding as before, we may then realize a pseudolinear estimator by approximating the gradient with the input–output quantities of the ARMA_m system: $\xi_{h,}' \approx [\mathbf{z}_{t-1}', \mathbf{e}_{t-1}'] \forall h$, i.e. by avoiding the filtering with $\Theta(B)^{-1}J_{ij}$. Here, by extending at the iterative level the recursive analysis of Chen and Guo (1987), we may state that this approximation does not affect the properties of the NLS estimator only if the MA matrix behaves like a passive (or dissipative) filter.

$$\mathrm{Re}\left[\Theta(e^{i\lambda})^{-1} - \frac{1}{2}I_m\right] > 0, \qquad \lambda \in [-\pi, \pi] \tag{8}$$

In his analysis Spliid (1983) has not considered this aspect and wrongly concluded that PLR algorithms always converge. With respect to the STF model, we may note that condition (8) is much more involved, and it might not be easy to check and satisfy.

Finally, the structure of the iterative pseudolinear regression which simultaneously estimates all the system parameters $[\Phi : \Theta : \Sigma]$, and so approximates the efficient maximum likelihood estimator, is

$$X' = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N], \qquad [\mathbf{z}'_{t-1}, \mathbf{e}'_{t-1}] = \mathbf{x}'_t$$

$$\text{PLR} \quad \hat{\boldsymbol{\beta}}(k+1) = \{[I_m \otimes \hat{X}(k)]' [\hat{\Sigma}(k) \otimes I_N]^{-1} [I_m \otimes \hat{X}(k)]\}^{-1} [I_m \otimes \hat{X}(k)]' [\hat{\Sigma}(k) \otimes I_N]^{-1} \mathbf{z}$$

This form is somewhat different of that provided by Spliid (1983), and heuristically derived as an iterative multiple OLS estimator. From the theory of seemingly unrelated equations, an estimator of the dispersion matrix of $\hat{\boldsymbol{\beta}}(k)$ is given by

$$\hat{\Omega}(k) = \{[I_m \otimes \hat{X}(k)]' [\hat{\Sigma}(k) \otimes I_N]^{-1} [I_m \otimes \hat{X}(k)]\}^{-1}$$

but this only is an approximation of the consistent expression provided by Spliid (1983) and Hannan and McDougall (1988), based on the gradient $\Xi_t(\boldsymbol{\beta})$. However, under the same condition of convergence (8) (i.e. $\Theta(B) \approx I_m$) the statistic $\hat{\Omega}_N$ tends to be a good approximation, useful for inferential purposes.

AN ECONOMIC EXAMPLE

The problem considered for empirical comparisons turns on the analysis of the foreign sources of price inflation in Italy. We define five economic variables: Z_1 = exchange rate lira/dollar (\$), Z_2 = index of wholesale prices (PI), Z_3 = index of export prices (PX), Z_4 = index of import prices (PM), Z_5 = balance of foreign trade (B), t = monthly data 1973.01–1985.12 ($N = 156$).

Graphically, all the processes have evidenced components of trend. The analysis of correlograms and variances on differenced series have shown that stationarity in covariance may be reached with a difference of order one $(1 - B)$ for all the variables. The corresponding sample correlation functions are reported in Tables AI and AII in Appendix 3. Here, we may note the simultaneous correlation of the prices due to the fact that PX, PM are the prices of exported and imported goods (so that PI, PX, PM are synonymous). The series $(1 - B)PI_t$ and $(1 - B)B_t$ still exhibit a considerable autocorrelation (of AR(1) and MA(1) types, respectively), whereas the others are practically white noises. Following the Box–Jenkins approach, to identify the impulse response functions $v_{ij}(B)$, an analysis on prewhitened series is required. The corresponding univariate filters are $(1 + 0.621B)(1 - B)Z_{2t} = u_{1t}$, $(1 - B)Z_{5t} = (1 - 0.763B)u_{5t}$ and the resulting sample correlation functions are given in Table AIII (Appendix 3).

Identification and estimation of ARMA$_5$

Following the considerations above, in our data we identify an ARMA$_5$(1, 1) model. Note, instead, that the *aggregate* Tiao–Box statistics in Table AI (Appendix 3), at most, suggest an AR$_5$(2). The implementation of the PLR algorithm has required these steps:

(0) OLS estimation of an $AR_5(3)$: $\mathbf{z}_t = \sum_k^3 \Phi_k \mathbf{z}_{t-k} + \mathbf{e}_t$,
and generate $\hat{\mathbf{e}}_t(0) = \mathbf{z}_t - \sum_k^3 \hat{\Phi}_k(0)\mathbf{z}_{t-k}$
(1) OLS estimation of $\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \Theta \hat{\mathbf{e}}_{t-1}(0) + \mathbf{e}_t$,
and generate $\hat{\mathbf{e}}_t(1) = \mathbf{z}_t - \hat{\Phi}(1)\mathbf{z}_{t-1} - \hat{\Theta}(1)\hat{\mathbf{e}}_{t-1}(1)$
(2) OLS estimation of $\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \Theta \hat{\mathbf{e}}_{t-1}(1) + \mathbf{e}_t$, etc.

In step (0) we overcome the need for initial values of the parameters $\hat{\Phi}(0)$, $\hat{\Theta}(0)$ (which in the vector ARMA context are not readily available), by directly generating the series $\hat{\mathbf{e}}_t(0)$ with an autoregression of order $p^* > (p+q)$. This *general solution* of the initial value problem makes the pseudolinear algorithms more suitable than other efficient methods, not only from a practical viewpoint but also from an analytical one. Indeed, initial values are crucial in the search for the global minimum:

$$\begin{pmatrix} z_{1_t} \\ z_{2_t} \\ z_{3_t} \\ z_{4_t} \\ z_{5_t} \end{pmatrix} = \begin{pmatrix} . & . & \phi_{13} & . & . \\ . & \phi_{22} & . & . & . \\ . & . & \phi_{33} & . & . \\ \phi_{41} & . & . & . & \phi_{45} \\ . & \phi_{52} & . & \phi_{54} & . \end{pmatrix} \begin{pmatrix} z_{1_{t-1}} \\ z_{2_{t-1}} \\ z_{3_{t-1}} \\ z_{4_{t-1}} \\ z_{5_{t-1}} \end{pmatrix} + \begin{pmatrix} \theta_{11} & . & \theta_{13} & . & . \\ . & . & . & \theta_{24} & . \\ \theta_{31} & . & \theta_{33} & . & . \\ . & . & . & \theta_{44} & . \\ . & \theta_{52} & . & . & \theta_{55} \end{pmatrix} \begin{pmatrix} e_{1_{t-1}} \\ e_{2_{t-1}} \\ e_{3_{t-1}} \\ e_{4_{t-1}} \\ e_{5_{t-1}} \end{pmatrix} + \begin{pmatrix} e_{1_t} \\ e_{2_t} \\ e_{3_t} \\ e_{4_t} \\ e_{5_t} \end{pmatrix}$$

In a first estimation the algorithm has not converged due, probably, to the simultaneous correlation, the great number of parameters to be estimated (75) and the non-significance of many ϕ_{ij}, θ_{ij}. A simplification of the model, obtained by eliminating all non-significant regression coefficients (in the third iteration) and the introduction of a parameter of *stepsize* $0 < \varepsilon < 1$ such that $\tilde{\beta}(k) = \varepsilon(k)\hat{\beta}(k) + [1 - \varepsilon(k)]\hat{\beta}(k-1)$ have improved the situation. In 13 iterations, with constant $\varepsilon(k) = 1/2$, convergence was achieved (Table I). The introduction of ε practically transforms the PLR algorithm in a step-variable estimator. For this scheme Stoica *et al.* (1985) have indeed shown that the necessary condition (8) weakens to a condition of *positive reality*: $\mathrm{Re}[\Theta(z)^{-1}] > 0$, $|z| = 1$.

Table I. Identification and estimation of $ARMA_m$

$\hat{\phi}_{13} = 1.284$	(2.0)	$\hat{\phi}_{52} = 0.130$	(1.7)	$\hat{\theta}_{31} = 0.073$	(3.4)	$\hat{\sigma}_1 = 832.2$
$\hat{\phi}_{22} = 0.504$	(8.7)	$\hat{\phi}_{54} = -0.011$	(−3.6)	$\hat{\theta}_{33} = -0.70$	(−3.9)	$\hat{\sigma}_2 = 0.92$
$\hat{\phi}_{33} = 0.317$	(1.8)	$\hat{\theta}_{11} = 0.130$	(1.7)	$\hat{\theta}_{44} = -0.16$	(−2.0)	$\hat{\sigma}_3 = 65.8$
$\hat{\phi}_{41} = 0.125$	(2.6)	$\hat{\theta}_{13} = -1.70$	(−2.5)	$\hat{\theta}_{52} = -0.21$	(−2.1)	$\hat{\sigma}_4 = 306.0$
$\hat{\phi}_{45} = 5.900$	(3.6)	$\hat{\theta}_{24} = 0.010$	(2.5)	$\hat{\theta}_{55} = -0.67$	(−8.3)	$\hat{\sigma}_5 = 0.44$

Note: The statistics in parentheses are t-ratios.

Identification and estimation of STF

Reasoning on Tables AII and AIII, by applying the univariate-unidirectional Box–Jenkins schemes we identify by cell the STF model below. Here, the orders of many impulse response functions (in particular, those of rational polynomials $1/\delta_{ij}(B)$), are at the limit of acceptability. This *forcing* to non-linearity, however, is necessary to test the performance of

the estimator:

$$\begin{pmatrix} 1 & \cdot & \left(\frac{-\omega_0+\omega_1 B+\omega_2 B^2}{1+\delta_1 B}\right)B^{11} & \cdot & \left(\frac{-\omega_0}{1-\delta_1 B^4}\right)B^4 \\ \left(\frac{\omega_0}{1+\delta_1 B^2}\right)B & 1 & \cdot & \left(\frac{-\omega_0}{1-\delta_1 B^4}\right)B^4 & \cdot \\ \left(\frac{\omega_0}{1+\delta_1 B}\right)B & \omega_0 B^3 & 1 & \left(\frac{\omega_0}{1-\delta_1 B^8}\right)B^8 & \cdot \\ \left(\frac{\omega_0}{1+\delta_1 B}\right)B & \omega_0 B & \omega_0 B^7 & 1 & \left(\frac{\omega_0}{1-\delta_1 B^3}\right)B \\ \cdot & \cdot & \left(\frac{\omega_0}{1-\delta_1 B^2}\right)B^6 & \left(\frac{-\omega_0}{1-\delta_1 B^3}\right)B & 1 \end{pmatrix} \begin{pmatrix} z_{1_t} \\ z_{2_t} \\ z_{3_t} \\ z_{4_t} \\ z_{5_t} \end{pmatrix} = \begin{pmatrix} (1+\theta_1 B)a_{1_t} \\ \left(\frac{1}{1+\phi_1 B}\right)a_{2_t} \\ (1-\theta_1 B)a_{3_t} \\ (1-\theta_1 B^3)a_{4_t} \\ (1-\theta_1 B)a_{5_t} \end{pmatrix}$$

As in the previous model, the *philosophy* of pseudolinearity may be used to obtain initial values of the parameters. Indeed, by means of a set of 'linear' equations

$$\text{(i) } z_{i_t} = \boldsymbol{\delta}_i' \mathbf{z}_{i_{t-1}} + \boldsymbol{\omega}_{ij}' \mathbf{z}_{j_{t-b}} + n_{ij_t}, \qquad \text{(ii) } \hat{n}_{i_t} = \boldsymbol{\alpha}_i' \hat{\mathbf{n}}_{i_{t-1}}, \qquad \text{(iii) } \hat{n}_{i_t} = \boldsymbol{\phi}_i' \hat{\mathbf{n}}_{i_{t-1}} + \boldsymbol{\theta}_i' \hat{\mathbf{e}}_{i_{t-1}} + a_{i_t}$$

we sequentially obtain the OLS estimates $\hat{\boldsymbol{\delta}}_{ij}(0)$, $\hat{\boldsymbol{\omega}}_{ij}(0)$, $\hat{e}_{i_t}(0)$, $\hat{\boldsymbol{\phi}}_i(0)$, $\hat{\boldsymbol{\theta}}_i(0)$. In the second step we first generate a white noise process $\{e_{i_t}\}$ with an autoregression of order $p_i^* > (p_i + q_i)$. Then the ARMA parameters are estimated vai PLR. In the various iterations pseudolinear regressors $\mathbf{x}_{i_t}$ may be obtained with the three-step filtering procedure described above.

As a consequence of the non-linear forcing, the estimation has required an *ad hoc* search for the function $v_{24}(B)$ (fixing all the other coefficients), to find the narrow band of convergence. Without stepsize $\varepsilon(k) = 1/2$, other polynomials like $1/\delta_{34}(B)$ diverge in an oscillatory manner. In any case, in eight iterations convergence was achieved (Table II). The validity of the disaggregate identification is shown by the statistical significance of the estimates and by the fact that their signs coincide with what is expected from the correlograms.

Rapprochement of ARMA$_m$ and TFS models

In the previous subsection may rational impulse response functions were at the limit of specification. A linear modelling like $v_{ij}(B) = \omega_{ij}(B)$ is more realistic and flexible, and can improve the speed of convergence of the estimators. The model that follows looks like a closed-loop system of simultaneous ARMAX equations and thus provides a substantial

Table II. Identification and estimation of STF

$\hat{\delta}_{13}=$	0.387	(1.8)	$\hat{\delta}_{21}=$	0.625	(3.9)	$\hat{\omega}_{32}=$	0.874	(1.9)	$\hat{\omega}_{43}=$	0.341	(2.2)	$\hat{\omega}_{54}=-0.009$ (−3.1)
$\hat{\omega}_{13}^0=$	−0.826	(−3.6)	$\hat{\omega}_{21}=$	0.011	(4.4)	$\hat{\delta}_{34}=$	−0.498	(−2.2)	$\hat{\delta}_{45}=$	−0.468	(−2.0)	$\hat{\theta}_5=-0.600$ (−7.2)
$\hat{\omega}_{13}^1=$	0.862	(2.9)	$\hat{\delta}_{24}=$	−0.848	(−1.8)	$\hat{\omega}_{34}=$	0.114	(3.6)	$\hat{\omega}_{45}=$	6.53	(3.9)	$\hat{\sigma}_1=631.5$
$\hat{\omega}_{13}^2=$	0.936	(3.8)	$\hat{\omega}_{24}=$	−0.005	(−1.5)	$\hat{\theta}_3=$	−0.341	(−4.3)	$\hat{\theta}_4=$	−0.216	(−2.7)	$\hat{\sigma}_2=$ 0.79
$\hat{\delta}_{15}=$	−0.658	(−2.1)	$\hat{\phi}_2=$	0.541	(8.6)	$\hat{\delta}_{41}=$	0.556	(1.2)	$\hat{\delta}_{53}=$	−0.590	(−2.2)	$\hat{\sigma}_3=$ 55.3
$\hat{\omega}_{15}=$	−5.87	(−2.3)	$\hat{\delta}_{31}=$	0.143	(0.7)	$\hat{\omega}_{41}=$	0.083	(1.7)	$\hat{\omega}_{53}=$	0.019	(3.2)	$\hat{\sigma}_4=265.6$
$\hat{\theta}_1=$	0.232	(2.8)	$\hat{\omega}_{31}=$	0.099	(4.3)	$\hat{\omega}_{42}=$	2.81	(2.2)	$\hat{\delta}_{54}=$	−0.501	(−2.0)	$\hat{\sigma}_5=$ 0.39

Table III. Identification and estimation of ARMAX$_m$

$\hat{\omega}^1_{13}=$	-0.652	(-2.8)	$\hat{\phi}_2=$	0.503	(8.0)	$\hat{\omega}^2_{34}=$	-0.073	(-2.4)	$\hat{\omega}^1_{45}=$	6.49	(4.0)	$\hat{\omega}^2_{54}=$	0.007	(2.3)
$\hat{\omega}^2_{13}=$	0.655	(2.7)	$\hat{\omega}^1_{24}=$	-0.007	(-1.9)	$\hat{\omega}^3_{34}=$	-0.101	(-3.2)	$\hat{\omega}^2_{45}=$	-4.78	(-3.0)	$\hat{\theta}_5=$	-0.551	(-6.6)
$\hat{\omega}^3_{13}=$	1.12	(4.6)	$\hat{\omega}^2_{24}=$	0.010	(2.7)	$\hat{\omega}_3=$	-0.325	(-4.0)	$\hat{\omega}_4=$	-0.243	(-2.9)	$\hat{\sigma}_1=$	622.4	
$\hat{\omega}^1_{15}=$	-5.53	(-2.2)	$\hat{\omega}^1_{31}=$	0.083	(3.7)	$\hat{\theta}^1_{41}=$	0.081	(1.7)	$\hat{\omega}^1_{53}=$	-0.013	(-2.2)	$\hat{\sigma}_2=$	0.78	
$\hat{\omega}^2_{15}=$	7.62	(2.9)	$\hat{\omega}^2_{31}=$	0.038	(1.9)	$\hat{\omega}^2_{41}=$	0.073	(1.5)	$\hat{\omega}^2_{53}=$	0.023	(3.9)	$\hat{\sigma}_3=$	51.3	
$\hat{\theta}_1=$	0.231	(2.8)	$\hat{\omega}^1_{32}=$	0.814	(1.7)	$\hat{\omega}^1_{42}=$	3.67	(2.8)	$\hat{\omega}^3_{53}=$	-0.019	(-3.1)	$\hat{\sigma}_4=$	242.6	
$\hat{\omega}^1_{21}=$	0.009	(3.6)	$\hat{\omega}^2_{32}=$	0.650	(1.3)	$\hat{\omega}^2_{42}=$	-2.15	(-1.9)	$\hat{\omega}^4_{53}=$	0.023	(3.6)	$\hat{\sigma}_5=$	0.374	
$\hat{\omega}^2_{21}=$	0.006	(2.3)	$\hat{\omega}^1_{34}=$	0.114	(3.5)	$\hat{\omega}^1_{43}=$	0.376	(2.4)	$\hat{\omega}^1_{54}=$	-0.009	(-3.0)			

compromise of the models considered so far:

$$\phi_i(B)z_{i_t} - \left(\sum_{j\neq i}^{m} \omega_{ij}(B)z_{j_{t-b}}\right) = \theta_i(B)a_{i_t} \qquad i = 1, 2, \ldots, m$$

$$\text{ARMAX}_m \qquad [\ddot{\Phi}(B) - \Omega(B)]\,\mathbf{z}_t = \ddot{\Theta}(B)\mathbf{a}_t$$

The ω^k_{ij} parameter are identified in the same position as the significant sample cross-correlation coefficients. Convergence was achieved in four iterations and estimation results are given in Table III.

The feasibility of a disaggregated identification for STF models, which emerges in this section, shows the adaptability of the STF structure and its superiority with respect to the ARMA$_m$ one. This last, in fact, is rather rigid and can be completely specified only in an iterative cycle of estimation and testing. In practice, from the sample correlation functions it is not possible to identify which coefficients ϕ^k_{ij}, θ^k_{ij} would turn out to be significant in estimation. These are, however, heuristic considerations. Parametric comparisons which relate to the operative behaviour of time series models will be developed in the sequel.

$$\begin{pmatrix} 1 & . & -\omega_1 B^{11} + \omega_2 B^{12} + \omega_3 B^{13} & . & -\omega_1 B^4 + \omega_2 B^8 \\ \omega_1 B + \omega_2 B^3 & (1+\phi_1 B) & . & -\omega_1 B^4 + \omega_2 B^8 & . \\ \omega_1 B + \omega_2 B^2 & \omega_1 B^3 + \omega_2 B^{12} & 1 & \omega_1 B^8 - \omega_2 B^9 - \omega_3 B^{16} & . \\ \omega_1 B + \omega_2 B^2 & \omega_1 B - \omega_2 B^6 & \omega_1 B^7 & 1 & \omega_1 B - \omega_2 B^4 \\ . & . & -\omega_1 B^5 + \omega_2 B^6 - \omega_3 B^8 + \omega_4 B^{11} & -\omega_1 B + \omega_2 B^4 & 1 \end{pmatrix} \begin{pmatrix} z_{1_t} \\ z_{2_t} \\ z_{3_t} \\ z_{4_t} \\ z_{5_t} \end{pmatrix} = \begin{pmatrix} (1+\theta_1 B)a_{1_t} \\ a_{2_t} \\ (1-\theta_1 B)a_{3_t} \\ (1-\theta_1 B^3)a_{4_t} \\ (1-\theta_1 B)a_{5_t} \end{pmatrix}$$

FORECASTING AND COMPARISONS

This section is concerned with the analysis of forecasting algorithms and with the evaluation of the forecasting performance of the multivariate systems considered previously. Prediction techniques for vector ARMA and ARMAX models are straightforward generalizations of those developed in the univariate context. However, for STF models there are inherent complications. Below we shall propose a very simple and practical algorithm for STF which relates to its pseudolinear representation and estimation. Finally, empirical comparisons of the estimated models will be developed with typical statistics of prediction errors.

Jenkins and Alavi (1981) have provided a vector extension of the univariate prediction

algorithms of Box and Jenkins (1970). In practice if $\Im_t$ is the algebra of the events generated by the set of information $\{z_{t-k}\}_0^\infty$ (past and present), the optimal l-step-ahead predictor $\hat{\mathbf{z}}_t(l) = E[\mathbf{z}_{t+l} \mid \Im_t]$ and the related conditional dispersion matrix $\hat{\Sigma}(l) = E\{[\mathbf{z}_t(l) - \mathbf{z}_{t+l}]^2 \mid \Im_t\}$ are given by

$$\hat{\mathbf{z}}_t(l) = \Phi_1 \hat{\mathbf{z}}_t(l-1) + \cdots + \Phi_p \mathbf{z}_{t+l-p} + \Theta_l \mathbf{e}_t + \cdots + \Theta_q \mathbf{e}_{t+l-q}, \qquad \hat{\Sigma}(l) = \sum_{k=0}^{l-1} \Psi_k \Sigma \Psi_k' \tag{9}$$

As regards the STF scheme, the extension of the unidirectional techniques of Box–Jenkins leads to a great complication of calculations. This is the case of the approach that transforms the model into a system of ARMAX equations, with a linearization by row:

$$\left[\phi_i(B) \prod_{j \neq i}^{m} \delta_{ij}(B)\right] z_{i_t} = \sum_{j \neq i}^{m} \left[\omega_{ij}(B)\phi_i(B) \prod_{h \neq j}^{m} \delta_{hj}(B)\right] z_{j_{t-b}} + \left[\theta_i(B) \prod_{j \neq i}^{m} \delta_{ij}(B)\right] a_{i_t}$$

Algorithm (9) could then be applied but the number of matrices involved increases rapidly with m. Moreover, searching for the common linear factors to be cancelled may not be easy.

Similar problems are encountered in utilizing the infinite linear representation which results by expanding in Taylor series the impulse functions $v_{ij}(B)$, $\psi_i(B)$:

$$\hat{\mathbf{z}}_t(l) = \sum_{k=0}^{l-b+1} V_k \hat{\mathbf{z}}_t(l-b-k) + \sum_{k=l-b}^{\infty} V_k \mathbf{z}_{t+l-b-k} + \sum_{k=l}^{\infty} \ddot{\Psi}_k \mathbf{a}_{t+l-k}$$

Now, a substantial simplification is afforded by utilizing the pseudolinear representation (5). Indeed, if $b = \min_{ij}(b_{ij}) > 0$ the predictor implemented on that form coincides with the solution of m^2 single-equation sub-problems:

$$\hat{z}_{i_t}(l) = \sum_{j \neq i}^{m} \hat{w}_{ij_t}(l) + \hat{n}_{i_t}(l), \qquad i = 1, 2, \ldots, m$$

$$\hat{w}_{ij_t}(l) = \delta_{ij}^1 \hat{w}_{ij_t}(l-1) + \cdots + \delta_{ij}^r w_{ij_{t+l-r}} + \omega_{ij}^0 \hat{z}_{j_t}(l-b) + \cdots + \omega_{ij}^s z_{j_{t+l-b-s}}$$

$$\hat{n}_{i_t}(l) = \phi_i^1 \hat{n}_{i_t}(l-1) + \cdots + \phi_i^p n_{i_{t+l-p}} + \theta_i^l a_{i_t} + \cdots + \theta_i^q a_{i_{t+l-q}}$$

A complication in this approach seems represented by the fact that the starting auxiliary variables w_{ij_t}, n_{i_t} are available only by filtering the whole sample of data $\{Z_t\}_1^n$. However, as we have seen, these quantities are generated in the estimation phase in an adaptive manner. The variance of the STF predictor is still given by equation (9), where the Ψ_k weights are computed as above.

Empirical comparisons of the forecasting performance can be developed in two ways: the size of the prediction regions and the value of the errors $|\hat{\mathbf{z}}_t(l) - \mathbf{z}_{t+l}|$. About the first indicator, note that $[\mathbf{z}_{t+l} \mid \Im_t] \sim N_m[\hat{\mathbf{z}}_t(l), \hat{\Sigma}(l)]$. Hence the prediction region of probability $(1-\alpha)$ may be obtained from the quadratic relationship

$$[\hat{\mathbf{z}}_t(l) - \mathbf{z}_{t+l}]' \hat{\Sigma}(l)^{-1} [\hat{\mathbf{z}}_t(l) - \mathbf{z}_{t+l}] = \chi_\alpha^2(m)$$

It is well known that the size of the corresponding ellipsoid of concentration is proportional to the determinant $|\hat{\Sigma}(l)|$, and in Table IV some values of this are displayed. We may see that for l-small the volume is largely determined by $|\Sigma|$. Moreover, the values associated with STF-ARMAX$_m$ increase more gradually.

The limit of this test is that it depends on the quantities Σ, Ψ_k, estimated on past data. Moreover, it does not directly relate to the operative behaviour of the models. Out-of-sample comparisons based on mean absolute prediction errors (MAPE; see Liu, 1987; and DeGooijer

Table IV. Comparison of the variance $|\hat{\Sigma}(l)|$ of prediction errors

l	$ARMA_m$	STF	$ARMAX_m$
1	3 942 899	1 831 615	1 645 809
3	12 584 743	6 457 322	5 957 136
6	12 944 955	9 209 172	8 478 442
12	12 950 172	14 168 371	14 587 210

Note: Recall that the generalized variance $\hat{\Sigma} = \hat{\Sigma}(1)$.

and Klein, 1989), avoid these problems. We propose a second version of such a statistic:

$$\text{MAPE}_1(h \mid t) = \frac{1}{h} \sum_{t=1}^{h} \left| \frac{\hat{Z}_t(l) - Z_{t+l}}{Z_{t+l}} \right|, \qquad \text{MAPE}_2(l \mid h) = \frac{1}{h} \sum_{\tau=1}^{h} \left| \frac{\hat{Z}_{t+\tau}(l) - Z_{t+\tau+l}}{Z_{t+\tau+l}} \right|$$

where $Z_t = (1 - B)^{-d} z_t$. Note that in MAPE_1 the forecast origin (t) is held fixed and what varies is the forecast horizon (l), whereas MAPE_2 deals with the opposite situation. This is motivated by the need for a discrepancy index which must be independent of the particular t and able to provide significant results even for small l. On the other hand, a greater number of post-sample observations is required to compute MAPE_2.

Table V gives values of the two indicators on different lags h, l and for each individual series Z_{it}. The forecast horizon of MAPE_1 is $t = 1986.01 - 1986.12$ while for MAPE_2 it is $t = 1986.01 - 1987.12$ with $h = 12$. The conclusions that can be drawn from the analysis of Tables IV and V are important:

(1) With in-sample comparisons the great superiority of the STF model, with respect to the $ARMA_m$ one, is well established by $|\hat{\Sigma}_a| < 1/2\,|\hat{\Sigma}_e|$ (Table IV), but the same conclusion cannot be stated in the out-of-sample context. This contradiction is due to the fact that the first analysis is conducted on the stationary series $\mathbf{a}_t, \mathbf{e}_t$, whereas the second is made on the integrated series Z_{it}. In practice, due to the predominant role of the unit roots $(1 - B)$ on the stable components, multivariable models with the same degree (d) of nonstationarity tend to have a close operative behaviour.

(2) Specific forecasting of the differenced series z_{it} have re-established the superiority of the

Table V. Comparison of mean absolute prediction errors

→		M	A	P	E_1			M	A	P	E_2	
↓	h	Z_1	Z_2	Z_3	Z_4	Z_5	l	Z_1	Z_2	Z_3	Z_4	Z_5
A	1	0.013	0.004	0.023	0.006	0.65	1	0.023	0.001	0.010	0.011	0.35
R	3	0.044	0.009	0.020	0.011	0.57	3	0.037	0.005	0.018	0.012	0.31
M	6	0.027	0.012	0.015	0.012	0.49	6	0.063	0.009	0.015	0.018	0.32
A_m	12	0.062	0.011	0.017	0.026	0.81	12	0.095	0.015	0.026	0.032	0.29
S	1	0.015	0.003	0.016	0.004	0.71	1	0.017	0.001	0.008	0.007	0.40
T	3	0.032	0.007	0.011	0.005	0.55	3	0.028	0.003	0.011	0.010	0.32
F	6	0.021	0.008	0.008	0.006	0.51	6	0.051	0.006	0.015	0.014	0.23
	12	0.054	0.007	0.012	0.019	0.73	12	0.076	0.008	0.020	0.023	0.21

STF model. Indeed, the term Φ^l, utilized by ARMA_m in computing $\hat{z}_t(l+1)$, decays rapidly with l so that MAPE $\to 1$. On the other hand, by stressing the role of time delays b_{ij}, STF has a broader prediction horizon. This feature can also be seen in Table V since indicators associated with STF increase more slowly. A substantial improvement of this performance could clearly be obtained with the specifications $V_0 = L$, lower triangular, or $V_0 = V_0'$ (in this case, a constrained estimation would be necessary).

(3) Although a significant superiority of the STF scheme can be deduced from Table 5, a relevant aspect appears to be the poor performance of both the models in forecasting the series Z_{5_t}. The reason lies in the fact that Balance of Foreign Trade (B) is the direct aggregation of several autonomous and temporary components (imports, exports, etc.) and so it exhibits nonstationarity in covariance. These considerations enable us to state that methodological research on multivariate systems is important but the future of time series analysis is represented by nonlinear and nonstationary models (see Young, 1989; Priestley, 1988).

In summary, the secret of the STF structure lies in the specialized treatment of auto- and cross-dynamic regression (or ACR and CCR), more specifically in what the system engineers call the interaction matrix, that is, the matrix $D = \{b_{ij}\}$ of time delays. Operatively, the vector ARMA model remains a useful practical tool on which simplified and *automatic* techniques of identification, based on scalar information criteria, can be applied. On the other hand, given the multivariate structure of residuals, its estimation problems (either in pseudolinear or in non-linear form) increase rapidly with the dimension m.

APPENDIX 1 (SUM OF POLYNOMIALS)

Proposition

Let $\{\phi_i(z)\}_1^m$ be m monic polynomials of orders $p_i \leqslant p$, with real coefficients and roots lying outside the unit circle: $\phi_i(0) = 1$, $\phi_i(z) \neq 0 \,|\, z| \leqslant 1$ for all i. Their sum defines a polynomial $\phi(z)$ with roots of the same kind, in general only if $p = \max(p_i) < 3$.

Proof

It is easy to see that $\phi(z)$ has degree p and coefficients

$$\phi(z) = \sum_i^m \phi_i(z) = m + \sum_j^p \left(\sum_i^m \phi_i\right) z^j$$

If we consider the parameters of $\phi_i(z)$ as a vector $\boldsymbol{\phi}_i \in \Re^p$, associated with $\phi(z)$ we have $\boldsymbol{\phi} = \Sigma_i^m \boldsymbol{\phi}_i$. Moreover, by hypothesis, $\boldsymbol{\phi}_i \in S_i \subset \Re^p$, where S_i is the region of stability of order p_i. Since $\phi(z)$ and $\phi(z)/m$ have the same roots it is clear that $\boldsymbol{\phi}/m \in S_p$ only if $S_i \subseteq S_{i+1}$ and S_p is a *convex* set. In this case we would have $\boldsymbol{\phi}_i \in S_p$, and $\boldsymbol{\phi}/m$ would form a *convex combination* of $\boldsymbol{\phi}_i$ in S_p. $S_i \subseteq S_{i+1}$ because a polynomial of degree p_i is a particular case of one of degree $p_i + 1$. However, S_p is convex only for $p \leqslant 2$ (see Box and Jenkins, 1970).

APPENDIX 2 (STRUCTURAL IDENTIFICATION)

The first condition (i) assures that the matrices $[I - V(z)]$, $\ddot{\Psi}(z)$ be left coprime, such that, as Hannan (1969) has shown, their sole admissible greatest common left divisor be a *unimodular*

matrix $U(z)$. The diagonality of $\ddot{\Psi}(z)$, however, compels $U(z)$ to be also diagonal and, by definition of unimodularity ($U(z)$ linear and $\mathrm{Det}U(z)$ constant), it must in addition be constant: $\dot{U}$.

Although any matrix $\tilde{\Psi}(z) = \Psi(z)\dot{H}(z)Q$, with $\dot{H}(z) = \mathrm{Diag}[h_i(z)/h_i(z^{-1})]$ and Q orthogonal (arbitrary), will also satisfy the spectral representation of $\Gamma(z)$, because $\dot{H}(z)QQ'\dot{H}(z^{-1}) = I$. The matrix $\dot{H}(z)$ (and so $\ddot{\Psi}(z)$) does not simultaneously enjoy the properties of stationarity and invertibility (1). Indeed, if $h_i(z)$ has roots in $|z| > 1$, $h_i(z^{-1})$ must have roots in $|z| \leqslant 1$. Therefore it must be $\dot{H}(z) = \dot{U}$ constant. Moreover, it is easy to assess that conditions (iii) constrain $\dot{U} = Q = I$.

Finally, the specifications in (iii), in particular $\Psi_0 = \Psi_0'$, are identified because, by assuming the process $\mathbf{z}_t$ uncorrelated, we would have $\Gamma_0 = \Psi_0\Psi_0'$, that is, $\Psi_0 = \sqrt{\Gamma_0} = P\sqrt{\Lambda}P'$, which is symmetrical and positive definite.

APPENDIX 3 (SAMPLE CORRELATION FUNCTIONS)

Table AI. Autocorrelations and Tiao–Box statistics

Lag	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
z_1	0.24	0.07	0.05	0.08	0.12	0.00	0.01	0.06	−0.05	−0.06	−0.13	−0.06	0.05	0.14	0.03	0.08	−0.03	0.02	−0.03	−0.03
z_2	0.62	0.41	0.28	0.23	0.20	0.14	0.11	0.09	0.00	−0.03	0.00	0.08	0.02	0.01	−0.04	−0.01	0.00	0.02	0.00	0.04
z_3	−0.16	−0.02	0.01	0.04	0.00	0.11	−0.01	−0.03	−0.04	0.04	−0.10	0.09	0.03	0.07	−0.01	0.00	0.05	−0.05	0.04	0.01
z_4	−0.09	0.08	−0.18	0.04	0.09	0.12	0.14	−0.18	−0.04	−0.12	0.15	0.10	−0.06	−0.05	−0.03	0.06	−0.02	0.12	−0.05	−0.04
z_5	−0.39	−0.11	−0.07	−0.09	0.08	−0.11	0.04	0.16	0.02	−0.21	0.03	0.21	−0.07	0.00	−0.12	0.14	−0.01	−0.15	0.05	0.06
$M(k)$	154	69	39	40	39	34	36	29	25	32	38	30				($\chi^2_{01}(25) = 44.3$)				

Note: Typical standard error under null hypothesis is $1/\sqrt{N} = 0.08$ (see Haugh & Box, 1977).

Table A2. Cross correlations (differenced series)

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$z_1 \rightarrow z_2$	0.34	0.45	0.24	0.23	0.09	0.10	0.08	0.07	0.14	0.04	−0.07	−0.10	−0.14	−0.10	0.00	0.02	−0.06	0.02	0.00	0.04
$z_1 \leftarrow z_2$	0.34	0.07	0.08	0.05	0.06	0.02	0.01	0.04	0.06	−0.01	0.04	0.10	0.14	0.14	0.05	0.05	0.14	0.13	0.08	−0.01
$z_1 \rightarrow z_3$	0.35	0.27	0.18	−0.03	0.00	0.05	0.10	−0.02	0.03	−0.13	0.01	−0.07	−0.01	−0.02	−0.05	0.10	0.01	−0.03	−0.06	−0.06
$z_1 \leftarrow z_3$	0.35	0.03	0.01	−0.05	−0.01	−0.07	0.06	−0.10	0.07	−0.02	−0.03	−0.24	0.18	0.20	0.05	0.04	0.01	0.14	0.06	0.00
$z_1 \rightarrow z_4$	0.26	0.25	0.22	0.04	0.06	0.00	0.04	−0.07	0.03	−0.01	−0.11	−0.08	−0.06	−0.05	−0.10	−0.05	0.10	0.01	−0.02	−0.07
$z_1 \leftarrow z_4$	0.26	0.00	0.12	−0.04	−0.09	0.10	0.11	0.00	−0.04	−0.16	0.12	0.00	0.12	0.02	0.12	−0.01	0.05	−0.09	−0.05	0.04
$z_1 \rightarrow z_5$	0.11	−0.08	0.01	−0.02	0.02	0.13	−0.10	0.10	0.04	−0.15	0.04	0.09	0.06	−0.04	−0.04	0.05	0.03	−0.10	−0.14	0.11
$z_1 \leftarrow z_5$	0.11	−0.11	0.07	0.10	−0.26	−0.11	0.13	−0.02	0.17	−0.07	−0.10	0.10	−0.02	−0.09	0.12	−0.01	0.03	−0.02	−0.10	0.04
$z_2 \rightarrow z_3$	0.38	0.08	0.10	0.18	0.11	0.06	0.02	0.05	0.02	0.01	0.09	0.04	0.14	−0.10	0.02	0.03	0.03	−0.04	0.01	−0.06
$z_2 \leftarrow z_3$	0.38	0.25	0.25	0.10	0.02	0.07	0.04	0.09	0.09	0.01	−0.06	−0.11	−0.06	0.05	0.09	−0.04	−0.02	0.08	0.11	−0.01
$z_2 \rightarrow z_4$	0.31	0.24	0.16	0.13	0.08	0.03	−0.06	0.08	0.01	−0.07	0.03	0.01	0.07	−0.02	0.03	−0.01	0.09	−0.01	−0.01	0.06
$z_2 \leftarrow z_4$	0.31	0.27	0.22	0.20	−0.01	0.07	0.03	0.05	0.16	−0.02	−0.05	−0.04	−0.09	0.10	0.05	0.05	−0.10	0.05	−0.12	0.04
$z_2 \rightarrow z_5$	−0.09	−0.05	0.05	−0.02	0.10	0.03	0.05	−0.03	−0.13	0.09	0.01	0.03	−0.07	−0.02	0.02	0.04	0.01	0.02	0.08	−0.06
$z_2 \leftarrow z_5$	−0.09	−0.08	−0.12	0.11	−0.04	−0.12	0.04	0.02	−0.01	0.01	0.01	0.06	−0.05	−0.11	0.02	0.08	−0.03	−0.02	0.00	0.02
$z_3 \rightarrow z_4$	0.26	0.15	0.08	0.01	−0.01	0.07	−0.15	0.21	−0.06	−0.06	−0.06	−0.02	0.10	−0.09	0.04	−0.02	0.03	0.04	0.07	0.01
$z_3 \leftarrow z_4$	0.26	0.04	0.11	−0.01	0.10	−0.08	0.09	−0.04	0.29	−0.24	−0.03	0.06	0.13	0.06	0.06	0.06	−0.25	0.15	−0.01	−0.07
$z_3 \rightarrow z_5$	0.02	−0.05	−0.02	−0.02	0.07	−0.19	0.28	0.00	−0.19	0.02	−0.01	0.22	−0.11	−0.08	0.09	0.04	−0.18	0.11	−0.04	−0.06
$z_3 \leftarrow z_5$	0.02	0.02	0.11	−0.05	−0.14	−0.02	0.02	0.03	0.02	0.00	0.01	−0.02	−0.18	0.15	0.10	−0.09	−0.01	−0.03	0.02	−0.05
$z_4 \rightarrow z_5$	−0.14	−0.22	0.10	−0.05	0.23	−0.10	0.01	0.01	−0.06	−0.11	0.20	−0.02	−0.03	−0.04	0.06	0.01	−0.08	0.01	0.06	0.13
$z_4 \leftarrow z_5$	−0.14	0.26	−0.10	0.13	−0.20	0.01	−0.05	−0.06	0.15	−0.05	0.05	−0.18	0.11	−0.09	0.21	−0.07	−0.08	−0.03	−0.01	0.06

Table AIII. Cross correlations (prewhitened series)

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$z_1 \rightarrow u_2$	0.37	0.31	−0.04	0.12	−0.07	0.06	0.02	0.04	0.11	−0.05	−0.12	−0.09	−0.10	−0.02	0.08	0.02	−0.10	0.07	−0.02	0.05
$z_1 \leftarrow u_2$	0.37	0.02	0.07	0.01	0.07	0.01	−0.01	−0.01	0.09	−0.04	−0.02	0.02	0.07	0.13	0.03	−0.05	0.07	0.11	0.11	−0.01
$z_1 \rightarrow u_5$	0.02	−0.08	−0.06	−0.08	−0.04	0.12	−0.02	0.10	0.12	−0.09	−0.03	0.09	0.15	0.07	0.01	0.07	0.10	−0.03	−0.19	−0.02
$z_1 \leftarrow u_5$	0.02	−0.14	0.00	−0.11	−0.31	0.01	0.19	0.06	0.11	−0.12	−0.04	0.09	−0.03	−0.01	0.12	−0.03	−0.01	−0.06	−0.04	0.10
$u_2 \rightarrow z_3$	0.41	0.02	−0.02	0.14	0.09	0.07	−0.02	0.05	0.01	−0.06	0.09	−0.06	0.26	−0.14	0.00	0.01	0.08	−0.06	0.06	−0.05
$u_2 \leftarrow z_3$	0.41	0.02	0.13	−0.06	−0.06	0.07	−0.01	0.09	0.04	−0.05	−0.10	−0.08	−0.01	0.12	0.08	−0.12	0.01	0.12	0.07	−0.10
$u_2 \rightarrow z_4$	0.20	0.18	0.10	0.10	0.08	0.09	−0.14	0.09	0.06	−0.11	0.03	−0.04	0.11	−0.05	0.04	−0.09	0.12	0.00	−0.06	0.12
$u_2 \leftarrow z_4$	0.20	0.11	0.07	0.08	−0.17	0.10	−0.01	0.05	0.16	−0.15	−0.05	−0.02	−0.08	0.20	−0.02	0.02	−0.16	0.14	−0.20	0.15
$u_2 \rightarrow u_5$	−0.11	−0.21	−0.07	−0.18	−0.01	−0.01	0.09	0.14	−0.17	−0.01	−0.01	0.10	0.00	−0.05	−0.06	0.01	0.01	−0.04	0.15	0.01
$u_2 \leftarrow u_5$	−0.11	−0.02	0.02	0.13	−0.19	−0.04	0.13	−0.02	−0.02	0.02	−0.01	−0.01	−0.13	0.00	0.17	0.02	−0.10	0.03	0.02	0.01
$z_3 \rightarrow u_5$	0.04	−0.04	−0.05	−0.07	0.03	−0.19	0.19	0.14	−0.12	−0.07	−0.06	0.21	0.04	−0.06	0.06	0.08	−0.14	0.02	−0.03	−0.09
$z_3 \leftarrow u_5$	0.04	0.02	−0.02	−0.19	−0.16	0.00	0.04	0.02	−0.01	−0.04	−0.05	−0.08	−0.07	0.19	0.02	−0.13	−0.04	−0.04	−0.01	−0.04
$z_4 \rightarrow u_5$	−0.03	−0.28	−0.10	−0.13	0.16	0.01	0.02	0.03	−0.05	−0.18	0.10	0.06	0.02	−0.03	0.05	0.06	−0.05	−0.02	0.06	0.20
$z_4 \leftarrow u_5$	−0.03	0.19	−0.15	−0.04	−0.25	−0.02	−0.04	0.02	0.11	−0.09	−0.03	−0.12	0.12	−0.01	0.12	−0.16	−0.10	−0.01	0.04	0.07

ACKNOWLEDGEMENTS

The author is grateful to the departmental editor and to a referee for their helpful suggestions.

REFERENCES

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day (1970).

Box, G. E. P. and MacGregor, J. R., 'The analysis of closed-loop dynamic stochastic systems', *Technometrics* **16** (1974), 391–8.

Chen, H. F. and Guo, L., 'Optimal adaptive control and consistent parameter estimates for ARMAX models with quadratic cost', *SIAM Jour. on Control*, **25** (1987), 845–67.

DoGooijer, J. G. and Klein, A., 'Forecasting Antwerp maritime steel traffic flow', *Jour. of Forecasting*, **8** (1989), 381–96.

Faurre, P., Clerget, M. and Germain, F., *Operateurs Rationnels Positifs*, Paris: Dunod (1979).

Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, New York: Academic Press (1977).

Grillenzoni, C., 'Costruzione di un modello a serie storiche multivariato: un approccio disaggregato-simultaneo', *Statistica*, **43** (1983a), 323–53.

Grillenzoni, C., 'La distribuzione asintotica dei *gains* in funzioni di impulso-risposta non lineari', *Statistica*, **43** (1983b), 451–74.

Hannan, E. J. 'The identification of vector mixed ARMA systems', *Biometrika*, **56** (1969), 223–5.

Hannan, E. J., 'The statistical theory of linear systems', in Krishnaiah, P. R. (ed.), *Development in Statistics* **2**, New York: Academic Press (1979).

Hannan, E. J. and McDougall, A. J., 'Regression procedures for ARMA estimation', *Journal of American Statistical Association*, **83** (1988), 490–98.

Hanssens, D. M. and Liu, L. M., 'Lag specification in rational distributed lag structural models', *Jour. of Business and Econ. Statistics*, **1** (1983), 316–25.

Haugh, L. D. and Box, G. E. P., 'Identifiction of dynamic regression models connecting two time series', *Jour. of American Statistical Association*, **72** (1977), 121–30.

Jenkins, G. M., *Practical Experience with Modelling and Forecasting Time Series*, St. Helier: G.J.P. Publications (1979).

Jenkins, G. M. and Alavi, A. S., 'Some aspects of modeling and forecasting multivariate time series', *Jour. of Time Series Analysis*, **2** (1981), 1–49.

Ljung, L., 'Estimation of parameters in dynamical systems' in Hannan, E. J., Krishnaiah, P. R. and Rao, M. M. (eds), *Handbook of Statistics* **5**, Amsterdam: North-Holland (1985).

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

Lii, K. S., 'Transfer function model order and parameter estimation', *Jour. of Time Series Analysis*, **6** (1985), 153–69.

Liu, L. M., 'Sales forecasting using multi-equation transfer function models', *Jour. of Forecasting*, **6** (1987), 223–38.

Liu, L. M. and Hanssens, D. M., 'Identification of multiple input transfer function models', *Comm. in Statistics*, **A11** (1982), 297–314.

Liu, L. M. and Hudak, G. B., 'United econometric model building using simultaneous transfer function models', in Anderson, O. D. (ed.), *Time Series Analysis: Theory and Practice* **7**, Amsterdam: North-Holland (1985).

Maloney, M. T. and Ireland, M. E., 'Fiscal versus monetary policy', *Jour. of Econometrics*, **13** (1980), 253–66.

Pierce, D. A., 'Least squares estimation in dynamic disturbance time series models', *Biometrika*, **59** (1972), 73–8.

Poskitt, D. S., 'A method for the estimation and identification of transfer functions', *Jour. of Royal Statistics Society, Ser. B*, **51** (1989), 29–46.

Priestley, M. B., 'The frequency domain approach to the analysis of closed-loop systems', in Brillinger, D. R. and Krishnaiah, P. R. (eds), *Handbook of Statistics* **3**, Amsterdam: North Holland (1983).

Priestley, M. B., *Nonlinear and Nonstationary Time Series Analysis*, New York: Academic Press (1988).
Rozanov, Y. A., *Stationary Stochastic Processes*, San Francisco: Holden-Day (1967).
Solo, V., *Time Series Recursions and Stochastic Approximation*, PhD dissertation, Canberra: The Australian National University (1978).
Stoica, P., Söderström, T., Ahlen, A. and Solbrand, G., 'On the convergence of pseudolinear regression algorithm', *Int. Jour. of Control*, **40** (1985), 1429–1444.
Spliid, H., 'A fast estimation method for vector ARMAX models', *Jour. of American Statistical Association*, **78** (1983), 843–9.
Tiao, G. C. and Box, G. E. P., 'Modeling multiple time series with application', *Jour. of American Statistical Association*, **76** (1981), 802–16.
Tiao, G. C. and Tsay, R. S., 'Multiple time series modeling and extended sample cross correlation', *Jour. of Business and Econ. Statistics*, **1** (1983), 43–56.
Tsypkin, Y., *Adaptation and Learning in Automatic Systems*, New York: Academic Press (1971).
Young, P. C. 'Recursive estimation, forecasting, and adaptive control', in Leondes, C. T. (ed.), *Control and Dynamic Systems*, Vol. 30, 119–65, Orlando, FLA: Academic Press (1989).
Young, P. C. and Jakeman, A. J., 'Refined instrumental variables methods of recursive time-series analysis: Part I, SISO systems', *Int. Jour. of Control*, **29** (1979a), 1–30.
Young, P. C. and Jakeman, A. J., 'Refined instrumental variables methods of recursive time-series analysis: Part II, multivariable systems', *Int. Jour. of Control*, **29** (1979b), 621–44.
Young, P. C., Jakeman, A. J. and McMurtrie, R., 'An instrumental variable method for model order identification', *Automatica*, **26** (1980), 281–94.
Zellner, A. and Palm, R., 'Time series analysis and simultaneous equation econometric models', *Jour. of Econometrics*, **2** (1974), 1–17.

Author's biography:
Carlo Grillenzoni is a Lecturer in Statistics in the Department of Economics at the University of Modena. Currently he is concluding his Doctoral Program in Statistics in the Department of Statistics at the University of Padova.

Author's address:
Dr Carlo Grillenzoni, Dipartimento di Economia Politica, Università di Modena, Via Giardini 454, 41100 Modena, Italy.