



ELSEVIER

Journal of Econometrics 73 (1996) 355–376

**JOURNAL OF
Econometrics**

Testing for causality in real time

Carlo Grillenzoni

Architecture University Institute of Venice, Cà Tron, 30125 Venezia, Italy

(Received July 1993; final version received January 1995)

Abstract

A framework for testing in real time (on-line) the statistical significance of the causality between nonstationary random processes is developed. The process representation is that of transfer function (TF-ARMA) models; the causality parameters are prediction error variances and dynamic multipliers; the estimation algorithm is that of recursive nonlinear least squares (RNLS). The basic step is made by analyzing the asymptotic distribution of this estimator under an assumption of stationary, but in operative conditions given by discounting past observations with exponential weights (EW). An empirical example, based on real economic time series, illustrates and checks the method of on-line inference.

Key words: Recursive estimators; Dynamic multipliers; Nonstationary processes; Tests of causality; Significance bands

JEL classification: C22; C32; C52

1. Introduction

Recursive estimators with adaptive implementation are well-established methods for estimating dynamic models whose parameters change over time. As shown in the books of Ljung and Söderström (1983) and Goodwin and Sin (1984), the various recursive algorithms have formal and practical connections; however, they are somewhat different from a statistical viewpoint. While the Kalman filter (KF) assumes linear dynamics for the parameters and updates its covariance matrix with fixed quantities, recursive least squares (RLS) with weighted observations do not assume explicit laws of evolution and their covariance matrix changes adaptively. Weighted RLS are then analogous to modern techniques of nonparametric regression; specifically, the problem of designing the discounting rate of observations is similar to the problem of choice of the *bandwidth* in kernel-type estimators (see Härdle, 1990). In both cases,

a suitable trade-off between bias and variance of the regression function must be achieved.

In this paper we focus on the RLS algorithm with exponentially weighted (EW) observations, i.e., in which the weighting sequence of $\{Z_{t-k}\}$, $0 < k < t$, is given by $\{\lambda^{t-k}\}$, $0 < \lambda < 1$. Unlike rectangular windows, this form of discounting is easy to manage on-line and, owing to its concentration around most recent observations, is powerful in estimating models subject to sudden changes and nonlinear oscillations. More generally, it is possible to show that the EW-RLS includes the KF as a particular case, being the underlying dynamics of parameters conditionally Gaussian (see Grillenzoni, 1994). The statistical behaviour of these algorithms in estimating regression models subject to several conditions of evolution has been investigated by Benveniste (1987), Niedzwiecki (1988), and Gunnarsson and Ljung (1989). These studies have provided theoretical rules for the optimal design of the factor λ . In econometrics, Zellner, Hong, and Min (1990) have developed an original Bayesian design of the EW-RLS and have compared its forecasting ability with that of RLS and KF having an analogous design.

In this article we focus on the transfer function (TF) model of Box and Jenkins (1976), estimated by recursive nonlinear least squares (RNLS) of the Gauss–Newton type. Even though the analysis is developed only under the assumption of constant parameters, complex indicators of causality concerning the *predictive effect* and the *multiplicative impact* of the input on the output are considered. The asymptotic distributions of adaptive statistics of F , χ^2 , T type are investigated, finding that their degrees of freedom crucially depend on the effective sample size of the algorithm given by $(1 + \lambda)/(1 - \lambda)$.

An inferential framework for testing in *real time* the statistical significance of the parameters of time-varying dynamic models can be easily developed with the results of the paper. In an extended numerical application on a data-set published by Lütkepohl (1991) we investigate the relationships of causality between two nonstationary economic processes. However, other important applications are represented by (i) testing for the stability over time of the coefficients of time series models and (ii) detection of ruptures in dynamical systems applied to industrial processes. In all cases, the statistical procedures directly monitor the trajectories of recursive parameter estimates.

2. Off-line analysis

Consider two stationary stochastic processes $\{y_t, x_t\}$ having a cross-covariance function $E(y_t x_{t-k}) = \gamma_{xy}(k)$ null for every $k < b \geq 0$ (i.e., without feedback $y_t \Rightarrow x_t$) and absolutely summable for $k \geq b$ (i.e., y_t, x_t are jointly ergodic). Assuming zero means and gaussian distributions, these features lead to the representation

TF

$$y_t = \frac{(\omega_0 + \omega_1 B + \dots + \omega_s B^s)}{(1 - \delta_1 B - \dots - \delta_r B^r)} x_{t-b} + \frac{(1 + \theta_1 B + \dots + \theta_q B^q)}{(1 - \phi_1 B - \dots - \phi_p B^p)} a_t, \quad a_t \sim \text{IN}(0, \sigma_a^2), \quad (1a)$$

ARMA

$$(1 - \tilde{\phi}_1 B - \dots - \tilde{\phi}_p B^p) y_t = (1 + \tilde{\theta}_1 B + \dots + \tilde{\theta}_q B^q) e_t, \quad e_t \sim \text{IN}(0, \sigma_e^2), \quad (1b)$$

where (A1) $(\delta, \phi, \theta, \tilde{\phi}, \tilde{\theta})(B)$ are stable polynomial in the lag operator B ($Bx_t = x_{t-1}$); (A2) $\omega(B)B^b$ is a nonmonic polynomial with bounded coefficients and delay factor $0 \leq b < \infty$; finally, (A3) the pairs (ω, δ) , (θ, ϕ) , $(\tilde{\theta}, \tilde{\phi})$ are relatively prime. It is quite clear that under the stated assumptions the parametric covariance functions of $\{y_t, x_t\}$ decay exponentially to zero (for details, see Box and Jenkins, 1976, Part III).

The two main elements that characterize the causal action $x_t \Rightarrow y_t$ are the predictive effect and the multiplicative impact. With respect to the representation (1) these are summarised by the parameters

$$\Delta = (\sigma_e^2 - \sigma_a^2) = E[y_t^2 | y_{t-i}; i > 0] - E[y_t^2 | y_{t-i}, x_{t-j}; i > 0, j \geq 0], \quad (2a)$$

$$g = \left(\frac{\omega_0 + \omega_1 + \dots + \omega_s}{1 - \delta_1 - \dots - \delta_r} \right) = \sum_{k=0}^{\infty} v_k, \quad v_k = \sum_{j=1}^r \delta_j v_{k-j} + \omega_k, \quad (2b)$$

where Δ is the reduction in the variance of the one-step-ahead prediction error, $\{v_k\}$ is the sequence of dynamic multipliers, and g is the steady-state gain (that is, the global change in y_t yielded by an unitary increment of x_t).

Both (2a) and (2b) are indexes of causality, but with different operative meanings. The gain is a typical parameter of control, and when $|g| > 0$, we also have $\Delta > 0$; the converse, however, is not necessarily true. Examples in this sense are provided by response functions of the type $v(B) = (\omega_0 - \omega_1 B^1)B^b$ with $\omega_0 \approx \omega_1$, that may be encountered in many applications (see Grillenzoni, 1991a). More generally, the tendency to yield $g \approx 0$ concerns sequences $\{v_k\}$ which decay rapidly and nonmonotonically, i.e., corresponding to polynomials $\delta(B)$ with negative or complex roots far from the unit circle. These remarks about Δ, g enable us to qualify the concept of causality of Wiener–Granger (see Granger and Newbold, 1986) that was originally based on (2a) alone. In particular, from the viewpoint of economic policy it may be necessary to establish if an input x_t has a positive impact on the output y_t , and not whether x_t may improve the predictions of y_t .

Given the complexity of the likelihood function associated with (1), a suitable estimator for the vector $\beta' = [\delta_1 \dots \delta_r, \omega_0, \omega_1 \dots \omega_s, \phi_1 \dots \phi_p, \theta_1 \dots \theta_q]$ is

that of nonlinear least squares. Denoting by (N, k) the number of observations and iterations it becomes

NLS

$$\hat{\beta}_N(k + 1) = \hat{\beta}_N(k) + \alpha_k \left[\sum_{t=1}^N \hat{\xi}_t(k) \hat{\xi}'_t(k) \right]^{-1} \left[\sum_{t=1}^N \hat{\xi}_t(k) \hat{a}_t(k) \right], \tag{3a}$$

$$\xi_t(\beta)' = \left[\frac{\phi(B)}{\theta(B)\delta(B)} (m_{t-1} \dots m_{t-r}, x_{t-b} \dots x_{t-b-s}); \right. \\ \left. \frac{1}{\theta(B)} (n_{t-1} \dots n_{t-p}, a_{t-1} \dots a_{t-q}) \right], \tag{3b}$$

where $m_t = [\omega(B)/\delta(B)]x_{t-b}$, $n_t = [\theta(B)/\phi(B)]a_t$ are auxiliary regressors, $\xi_t(\beta) = -\partial a_t(\beta)/\partial \beta$ is the gradient, and $0 < \alpha_k < \infty$ is the stepsize. Clearly, under assumptions (A), the process $\{\xi_t\}$ is still stationary and ergodic; this provides the basis for a theorem established by Pierce (1972) and reconsidered in Poskitt (1989).

Theorem 1. Let $\{x_t, y_t\}$ be ergodic processes that satisfy the stable and identified representation (1); then the estimator (3) is consistent for β and, more generally,

$$N^{1/2} [\hat{\beta}_N(k) - \beta] \xrightarrow{L} N[0, E(\xi_t \xi'_t)^{-1} \sigma^2] \text{ as } k, N \rightarrow \infty,$$

where $E(\xi_t \xi'_t)$ is block-diagonal for the independence of $\{m_t, x_t\}$ and $\{n_t, a_t\}$.

Since quadratic transformations of ergodic processes are also ergodic, the above result follows by combining the limit theorems for the sums of ergodic sequences with the properties of convergence (minimization) of the algorithm (3) (see Grillenzoni, 1991b). Finite-sample properties have not been widely investigated in the statistical literature, but they should be similar to those established for ARMA models.

Given the parsimony of representation (1), with (3) we may define efficient and asymptotically unbiased estimators for the causality parameters (2), namely

$$\hat{\Delta}_N(k) = [\hat{\sigma}_e^2(k) - \hat{\sigma}_a^2(k)], \quad \hat{g}_N(k) = \sum_{i=0}^s \hat{\omega}_i(k) \Big/ \sum_{j=0}^r \hat{\delta}_j(k). \tag{4}$$

As a consequence of the theorem they enjoy optimal properties, in particular we have:

Corollary 1. Under the same conditions as Theorem 1, the estimator (4) is consistent for the gain $g = g(\beta)$ and, more generally,

$$N^{1/2} [\hat{g}_N(k) - g] \xrightarrow{L} N \left[0, \left(\frac{\partial g}{\partial \beta} \right)' E(\xi_t \xi'_t)^{-1} \sigma^2 \left(\frac{\partial g}{\partial \beta} \right) \right] \text{ as } k, N \rightarrow \infty.$$

Proof. Since $g(\cdot)$ is continuous, by Slutsky theorem we have that

$$(\hat{\beta}_N \xrightarrow{P} \beta) \Rightarrow (\hat{g}_N \xrightarrow{P} g) \text{ as } k, N \rightarrow \infty.$$

To prove the distribution, we consider a Taylor expansion of $\hat{g}(\cdot)$ in β ,

$$g(\hat{\beta}_N) = g(\beta) + \left(\frac{\partial g}{\partial \beta}\right)' (\hat{\beta}_N - \beta) + O_p(N^{-1}), \tag{5}$$

where $|\hat{\beta}_N - \beta| = O_p(N^{-1/2})$ for the consistency of $\hat{\beta}_N(k)$. Now, multiplying by $N^{1/2}$ we get

$$N^{1/2}(\hat{g}_N - g) = N^{1/2}(\hat{\beta}_N - \beta)' \left(\frac{\partial g}{\partial \beta}\right) + O_p(N^{-1/2}).$$

The above equation can also be rewritten as $y_N = x'_N c + o_p(1)$, where c is a constant vector. Since $x_N \xrightarrow{L} x$ Gaussian, from the Cramer-Wold lemma we have $y_N \xrightarrow{L} x'c$ with the same distribution; the asymptotic dispersion being $c'E(xx')c$.

While Corollary 1 provides the necessary background for making inference on the steady state gain g , it is well-known that powerful procedures for testing the significance of Δ involve statistics of F -type. Indeed, under the conditions of the theorem, in particular that of independent normal (IN) disturbances, and the null hypothesis $H_0: \sigma_e^2 = \sigma_a^2$, for N sufficiently large (but finite) we have

$$\hat{F}_N(k) = \left[\frac{(N - b)\hat{\Delta}_N(k)/m}{(N - b)\hat{\sigma}_a^2(k)/(N - b - m - n)} \right] \xrightarrow{L} F(m, N - b - m - n), \tag{6}$$

as $k \rightarrow \infty$. The quantities $m = (r + s + 1)$ and $n = (p + q)$ provide the degrees of freedom.

3. On-line analysis

In this section we begin to extend the above results to models estimated by recursive algorithms with discounted observations. As shown in Ljung and Gunnarsson (1990), this implementation is *sufficient* for tracking time-varying parameters, although it is not generally optimal. In order to simplify the analysis and to obtain statistical results as rigorous as possible, we commence with a linear model of ARX type,

$$\begin{aligned} y_t &= (\phi_{1t}y_{t-1} + \dots + \phi_{pt}y_{t-p} + \omega_{0t}x_{t-b} + \dots + \omega_{st}x_{t-b-s}) + a_t \\ &\stackrel{\text{def}}{=} \beta_t' z_t + a_t, \quad a_t \sim \text{IN}(0, \sigma_a^2), \end{aligned} \tag{7}$$

where $z'_t = [y_{t-1} \dots, \dots x_{t-b-s}]$ is the vector of ‘regressors’. For this model, the RLS algorithm with exponentially weighted observations can be defined in two equivalent ways (Brown, Durbin, and Evans, 1975; Ljung and Söderström, 1983):

$$\hat{\beta}_t = R_t^{-1} \sum_{\tau=1}^t (z_\tau \lambda^{t-\tau} y_\tau), \quad \hat{\beta}_t = \hat{\beta}_{t-1} + R_t^{-1} z_t \tilde{a}_t, \tag{8a}$$

$$R_t = \sum_{\tau=1}^t (z_\tau \lambda^{t-\tau} z'_\tau), \quad R_t = \lambda \cdot R_{t-1} + z_t z'_t, \tag{8b}$$

$$S_t = \sum_{\tau=1}^t \lambda^{t-\tau} (y_\tau - \hat{\beta}'_t z_\tau)^2, \quad S_t = \lambda \cdot S_{t-1} + \tilde{a}_t \tilde{a}_t, \tag{8c}$$

$$\tilde{a}_t = (y_t - z'_t \hat{\beta}_{t-1}), \quad \hat{a}_t = \lambda^{-1} [\lambda + z'_t R_{t-1}^{-1} z_t] \tilde{a}_t, \tag{8d}$$

where \tilde{a}_t are prediction errors and $\hat{a}_t = (y_t - \hat{\beta}'_t z_t)$ are recursive residuals; S_t is the weighted sum of squared residuals, generated with the latest parameter estimate.

The properties of this algorithm have been investigated in depth only under an assumption of constant parameters, a condition of ergodicity for the input, namely,

$$(B1) \quad \beta_t = \beta, \quad \sigma_t = \sigma, \quad (B2) \quad \sum_k |\gamma_{xx}(k)| < \infty, \tag{9}$$

and $\lambda \equiv 1$. The general conclusion was that the RLS algorithm is asymptotically equivalent to the OLS estimator. In the following we briefly comment on (8) when $\lambda < 1$:

1) The two versions of the algorithm differ only at the computational level, provided that both initial values β_0 and R_0^{-1} are zero. The right one clearly indicates that in order to track the changes $(\beta_t - \beta_{t-1})$, the condition $R_t < \infty$ (i.e., $\lambda < 1$) must hold uniformly. Since β_0 and R_0 have a significant role in the tracking (see Grillenzoni, 1994), in the sequel we shall only refer to the recursive implementation $\hat{\beta}_t = \hat{\beta}_{t-1} + R_t^{-1} z_t \tilde{a}_t$.

2) The expression (8a) on the left can be formally obtained by minimizing the functional $S_t(\beta) = \sum_1^t \lambda^{t-\tau} a_\tau^2(\beta)$ with respect to β . Even though S_t on the left is only a function of the latest estimate $\hat{\beta}_t$, the version on the right (derived in Ljung and Söderström, 1983, p. 434) shows that $\hat{\sigma}_t^2 = S_t / \sum_1^t \lambda^{t-\tau}$ is suitable for tracking the noise variances $\{\sigma_\tau^2\}$ in the model (7).

3) The standardized prediction errors $\hat{u}_t = (\tilde{a}_t \cdot \hat{a}_t)^{1/2} = \tilde{a}_t [\lambda / (\lambda + z'_t R_{t-1}^{-1} z_t)]^{1/2}$ have been emphasized by Brown, Durbin, and Evans (1975) and Dufour (1982) by showing that, under (9) and $\lambda = 1$, their conditional distribution is $(\hat{u}_t | z_t, z_{t-1} \dots) \sim \text{IN}(0, \sigma^2)$. However, given the relationship (8d), as $\lambda \rightarrow 1$ these properties hold asymptotically even for the terms $\{\tilde{a}_t, \hat{a}_t\}$.

The central purpose of this paper is that of making *on-line inference*, in the sense of deriving standard errors for the recursive estimates $\{\hat{\beta}_t\}$ and testing for their statistical significance. In this context, the basic null hypothesis for the parameters is given by $H_0: \beta_t = \beta$, which belongs, in a strict sense, to the condition of stationary $\beta_t = \beta$. It is necessary, therefore, to analyse the distribution of estimator (8a) under the assumptions (9) and to evaluate the action of the discounting rate λ on the mean squared error $E \|\hat{\beta}_t - \beta\|^2$. The next lemma quantifies the effect of the factor $(1 - \lambda)$ on basic statistics involved in the EW-RLS algorithm; the proof is given in Appendix 1.

Lemma 1. Let $\{y_t, x_t\}$ be Gaussian ergodic sequences, with covariance functions (auto and cross) decaying exponentially at a rate $0 < \mu \leq \lambda < 1$; then the sample covariance

$$\hat{\gamma}_t(0) = (1 - \lambda) \sum_{\tau=1}^t \lambda^{t-\tau} x_\tau y_\tau \xrightarrow{P} \gamma_{xy}(0) + O_p(1 - \lambda)^{1/2} \quad \text{as } t \rightarrow \infty.$$

Since rigorous asymptotic results for the RLS algorithm applied to ARX models are only available under consistency conditions, we need to let $\lambda \rightarrow 1$. In order to avoid the convergence to the OLS estimator, however, we have to rely on a double limit operation $\lim_{\lambda \rightarrow 1} \lim_{t \rightarrow \infty} f(t, \lambda)$ in which there is not exchangeability of the order, and the accumulation point of \lim_{λ} is external to the interval of definition of $\lambda \in (0, 1)$ open. In practice, since $(1 - \lambda)$ has the same role as does the *bandwidth* in a kernel-type estimator, analogously to the analysis of non-parametric regression one should have $\lim_{\lambda \rightarrow 1} \lim_{t \rightarrow \infty} [1/(1 - \lambda)t] = 0$ (see Härdle, 1990). We now reconsider a result outlined by Niedzwiecki (1988).

Proposition 1. Given the linear model (7), under the stationary assumption (9) the recursive estimator (8) is nonconsistent when $\lambda < 1$, and more generally,

$$(1 - \lambda)^{-1/2} [\hat{\beta}_t - \beta] \xrightarrow{L} N[0, \frac{1}{2} E(z_t z_t')^{-1} \sigma^2] \quad \text{as } t \rightarrow \infty, \lambda \rightarrow 1, \quad (10a)$$

i.e.,

$$\lim_{t \rightarrow \infty} E[(\hat{\beta}_t - \beta)(\hat{\beta}_t - \beta)'] = \left(\frac{1 - \lambda}{1 + \lambda} \right) E(z_t z_t')^{-1} \sigma^2 + O(1 - \lambda)^{3/2}. \quad (10b)$$

Proof. It may be easily seen that (8a) can be rewritten in the form $(1 - \lambda)R_t(\hat{\beta}_t - \beta) = (1 - \lambda) \sum_{\tau=1}^t \lambda^{t-\tau} z_\tau a_\tau$. Now, by Lemma 1 the terms $(1 - \lambda)R_t - E(z_t z_t')$ and $[\hat{\beta}_t - \beta_t]$ are both of order $O_p(1 - \lambda, 1/t)^{1/2}$; thus we have

$$(1 - \lambda)^{-1/2} [\hat{\beta}_t - \beta] = E(z_t z_t')^{-1} (1 - \lambda)^{1/2} \sum_{\tau=1}^t \lambda^{t-\tau} z_\tau a_\tau - O_p[(1 - \lambda)^{1/2}, 1/t]. \quad (11)$$

The asymptotic normality and the related dispersion can be obtained by using calculations as in the Appendix 1; specifically, defining $E[(z_t')^2] = E[z_t z_t']$ we may get

$$\begin{aligned} & \lim_{\lambda \rightarrow 1} \lim_{t \rightarrow \infty} E \left[\left(\left((1 - \lambda)^{1/2} \sum_{\tau=1}^t \lambda^{t-\tau} z_\tau a_\tau \right)^2 \right) \right] \\ & = \lim_{\lambda \rightarrow 1} \lim_{t \rightarrow \infty} (1 - \lambda) \sum_{\tau=0}^{t-1} \lambda^{2\tau} E(z_\tau z_\tau') \sigma^2 = \frac{1}{2} E(z_t z_t') \sigma^2. \end{aligned}$$

The last follows by noting that $\lim_{\lambda \rightarrow 1} (1 + \lambda) = (1 + 0.\bar{9}) \equiv 2$ since $0.\bar{9} = 3 \cdot 0.\bar{3}$ and $0.\bar{3} = \frac{1}{3}$.

Proposition 1 highlights the trade-off between tracking capability and estimation accuracy induced by the factor λ ; in particular, as $\lambda \rightarrow 0$ the speed of adaptation of $\hat{\beta}_t$ increases but its MSE efficiency declines. Since as $\lambda \rightarrow 1$ the RLS algorithm converges in probability to the OLS estimator, a way for explaining the term $\frac{1}{2}$ in the dispersion of (10a) is to assume that λ varies more slowly than t , and its interval of definition does not contain the bound 1. Finally, the expression (10b) may be used in sampling form, by replacing $E(z_t z_t')$ by $(1 - \lambda)R_t$, provided that $\{\gamma_{zz}(k)\}$ decays at a rate faster than $\lambda < 1$ (see Lemma 1).

Proposition 2. Given the linear model (7), under the stationary assumption (9) the prediction errors and the recursive residuals (8d) behave like

$$\tilde{a}_t \xrightarrow{L} N[O(1 - \lambda)^{1/2}, \sigma^2 + O(1 - \lambda)], \quad E(\tilde{a}_t \tilde{a}_{t-k}) \rightarrow O(1 - \lambda) \quad \text{as } t \rightarrow \infty. \tag{12}$$

Proof. By definition, we have the orthogonal decomposition $\tilde{a}_t = z_t'[\beta - \hat{\beta}_{t-1}] + a_t$, from which it follows that $E(\tilde{a}_t) = O(1 - \lambda, 1/t)^{1/2}$; moreover,

$$\begin{aligned} E(\tilde{a}_t^2) &= \sigma^2 + E[z_t'(\hat{\beta}_{t-1} - \beta)]^2 \\ &= \sigma^2 + E\{z_t' E[(\hat{\beta}_{t-1} - \beta)(\hat{\beta}_{t-1} - \beta)' | z_t] z_t\}, \end{aligned}$$

and by (10b) we get

$$\lim_{t \rightarrow \infty} E(\tilde{a}_t^2) = \sigma^2 + E[z_t' E(z_t z_t')^{-1} z_t] \left(\frac{1 - \lambda}{1 + \lambda} \right) \sigma^2 + E(z_t z_t') O(1 - \lambda)^{3/2}.$$

Similarly, using the asymptotic expression of $E[(\hat{\beta}_t - \beta)(\hat{\beta}_{t-k} - \beta)']$ we have

$$\begin{aligned} \lim_{t \rightarrow \infty} E(\tilde{a}_t \tilde{a}_{t-k}) &= \left[\left(\frac{1 - \lambda}{1 + \lambda} \right) \lambda^k - (1 - \lambda) \lambda^{k-1} \right] \cdot E[z_t' E(z_t z_t')^{-1} z_{t-k}] \sigma^2 \\ &\quad + E(z_t' z_t) O(1 - \lambda)^{3/2}, \end{aligned}$$

which may be proved even for $\{\hat{a}_t\}$. In practice, as $t \rightarrow \infty$ and $\lambda \rightarrow 1$, the recursive terms $\{\tilde{a}_t, \hat{a}_t\}$ become best linear unbiased spherical (BLUS) residuals (see Dufour, 1982).

Under the condition of stationary (9) it may be checked that asymptotically unbiased estimators for the residual variance and the dispersion of $\hat{\beta}_t$ are given by

$$\hat{\sigma}_t^2 = \left[\sum_{\tau=1}^t \lambda^{t-\tau} (\tilde{a}_\tau \hat{a}_\tau) / \sum_{\tau=1}^t \lambda^{t-\tau} \right] \rightarrow (1 - \lambda) S_t, \quad \hat{V}_t = \left[\frac{1}{(1 + \lambda)} \hat{\sigma}_t^2 R_t^{-1} \right]. \tag{13}$$

Other estimators of σ^2 could be introduced, such as $\sum_1^t (\tilde{a}_\tau \hat{a}_\tau) / t$ or simply $(\tilde{a}_t \hat{a}_t)$. However, $\hat{\sigma}_t^2$ in (13) is a suitable compromise, consistent with the algorithm (8) and proper for tracking $\{\sigma_t^2\}$ if this does not have sudden changes.

4. On-line inference

In this section we extend previous results to the recursive nonlinear least squares (RNLS) estimator of the TF-ARMA model and we provide distributions for the on-line statistics of causality. The recursive version of (3) with weighted gradients $\{\lambda^{N-t} \hat{\xi}_t(k)\}$ may be obtained by equating the number of iterations and the number of processed data ($k = N$) = t , and proceeding as in the derivation of (8) (see Grillenzoni, 1991b). The resulting algorithm minimizes the weighted functional $S_t(\beta) = \sum_{\tau=1}^t \lambda^{t-\tau} a_\tau^2(\beta)$; leaving aside computational details for $\hat{\xi}(t)$ and the 'regressors' $\hat{z}(t)$, its basic expression is given by

EW-RNLS

$$R(t) = \lambda \cdot R(t - 1) + \hat{\xi}(t) \hat{\xi}(t)', \tag{14a}$$

$$\tilde{a}(t) = y_t - \hat{z}(t)' \hat{\beta}(t - 1), \tag{14b}$$

$$\hat{\beta}(t) = \hat{\beta}(t - 1) + R(t)^{-1} \hat{\xi}(t) \tilde{a}(t), \tag{14c}$$

$$\hat{a}(t) = y_t - \hat{\beta}(t)' \hat{z}(t), \tag{14d}$$

$$S(t) = \lambda \cdot S(t - 1) + \tilde{a}(t) \hat{a}(t), \tag{14e}$$

where $\hat{z}(t)' = [\hat{m}(t - 1) \dots \hat{m}(t - r), x_{t-b} \dots x_{t-s-b}, \hat{n}(t - 1) \dots \hat{n}(t - p), \hat{a}(t - 1) \dots \hat{a}(t - q)]$. The corresponding on-line estimators for the indexes (2) are given by

$$\hat{\Delta}(t) = (1 - \lambda)[S_o(t) - S_a(t)], \quad \hat{g}(t) = \left[\sum_{i=0}^s \hat{\omega}_i(t) / \sum_{j=0}^r \hat{\delta}_j(t) \right], \tag{15}$$

and in the following we investigate their asymptotic distributions, or that of related test statistics. Exact expressions are difficult to achieve, even under the assumption (9).

Proposition 3. Let $\{x_t, y_t\}$ be ergodic processes and (I) their stable and identified representation; then the sequences $\{\tilde{a}(t), \hat{a}(t)\}$ computed in (14) are such that

$$S(t) = \left[\sum_{\tau=1}^t \lambda^{t-\tau} \tilde{a}(\tau) \hat{a}(\tau) \right] \xrightarrow{P} W_\lambda + O_p(1 - \lambda) \quad \text{as } t \rightarrow \infty, \quad (16a)$$

where

$$W_\lambda \stackrel{L}{\approx} \left[\frac{\sigma^2}{(1 + \lambda)} \cdot \chi^2 \left(\frac{1 + \lambda}{1 - \lambda} \right) \right] \quad (\text{in law}). \quad (16b)$$

Proof. For simplicity, let be $\sigma = 1$. We first recall that under conditions of ergodicity and identification the recursive estimator (14c) with $\lambda \equiv 1$ is consistent for β (Ljung and Söderström, 1983). Thus, by Slutsky theorem we have $\tilde{a}(t), \hat{a}(t) = a_t(\beta) + O_p(1 - \lambda, 1/t)^{1/2}$ and $[\tilde{a}(t)\hat{a}(t)] = u_t + O_p(1 - \lambda, 1/t)$, where $u_t \sim \chi^2(1)$. If λ is close to unity, the variates $\hat{u}(t)^2 = [\tilde{a}(t)\hat{a}(t)]$ are also asymptotically independent.

Now, for t sufficiently large we may write $S(t) = \sum_{\tau=1}^t \lambda^{t-\tau} u_\tau + O_p(1 - \lambda)$ and we must find the limiting distribution of a linear combination of random variables $u_\tau \sim \chi^2(1)$. This analysis has some connection with the distribution of quadratic forms; however, this has been solved *exactly* only for particular cases of finite forms (see Solomon and Stephens, 1977; Mathai, 1983). Given the asymptotic structure of our problem it is convenient to utilize an approximate solution. A simple approach, particularly useful in the case of exponential weights, is that proposed by Patnaik (1949) which makes the first two moments of $S(t)$ agree with those of a known random variable W . This technique is also discussed in detail by Johnson and Kotz (1970, p. 165).

Thus, referring to $W = cU$, with c constant and $U \sim \chi^2(n)$, and equating mean and variance of S, W , we have

$$\sum_{\tau=1}^t \lambda^{t-\tau} = cn, \quad 2 \sum_{\tau=1}^t (\lambda^{t-\tau})^2 = 2c^2n.$$

Solving for c and n we get

$$c = \frac{\sum_{\tau=1}^t (\lambda^2)^{t-\tau}}{\sum_{\tau=1}^t \lambda^{t-\tau}}, \quad n = \left(\frac{\sum_{\tau=1}^t \lambda^{t-\tau}}{\sum_{\tau=1}^t (\lambda^2)^{t-\tau}} \right)^2$$

which belong to the distribution of $S(t) \stackrel{L}{\approx} c \cdot \chi^2(n)$. Now, the advantage of this approach lies in the fact that suitable asymptotic expressions of the unknown coefficients are available, in particular $c \rightarrow 1/(1 + \lambda)$ and $n \rightarrow (1 + \lambda)/(1 - \lambda)$ as

$t \rightarrow \infty$. These prove (16b) and provide approximate mean and variance of $S(t)$, namely $1/(1 - \lambda)$ and $2/(1 - \lambda^2)$. Finally, by the central limit theorem, we have $S(t) \xrightarrow{L} N(\cdot)$ as $t \rightarrow \infty$ and $\lambda \rightarrow 1$; thus, the order of approximation in (16b) is $O(1 - \lambda)$.

Remark 1. For values of λ , t sufficiently large we may derive, as a corollary of the proposition, the approximate (in law) F -statistic corresponding to the Δ -index,

$$\hat{F}_1(t) = \frac{[S_e(t) - S_a(t)]/m}{(1 - \lambda)S_a(t)/2} \xrightarrow{L} F \left[m; \left(\frac{1 + \lambda}{1 - \lambda} \right) - (m + n) \right], \tag{17a}$$

where $m = (r + s + 1)$, $n = (p + q)$, and $2 \approx [(1 + \lambda) - (m + n)(1 - \lambda)]$ are suitable 'degrees of freedom'. Similarly, for the F -statistic which is used for testing the hypothesis $H_0: \sigma_e = \sigma_a$, we may have

$$\hat{F}_2(t) = \frac{S_e(t)}{S_a(t)} \xrightarrow{L} F \left[\left(\frac{1 + \lambda}{1 - \lambda} \right); \left(\frac{1 + \lambda}{1 - \lambda} \right) \right]. \tag{17b}$$

The degrees of freedom of these distributions are consistent with the fact that, for decreasing values of λ , the variability of the recursive estimates $S(t)$, $\hat{\beta}(t)$ tends to increase. This means that larger critical values are needed to deal with greater sampling errors; in practice $f = (1 + \lambda)/(1 - \lambda)$ must decrease. It is worth noting that the value of f is such that $\lambda^f \approx 0$, hence it represents the effective sample size of the estimates. With respect to the most common choices of the weighting factor, we have $\lambda = (0.98, 0.97, 0.96) \Rightarrow f = (99, 66, 49)$, which correspond to medium samples.

Next result extends Proposition 1 and Corollary 1 to RNLS estimators.

Proposition 4. Let $\{x_t, y_t\}$ be ergodic processes and (1) their identified representation; then for the estimators $\hat{\beta}(t)$ (14c) and $\hat{g}(t)$ (15) we asymptotically have

$$(1 - \lambda)^{-1/2} [\hat{\beta}(t) - \beta] \xrightarrow{L} N[0, \frac{1}{2} E(\xi, \xi')^{-1} \sigma^2] \text{ as } t \rightarrow \infty, \lambda \rightarrow 1, \tag{18a}$$

i.e.,

$$\lim_{t \rightarrow \infty} E[\hat{g}(t) - g(\beta)]^2 = \left(\frac{\partial g}{\partial \beta} \right)' \left(\frac{1 - \lambda}{1 + \lambda} \right) E(\xi, \xi')^{-1} \sigma^2 \left(\frac{\partial g}{\partial \beta} \right) + O(1 - \lambda)^{3/2}. \tag{18b}$$

Proof. We recall that under conditions of ergodicity and identification the recursive estimator (14c) with $\lambda \equiv 1$ is asymptotically equivalent to the iterative version (3). This means that its statistical properties are summarized by

Theorem 1, and as $t \rightarrow \infty, \lambda \rightarrow 1$, the Slutsky theorem implies

$$[\hat{\beta}(t) \xrightarrow{P} \beta] \Rightarrow [\hat{\xi}(t) \xrightarrow{P} \xi_t(\beta)] \Rightarrow [(1 - \lambda)R(t) \xrightarrow{P} E(\xi_t, \xi_t')]. \tag{19}$$

The result (18a) can now be proved by focusing on the loss function $\hat{J}(t) = \sum_{\tau=1}^t \lambda^{t-\tau} \hat{\xi}_\tau(t) \hat{a}_\tau(t)$, which corresponds to $J_t(\beta) = \frac{1}{2} \partial S_t(\beta) / \partial \beta$ evaluated in $\hat{\beta}(t)$. While $J_t(\beta) \rightarrow 0$ in probability as $t \rightarrow \infty$, by the minimization properties of $\hat{\beta}(t)$ we must have $\hat{J}(t) \equiv 0$ for every $t < \infty$. Hence, taking a Taylor expansion of $\hat{J}(t)$ around β and multiplying by $(1 - \lambda)$ we get

$$\begin{aligned} & (1 - \lambda) \sum_{\tau=1}^t \lambda^{t-\tau} \xi_\tau a_\tau - (1 - \lambda) \sum_{\tau=1}^t \lambda^{t-\tau} (\xi_\tau \xi_\tau' + \Xi_\tau a_\tau) [\hat{\beta}(t) - \beta] \\ & = -O_p[(1 - \lambda)^2, 1/t], \end{aligned} \tag{20}$$

where $\Xi_t = -\partial \xi_t / \partial \beta'$ is the Hessian. Notice that $\{\Xi_t\}$ is a matrix process which is still ergodic under assumptions (A); moreover, as in (3b), it is only a function of the past events $\{x_{t-k}, a_{t-k}\}$. Thus, having $E(\Xi_t a_t) = 0$ for all t , from (19) we may get $(1 - \lambda) \sum_{\tau=1}^t \lambda^{t-\tau} (\xi_\tau \xi_\tau' + \Xi_\tau a_\tau) = E(\xi_t \xi_t') + O_p(1 - \lambda, 1/t)^{1/2}$, and substituting this in the expansion (20) provides us an expression similar to (11). Finally, proceeding as in the proof of Proposition 1, we may obtain (18a).

The proof of (18b) follows by noting that $\hat{g}(t) \xrightarrow{P} g(\beta)$ as $t \rightarrow \infty, \lambda \rightarrow 1$, and proceeding as in the proof of Corollary 1 with $\{k = N\} = t$. In particular, the basic elements are the expansion $[\hat{g}(t) - g(\beta)] = (\partial g / \partial \beta)' [\hat{\beta}(t) - \beta] + O_p(1 - \lambda, 1/t)$ and (18a).

Remark 2. Under the assumption of stationarity, an asymptotically unbiased estimator for the dispersion (18b) can be obtained from the on-line statistics $R(t), S(t)$,

$$\hat{V}(t) = \left[\frac{1}{(1 + \lambda)} \hat{\sigma}(t)^2 R(t)^{-1} \right], \quad \hat{\sigma}(t)^2 = (1 - \lambda)S(t). \tag{21}$$

It is worth noting that if $\hat{v}_{ii}(t)$ are the diagonal elements of $\hat{V}(t)$, then using Propositions 3 and 4 jointly, for λ close to unity we have the approximate (in law) T -statistics

$$\left[\frac{\hat{\beta}_i(t) - \beta_i}{\sqrt{\hat{v}_{ii}(t)}} \right] \underset{L}{\approx} T \left(\frac{1 + \lambda}{1 - \lambda} \right) \text{ as } t \rightarrow \infty. \tag{22}$$

Also in this case, confidence intervals constructed with (22) are 'aware' of the trade-off between tracking capability and estimation accuracy involved by λ . In fact, when λ decreases, T values increase for a given confidence level. Hence the probability of rejecting $H_1: \beta_t \neq 0$ when this is false (i.e., when $\hat{\beta}_t \neq 0$ for the action of \tilde{a}_t) is augmented. We may then conclude that under the assumption

$\beta_t = \beta$, the inferential procedures with recursive statistics (14) have the same structure as those developed in the iterative case. Since the effect of $\lambda < 1$ is similar to that induced by $t \ll \infty$, previous asymptotic expressions may be used, with further approximation, for finite samples. Substantial problems are represented by the computation of the degrees of freedom and by the possible serial correlation of recursive errors. A value of $\lambda \in [0.95 \div 0.99]$ is a suitable choice in the case of slowly time-varying parameters.

5. A numerical application

In this section we illustrate the on-line inferential framework discussed previously on a real data-set published by Lütkepohl (1991, p. 505). It consists of monthly observations from two financial processes: Y = short-term interest rate, X = long-term interest rate, for West Germany in the period t = January 1960 to December 1987 ($N = 336$). As is expected from the Keynesian economic theory, the relation of causality should be $X_t \Rightarrow Y_t$, just because the variable X is 'controlled' by monetary authorities via prime rate and open market operations.

Original data are displayed in Fig. 1, showing a clear situation of non-stationarity in covariance; stationarity in mean may be achieved with a first

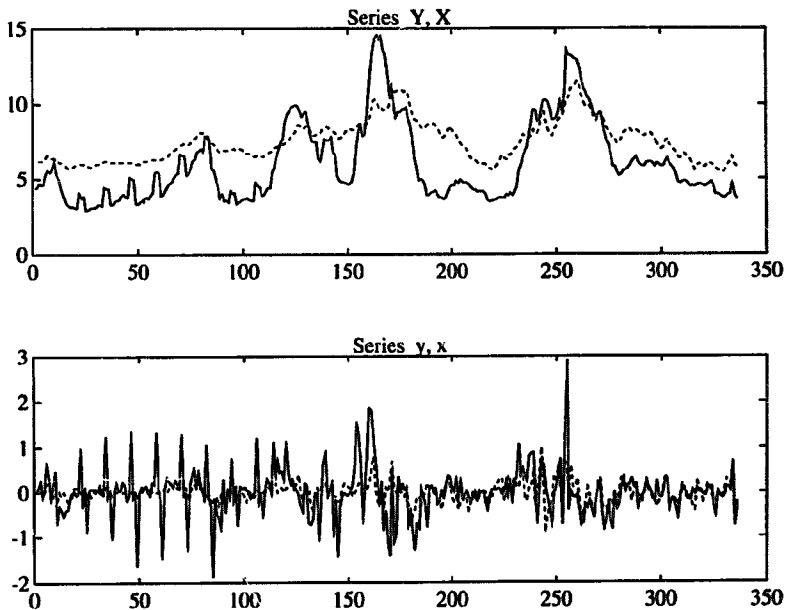


Fig. 1. Plot of series Y , y (—) and X , x (---) in levels and differences.

Table 1
Sample correlation functions and Ljung-Box statistics

Function	0	1	2	3	4	5	6	7	8	9	10	11	12	Q(24)
ACF(y)	1	0.32	0.19	0.03	0.12	0.09	0.07	-0.03	-0.02	-0.16	-0.03	-0.05	0.29	138.3
ACF(x)	1	0.48	0.08	-0.01	0.00	0.03	0.00	-0.05	0.01	0.01	0.06	0.09	0.08	104.4
CCF(y, x)	0.47	0.32	0.14	0.19	0.15	0.09	0.01	-0.08	0.07	-0.11	-0.10	-0.01	0.12	75.1
CCF(x, y)	0.47	0.20	0.11	0.02	-0.01	0.05	0.02	0.03	0.05	0.08	0.11	0.15	0.12	37.5

difference $\{y_t, x_t\} = (1 - B)\{Y_t, X_t\}$. The usefulness and the admissibility of this transformation is demonstrated in Appendix 2, which provides the analysis of the series in levels.

Table 1 reports the sample correlation functions, auto (ACF) and cross (CCF), of the series $\{y_t, x_t\}$ and the Ljung–Box statistics. Following the analysis of Box and Jenkins (1976, p. 349) the TF model (1a) has orders $(r, s, b) = (1, 0, 0)$ and $(p, d, q) = (1^{12}, 1, 1)$, where the AR operator is of seasonal type: $(1 - \phi B^{12})$. Table 2 reports the parameter estimates and the statistics of causality; with $\hat{\Delta} = -18\%$ and $\hat{g} = +1.4$ it confirms the relationship $X_t \Rightarrow Y_t$ found in Appendix 2.

Algorithm (14) was implemented with the coefficients $\lambda = 0.98$ and $R(0) = I/0.18$, which have yielded uniform and mild variability of recursive estimates—that is, suitable trade-off between tracking and accuracy on the entire sampling interval. Moreover, (14c) was initialized with $\hat{\beta}(0) = \hat{\beta}_N(k)$, the off-line estimates of Table 2. Fig. 2 shows the trajectories of the recursive estimates of $(\omega_0, \delta_1, \theta_1, \phi_{12})$, together with their *critical values* implied by (22). These are given by $T_{\alpha/2}(f) \cdot \sqrt{\hat{v}_{ii}(t)}$, where the confidence level was chosen as $(1 - \alpha) = 95\%$ and $f = 99$ are the degrees of freedom.

Despite the moderate values assigned to the tracking coefficients and the transformation of the original series, these graphs show the strongly nonstationary nature of the underlying economic processes. The greatest change is in the noise component $n_t = \phi_t^{-1}(B)\theta_t(B)a_t$, where the seasonal AR filter is gradually replaced by the MA one. This may be interpreted as the tendency for short-term variables, like speculative capital movements, to prevail in financial markets. Finally, while the estimates of the transfer ω_0 are uniformly greater than their critical values, those of the dynamics δ_1 are locally nonsignificant.

Fig. 3 shows the standardized prediction errors $\hat{u}(t) = [\hat{a}(t)\hat{a}(t)]^{1/2}$ together with their 95% confidence intervals $\pm 2 \cdot \hat{\sigma}(t)$ implied by Proposition 2. Apart from few outliers, their behaviour is more stable than the series $\{y_t\}$ in Fig. 1. Fig. 3 also provides the CUSUMQ statistics $\sum_{\tau=1}^t \hat{u}(\tau)^2 / \sum_{\tau=1}^N \hat{u}(\tau)^2$ together their 1% critical values. This last confirms, in terms of a powerful test, the pattern of variability of the parameters in Fig. 2.

Fig. 4 now displays the on-line F -statistics (17a) together with their asymptotic 1% critical value. The nonsignificance of the causality index $\hat{\Delta}(t)$ at the beginning of the sample is partly due to the choice of initial values $\hat{\sigma}_a^2(0)$ and

Table 2
Parameter estimates (with T -ratios) and statistics of causality

Model	ω_0	δ_1	θ_1	ϕ_{12}	RSS	Statistic
TF	0.809 (7.8)	0.424 (4.3)	0.279 (5.1)	0.364 (5.9)	60.4	$g = 1.4$ (5.1)
ARMA	—	—	0.345 (6.2)	0.399 (6.6)	73.3	$F = 35.1$ ($F_{1\%} = 4.7$)

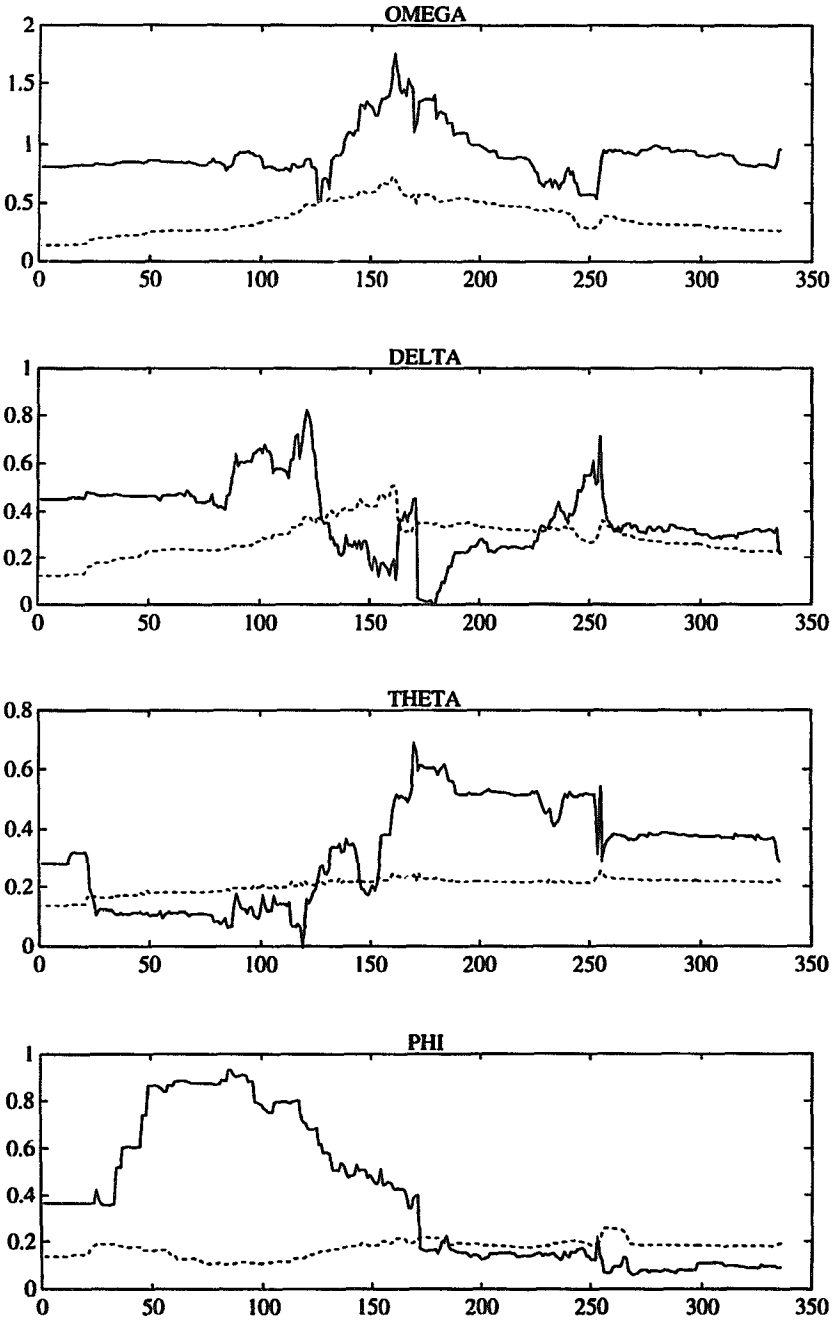


Fig. 2. Recursive estimates (—) and 5% critical values (---) of parameters ω_0 , δ_1 , θ_1 , ϕ_{12} .

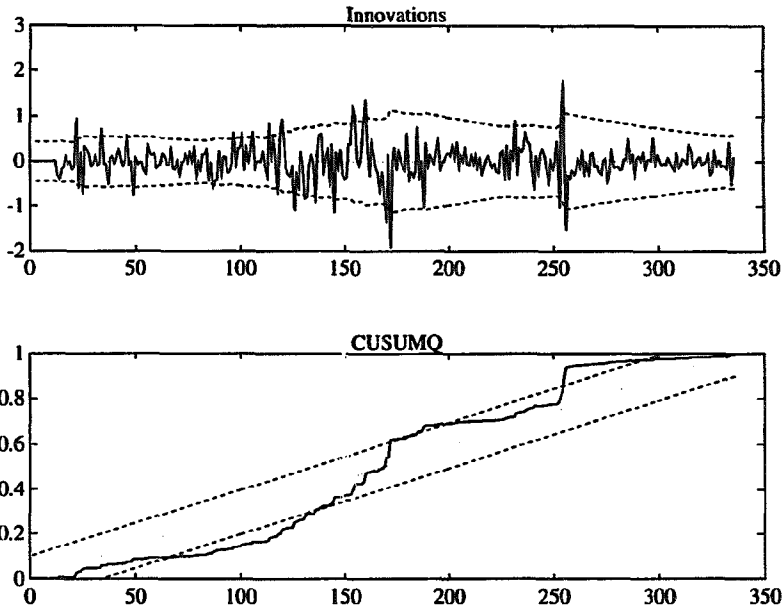


Fig. 3. Graphs of standardized prediction errors and CUSUMQ statistics (—) with 5% and 1% critical values (---), respectively.

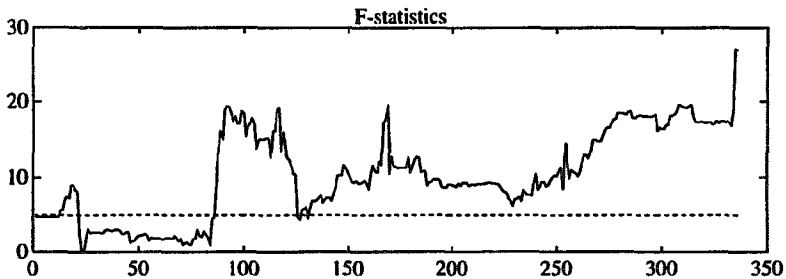


Fig. 4. Recursive F -statistics (17a) (—) and asymptotic 1% critical value (---).

$\hat{\sigma}_e^2(0)$. These were estimated as $\hat{\sigma}(0)^2 = 30^{-1} \sum_1^{30} \hat{u}(t)^2$, and turned out approximately equal.

Finally, Fig. 5 reports the recursive estimates of the gain together with their 5% critical values, derived from (18b). The hypothesis to test is $H_0: g_t = 0$ for $1 \leq t \leq 336$, and as anticipated by Fig. 2, the estimates $\hat{g}(t)$ are locally non-significant. This confirms the asymmetry of the causality parameters Δ, g ; in

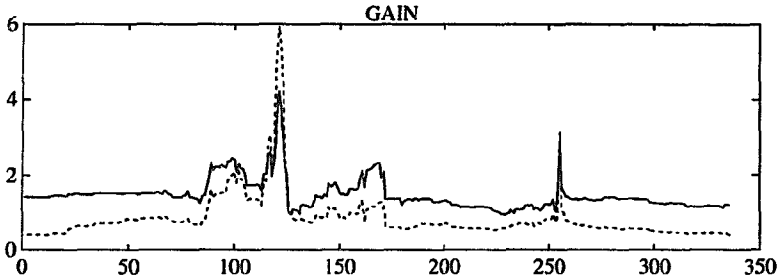


Fig. 5. Recursive estimates of the gain (—) and 5% critical values (---).

practice, to a very significant prediction effect there does not correspond an equivalent multiplicative impact.

A fundamental aspect in the above exercise is represented by the computation of suitable standard errors and degrees of freedom. In particular, the factor $1/(1 + \lambda)$ in the dispersion (21) has a crucial importance in tuning critical values. Without it, all recursive estimates would be nonsignificant, in contrast with the indications of the off-line analysis. In general, mean values of the statistics in Figs. 2–5 agree with the off-line estimates in Table 2, and this confirms the validity of our on-line inferential framework. This framework might also be used for testing the constancy over time of the regression coefficients by comparing the confidence intervals $\hat{\beta}_i(t) \pm T_{\alpha/2}(f) \sqrt{\hat{v}_{ii}(t)}$ at any pair of time instants. With respect to traditional tests of stability, the advantage of this approach is that it directly monitors the paths of the parameter estimates.

Critical aspects in the proposed methodology are represented by the finite-sample properties of estimates and power of the test of significance. Long simulation experiments are recommended to investigate these aspects, even though an intrinsic difficulty is the nonparametric nature of the EW-RLS algorithm and related statistics. We conclude the article with a small Monte Carlo experiment that aims to point out the validity of the recursive approach with respect to conventional methods of analysis.

An ARX(1,1) model with sinusoidal parameter functions was considered,

$$y_t = \phi(t)y_{t-1} + \omega(t)x_t + a_t, \quad x_t \sim \text{IN}(0, 2^2), \quad a_t \sim \text{IN}(0, 1), \quad (23a)$$

$$\phi(t) = 0.93 \sin(0.093t), \quad \omega(t) = 1.56 \sin(0.156t), \quad (23b)$$

$$t = 1, 2, \dots, 200,$$

and 30 replications were fitted with the algorithm (8). Adopting the initial condition $(\hat{\phi}_0 = \hat{\omega}_0) = 0$, suitable tracking coefficients were found to be $R_0 = 10 \cdot I_2$ and $\lambda = 0.7$. Fig. 6 shows the mean values of the recursive estimates (e.g., $\bar{\phi}_t = 30^{-1} \sum_{i=1}^{30} \hat{\phi}_{it}$), together with the mean values of their 5% significance bands. It may be noted that EW-RLS has a good tracking capability and

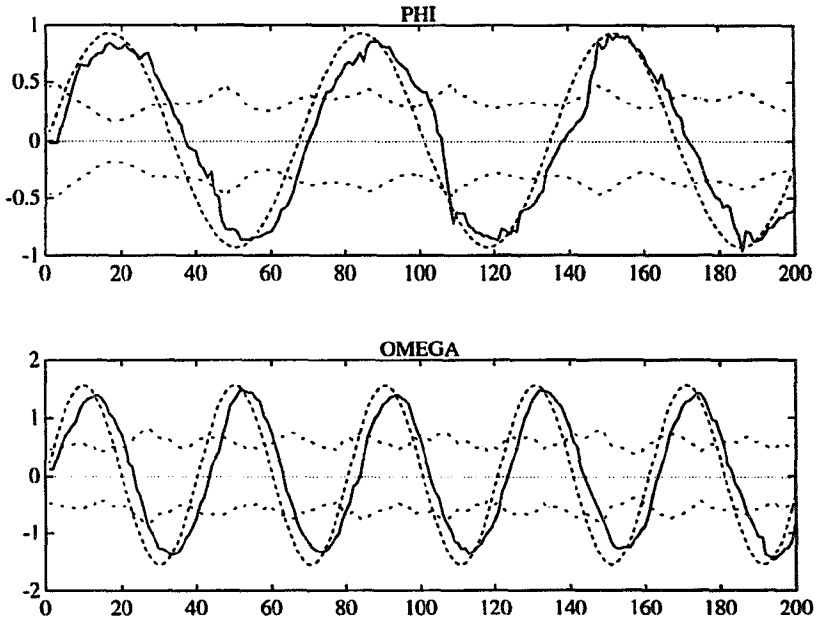


Fig. 6. Parameter functions (23b) (---), mean values of recursive estimates (—), and mean values of 5% critical values (· · ·).

a significant causal relationship $x_t \Rightarrow y_t$ is detected. On the contrary, conventional OLS estimates by providing $\bar{\phi}_N = -0.113$ (0.072), $\bar{\omega}_N = 0.024$ (0.131), mean standard errors are in parentheses, accept the independence of $\{y_t, x_t\}$.

Appendix 1

Proof of Lemma 1

It is easy to show the asymptotic unbiasedness $\lim_{t \rightarrow \infty} E[\hat{\gamma}_t(0)] = \gamma_{xy}(0)$, thus to prove the result we must show that the variance $E[\hat{\gamma}_t(0) - \gamma_{xy}(0)]^2 = O(1 - \lambda)$ when $t \rightarrow \infty$. By the property of Gaussian variates it is known that $E(x^4) = 3 \cdot E(x^2)$, hence for the mean square we have

$$\begin{aligned}
 E[\hat{\gamma}_t^2(0)] &= (1 - \lambda)^2 \sum_{i=1}^t \sum_{j=1}^i \lambda^{2t-i-j} E[x_i y_i x_j y_j] \\
 &= (1 - \lambda)^2 \sum_{i=1}^t \sum_{j=1}^i \lambda^{2t-i-j} [\gamma_{xy}^2(0) + \gamma_{xx}(i-j) \gamma_{yy}(i-j) \\
 &\quad + \gamma_{xy}(i-j) \gamma_{xy}(j-i)] \\
 &= C_t^1 + C_t^2 + C_t^3 \quad (\text{say}).
 \end{aligned}$$

Clearly, $\lim_{t \rightarrow \infty} C_t^1 = \gamma_{xy}^2(0)$ and by assumption $|\gamma_{xx}(i-j)\gamma_{yy}(i-j)| < \alpha \cdot \mu^{|i-j|}$ with α constant. Hence,

$$\begin{aligned} |C_t^2| &\leq \alpha(1-\lambda)^2 \sum_{i=1}^t \sum_{j=1}^t \lambda^{2t-i-j} \mu^{|i-j|} \\ &= \alpha(1-\lambda)^2 \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \lambda^{i+j} \mu^{|i-j|} \\ &\leq \alpha(1-\lambda)^2 \sum_{j=0}^{t-1} \lambda^{2j} + \alpha(1-\lambda)^2 \sum_{i=1}^{t-1} \sum_{j=0}^{i-1} \lambda^{i+j+i-j}. \end{aligned}$$

Now, the limit becomes $\lim_{t \rightarrow \infty} |C_t^2| \leq [\alpha(1-\lambda)/(1+\lambda) + \alpha(1-\lambda)^2 O(1)] = O(1-\lambda)$, and the same holds for C_t^3 (for details see Stoica and Nehorai, 1988). The result then follows by recalling that any real random sequence $\{z_t\}$ is as big as its standard deviation: $z_t = O_p(\sigma_t)$.

Appendix 2

Analysis of the series in level

Following Lütkepohl (1991, p. 378), the analysis of the Granger causality between integrated and cointegrated processes may be developed as in the stationary case. This is mainly due to the superconsistency property of the LS estimator when is applied to vector ARMA systems with unstable roots. Hence, fitting bivariate AR(p) models to the series $\{Y_t, X_t\}$, and selecting the order by minimizing the consistent criterion $BIC_N(p) = \log |\hat{\Sigma}_N| + p 4 \log(N-p)/(N-p)$ (where Σ is the residual covariance matrix), we have obtained

$$\begin{aligned} \begin{pmatrix} Y_t \\ X_t \end{pmatrix} &= \begin{pmatrix} 0.023 \\ (0.1) \\ 0.235 \\ (3.2) \end{pmatrix} + \begin{pmatrix} 1.19 & 0.602 \\ (20.1) & (4.3) \\ -0.006 & 1.46 \\ (0.3) & (26.5) \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} -0.231 & -0.571 \\ (4.1) & (4.2) \\ 0.023 & -0.504 \\ (0.9) & (9.3) \end{pmatrix} \begin{pmatrix} Y_{t-2} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} \hat{a}_t \\ \hat{e}_t \end{pmatrix}, \\ \hat{\Sigma}_N &= \begin{pmatrix} 0.236 & 0.037 \\ 0.037 & 0.035 \end{pmatrix}, \end{aligned}$$

where T -statistics are in parentheses. According to the characterization of the Granger causality in vector AR models (see Granger and Newbold, 1986), the above result establishes with sufficient statistical evidence the one-sided relationship $X_t \Rightarrow Y_t$.

Grillenzoni (1991a) has criticized the vector ARMA framework in modeling multiple time series, because it does not enable a precise identification of the dynamics of the individual processes. An alternative approach, that agrees with the econometric tradition, is provided by a system of simultaneous TF_s or ARMAX models. Following a disaggregate identification strategy we have checked that the series Y_t has a more complex representation:

$$\begin{aligned} & \begin{pmatrix} 1 - 1.364B + 0.415B^2 \\ (24.5) & (7.5) \end{pmatrix} \begin{pmatrix} 1 - 0.418B^{12} \\ (8.0) \end{pmatrix} Y_t \\ & = \begin{pmatrix} 0.492B - 0.878B^2 + 0.969B^3 - 0.559B^4 \\ (3.4) & (3.5) & (3.9) & (4.1) \end{pmatrix} X_t + \hat{a}_t, \quad \hat{\sigma}_a^2 = 0.197, \end{aligned}$$

whereas X_t was confirmed to be an exogenous AR(2) process. These results provide further evidence of the Keynes' theory of interest rates. The problem of the instantaneous causality between Y_t , X_t , raised by the correlation of residuals $\text{corr}(\hat{a}_t, \hat{e}_t) = 0.41$, may be solved by assigning it to the (most) significant of the one-sided relationships, i.e., to $X_t \Rightarrow Y_t$.

In order to achieve a more parsimonious representation for Y_t , a differencing of both series may be useful. In presence of cointegration, however, this transformation distorts certain features of the relationship between the variables and may raise noninvertibility in the subsequent ARMA representation. Tests for cointegration require a preliminary analysis of the degree of integration of the series. Now, performing some augmented Dickey–Fuller tests we have rejected at 99% the presence of a unit root in Y_t , whereas the same hypothesis was accepted at 99% for X_t . Specifically the test equation for Y_t was

$$\begin{aligned} Y_t = & 0.290 + 0.952Y_{t-1} + 0.359y_{t-1} + 0.169y_{t-4} + 0.361y_{t-12} + u_t, \\ & (4.4) \quad (95.1) \quad (7.4) \quad (3.4) \quad (7.3) \end{aligned}$$

where $y_t = (Y_t - Y_{t-1})$, and the test statistic $T_{DF} = (0.952 - 1)/0.01 = -4.8$ is lower than the 1% critical value -3.45 . Since Y_t and X_t have not the same degree of integration, they are not cointegrated (by definition); differencing is then an admissible transformation.

References

- Benveniste, A., 1987, Design of one-step and multistep adaptive algorithms for the tracking of time-varying systems, Ch. 2 in: V.H. Poor, ed., *Advances in statistical signal processing* (JAI Press, London).

- Box, G.E.P. and G.M. Jenkins, 1976, *Time series analysis: Forecasting and control* (Holden-Day, San Francisco, CA).
- Brown, R.L., J. Durbin, and J. M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society B* 37, 149–192.
- Dufour, J.M., 1982, Recursive stability analysis of linear regressions, *Journal of Econometrics* 19, 31–76.
- Goodwin, G.C. and K.S. Sin, 1984, *Adaptive filtering, prediction and control* (Prentice Hall, Englewood Cliffs, NJ).
- Granger, C.W.J. and P. Newbold, 1986, *Forecasting economic time series* (Academic Press, New York, NY).
- Grillenzoni, C., 1991a, Simultaneous transfer functions versus vector ARMA models, *Journal of Forecasting* 10, 477–499.
- Grillenzoni, C., 1991b, Iterative and recursive estimation of transfer functions, *Journal of Time Series Analysis* 12, 105–127.
- Grillenzoni, C., 1994, Optimal recursive estimation of dynamic models, *Journal of the American Statistical Association* 89, 777–787.
- Gunnarsson, S. and L. Ljung, 1989, Frequency domain tracking characteristics of adaptive algorithms, *IEEE Transactions on Signal Processing ASSP-37*, 1072–1089.
- Härdle, W., 1990, *Applied non-parametric regression* (Cambridge University Press, New York, NY).
- Johnson, N.L. and S. Kotz, 1970, *Continuous univariate distributions*, Vol. 2 (Houghton Mifflin, Boston, MA).
- Ljung, L. and S. Gunnarsson, 1990, Adaptation and tracking in system identification – A survey, *Automatica* 26, 7–21.
- Ljung, L. and T. Söderström, 1983, *Theory and practice of recursive identification* (MIT Press, Cambridge, MA).
- Lütkepohl, H., 1991, *Introduction to multiple time series analysis* (Springer Verlag, Berlin).
- Mathai, A.M., 1983, On linear combinations of independent exponential variables, *Communication in Statistics – Theory & Methods* 12, 625–632.
- Niedzwiecki, M., 1988, On tracking characteristics of weighted least squares estimators applied to nonstationary system identification. *IEEE Transactions on Automatic Control AC-33*, 96–98.
- Patnaik, P.B., 1949, The non-central χ^2 and F distributions and their applications, *Biometrika* 36, 202–232.
- Pierce, D.A., 1972, Least squares estimation of dynamic time series models, *Biometrika* 65, 297–303.
- Poskitt, D.S., 1989, A method for the estimation and identification of transfer function models, *Journal of the Royal Statistical Society B* 51, 29–46.
- Stoica, P. and A. Nehorai, 1988, On the asymptotic distribution of exponentially weighted prediction error estimators, *IEEE Transactions on Signal Processing ASSP-36*, 136–139.
- Solomon, H. and M.A. Stephens, 1977, Distribution of a sum of weighted chi-square variables, *Journal of the American Statistical Association* 72, 881–885.
- Zellner, A., C. Hong, and C. Min, 1990, Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques, *Journal of Econometrics* 49, 275–304.