EUI
WORKING
PAPERS IN
ECONOMICS

EUI Working Paper **ECO** No. 93/1

Forecasting Unstable and Non-Stationary Time Series

CARLO GRILLENZONI

European University Institute, Florence

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

Forecasting Unstable and Non-Stationary Time Series

CARLO GRILLENZONI

BADIA FIESOLANA, SAN DOMENICO (FI)

FORECASTING UNSTABLE AND NON-STATIONARY TIME SERIES

Carlo Grillenzoni

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

Abstract - Many of the time series encountered in social and economic applications are asymptotically unstable and intrinsically non-stationary, i.e. satisfy difference equations with roots greater than one (in modulus) and with time-varying coefficients. Time series modeling developed by Box & Jenkins (1976) solves these problems by imposing on data two groups of stationarity transformations : differencing and exponential (Box-Cox). Owing to a generalization of the Jensen inequality, these transformations are not optimal from the forecasting viewpoint and, sometimes, they may be entirely arbitrary. In this paper it is shown that there are no practical and methodological obstacles in modeling time series with unstable roots and changing coefficients. Paradoxically, instability has useful consequences for conventional least squares estimators since it improves their speed of convergence in probability. This property, named super-consistency, was thoroughly analyzed by statisticians in the '50 and in this paper it is investigated by means of recursive estimators applied to simulated data. Next, the effectiveness of adaptive recursive estimators in tracking time-varying unstable roots is shown in the context of an application to the airline data-set of Box-Jenkins. This framework is proper for forecasting time series with trends, cycles and seasonalities whose patterns change over time. Since it does not assume a-priori dynamics for such components, it may be used as a non-parametric alternative to the structured models of Harrison-Harvey.

Key Words - Unstable roots, time-varying coefficients, super-consistency, recursive estimators, tracking coefficients, adaptive forecasting, airline data-set.

1. Introduction

Unlike many of the time series encountered in natural and experimental sciences, social and economic time series frequently exhibit characteristics that are intrinsically *evolutive*. For example, real macroeconomic variables, such as output and consumption, behave as if they had no fixed mean, displaying growth and cyclical components. On the other hand, many financial time series, such as stock prices and and exchange rates, show "bursts" of variability together with local jumps in the levels. These examples qualify in two different ways the "evolutive" characteristic mentioned before :

1) the first class of data belongs to *unstable* or *long memory* time series, that is series in which the mean and the variance functions $\mu_t = E(Z_t)$, $\sigma_t = E(Z_t - \mu_t)^2$ increase gradually over time. These series can still be represented by stochastic difference equations with constant coefficients $Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = \varepsilon_t$, $\varepsilon_t \sim$ IID but in which some roots of the associated characteristic equation $z^p - \phi_1 z^{p-1} - \dots - \phi_p = 0$ are greater than unity in modulus : $|z_j| > 1$ for $j = 1, 2 \dots d$ with $d \leq p$. In this case, statistical analysis of the parameter estimators is complicated by the fact that the implied stochastic process $\{Z_t\}$ does not satisfy standard assumptions of asymptotic independence (or mixing) and moments existence (see White, 1958 and Anderson, 1959).

2) The second group of data, well exemplified by financial series, belongs to *intrinsic non-stationary* time series, i.e. data that can be viewed as realizations of heterogeneously distributed processes. Unlike the first group, these processes may have suitable mixing properties and finite second order moments; nevertheless, they cannot be well represented by linear models with constant parameters. Off-line inference over conventional models is supported by various extensions of the law of large numbers and the central limit theorem (see White and Domowitz, 1984); but it provides results of limited practical utility. On the other hand, on-line inference requires the use of recursive adaptive estimators which involve highly non-linear estimation problems (see Tjostheim 1986a,b).

It should be noted that unstable time series are also non-stationary, but in a weak sense, that is of a form that may be removed by suitable transformations. Rather than having time-varying moments, they do not posses moments at all, in the sense that these moments tend to be infinite.

In past decades the typical solution to problems of instability and non-stationarity was an attempt to reciprocally adapt available data and conventional models. Thus, data have been differenced ($Z_t - Z_{t-1}$), adjusted by outliers and subject to the Box-Cox transformation ($Z_t^\lambda - 1)/\lambda$ (that includes the log-transformation for $\lambda = 0$), in order to achieve linearity and stabilizing the variance. The ARIMA modeling designed by Box & Jenkins (1976) is the most important of these techniques; it has encountered great success by creating a good compromise between simplicity (or parsimony) of the models and their efficacy. From the model structure viewpoint the key element of this adaptation was the inclusion, in the difference equations, of linear factors with roots known a-priori as unity. Long practical experience with ARIMA models has raised two questions that are partially addressed by the recent research on time series developed in econometrics and statistics :

1

i) A-priori assumption of unit roots may be arbitrary since they form a set of measure zero relative to the entire parameter space of autoregressive models.

ii) Parametric transformations may not achieve stationarity since they are too general and implicitly assume that heterogeneity follows a well defined pattern.

Both criticisms reflect a latent demand for direct modeling of the original data; either in order to retain the interpretability of the models with respect to the underlying processes or to improve their forecasting capability. The question now is what kind of alternative models are possible without involving a parametric complexity excessively greater than that of standard ARIMA tools.

In this paper we provide a general framework for modeling unstable and non-stationary time series realizations of $ARIMA_t$ processes with *time-varying unstable roots*, i.e. roots which wander near the unit circle. This kind of representation is suitable for time series containing trends and periodicities with changing slopes and local jumps. For such series Perron (1989) has shown that the unit-root hypothesis (and in general that of ARIMA representation) must be rejected in favour of an alternative in which the trend is generated by a deterministic segmented line. This conclusion is clearly unsatisfactory from the forecasting standpoint and one might argue that the proper null hypothesis in this case, is provided by the unobserved components representation :

$$Z_t = \mu_t + S_t + z_t , \qquad z_t \sim \text{ARMA} \tag{1.1}$$

$$\mu_t = \mu_{t-1} + \delta_t + v_t , \qquad v_t \sim \text{IN}(0, \sigma_v^2)$$

$$\delta_t = \delta_{t-1} + w_t , \qquad w_t \sim \text{IN}(0, \sigma_w^2)$$

$$S_t = S_{t-s} + u_t , \qquad u_t \sim \text{IN}(0, \sigma_u^2)$$

in which the stochastic trend μ_t, its time-varying slope δ_t and the seasonal component S_t conform to the unit-root assumption. Models like the above were outlined by Harrison & Stephens (1976) as dynamic linear models (DLM) and developed by Harvey & Peters (1990) and Young *et al.*(1991) as structural time series models. Even though they concretely solve some problems of the ARIMA approach (retaining parsimony and introducing interpretability) their fundamental limits are represented by additive structure and their random-walk dynamics. More specifically, as we shall see in the application to the airline data-set, trend and seasonality may not be independent and their dynamics may be more complex, i.e. less smooth, than that implied by unit roots.

In our framework, based on time-varying unit-roots of period $s \geq 1$, it will be shown that components of trend and seasonality may simultaneously be captured by models of the type $Z_t = \Phi(t) Z_{t-s} + z_t$, where $\Phi(t)$ is a function of past events $\{Z_{t-1}, Z_{t-2}, \ldots\}$ and $\{z_t\}$ is a stable and nearly stationary process. In off-line inference, the mean value $\overline{\Phi} = \overline{E}[\Phi(t)] = \lim_{N \to \infty} N^{-1} \sum_{t=1}^{N} E[\Phi(t)]$ may consistently be estimated, even if its modulus is greater than one. In on-line estimation the path of $\Phi(t)$ may be tracked, with a satisfactory fit, by standard adaptive algorithms, such as recursive weighted least squares (RWLS). The various recursive methods can be unified into a general algorithm and their tracking

2

coefficients may be determined in an optimal way by minimizing a quadratic objective function based on one-step-ahead prediction errors. In the numerical example it will be shown that instability in a time series is of no substantial obstacle to the application of adaptive estimation methods.

2 . Unstable Time Series

In this section we consider basic issues of modeling unstable time series with constant coefficients; analyses will be developed on the dynamic representations, the forecasting algorithms and statistical properties of the estimators. Simulations on artificial data and application to the airline data-set of Box & Jenkins (1976, Chap.9) will be provided in order to clarify the methodological treatment.

2.1 Models - By definition, *long-memory* and *unstable* time series are realizations of stochastic processes which fail to satisfy standard requirements of asymptotic independence and existence of moments. More specifically, if $\{Z_t ; t = 0,\pm1,\pm2,\dots\}$ is a discrete time process and $\{f(Z_t)\}$ the associated density functions, one should have

$$\lim_{k \to \infty} f(Z_t|Z_{t-k}) \neq f(Z_t), \qquad \sup_t E[|Z_t - c|^r] = \infty$$

for every $0 < c, r < \infty$. More sophisticated definitions refer to joint density functions and probabilistic requirements of moments existence ; namely, for any $(t_1,\dots,t_n), n < \infty$

$$\left[f\left(Z_{t_1},\dots,Z_{t_n}|Z_{t_1-k},\dots,Z_{t_n-k}\right) - f\left(Z_{t_1},\dots,Z_{t_n}\right)\right] \neq O(1/k^\varepsilon), \quad \varepsilon > 0$$

$$\sup_t P(|Z_t| > z) = \sup_t \left[1 - \int_{-z}^{+z} f(Z_t)\, dZ_t \right] \neq o(1/z^{r+\varepsilon}) \quad \text{as} \quad |z| \to \infty$$

The latter always occurs in Cauchy processes, where $f(Z_t;\eta,\gamma) = \{\pi\gamma[1 + ((Z_t - \eta)/\gamma)^2]\}^{-1}$ and η, γ are parameters of location and shape, respectively.

The above properties are absolutely general and must be characterized in relation to the specific parametric representations of the processes. Given a short-memory time series $\{z_t\}$ a dynamic model is a mathematical device $\phi(\cdot)$ that transforms it into a white noise $\{a_t\}$. It is possible to extend this definition to the relationships between long-memory and short-memory time series (see Parzen , 1982), obtaining a sequence of filters

$$\{Z_t\} \quad \rightarrow \quad \begin{bmatrix} \text{Stabilizing} \\ \text{Filter } \Phi(\cdot) \end{bmatrix} \quad \rightarrow \quad \{z_t\} \quad \rightarrow \quad \begin{bmatrix} \text{Whitening} \\ \text{Filter } \phi(\cdot) \end{bmatrix} \quad \rightarrow \quad \{a_t\}$$

Referring to linear representations and allowing for the existence of periodicities of size $s > 1$ in the dynamics of the processes, we may specifically define

$$Z_t - \Phi_1 Z_{t-s} - \dots - \Phi_d Z_{t-ds} = z_t , \qquad \Phi_d(B^s)Z_t = z_t \qquad (2.1a)$$

$$z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = a_t , \qquad \phi_p(B)z_t = a_t \qquad (2.1b)$$

where (B) is the back-shift operator : $B^s Z_t = Z_{t-s}$. Thus, the previous references in terms

3

of memory-length may immediately be qualified as stability properties of linear difference equations, i.e. size of the zeros P_i^{-1}, ρ_j^{-1} of the filters $\Phi(B), \phi(B)$:

$$\Phi(B^s) = \prod_{i=1}^{d}(1 - P_i B^s), \qquad |P_i| \geq 1, \quad i = 1, 2 \dots d \qquad (2.2a)$$

$$\phi(B) = \prod_{j=1}^{p}(1 - \rho_j B), \qquad |\rho_j| < 1, \quad j = 1, 2 \dots p \qquad (2.2b)$$

Another definition of long-memory time series was introduced by Parzen (1982) in a somewhat heuristic way. Such series should satisfy the fitting condition

$$\tilde{\sigma}_N^2 = \left[\left(\sum_{t=d+1}^{N} z_t^2 \right) / \left(\sum_{t=d+1}^{N} Z_t^2 \right) \right] < \frac{8}{N}$$

where N is the sample size. Note that in terms of the coefficient of linear determination the condition is equivalent to $R_N^2 \geq .84, .92$ for medium sample sizes $N = 50, 100$. Even though the associated F-statistics are 1% significant for $p = 1, 2$, values $R_{50}^2 \geq .85, R_{100}^2 \geq .93$ are generated in 86%, 80% of realizations by the widely stable AR(1) model $Z_t = .95 Z_{t-1} + z_t$ with $z_t = .45 z_{t-1} + a_t$ and $a_t \sim \text{IN}(0, 1)$.

Classification of the memory-length of a time series based on the size of the roots in (2.2) is less ambiguos. For example, letting $d = 1, p = 0$ and solving the resulting model $Z_t = \Phi Z_{t-s} + a_t$ for $t = (\tau s) > 0$ with initial condition $Z_0 \neq 0$, we find

$$Z_\tau = \Phi^\tau Z_0 + \Phi^{\tau-1} a_s + \Phi^{\tau-2} a_{2s} + \dots + \Phi a_{(\tau-1)s} + a_{\tau s} \qquad (2.3)$$

With $|\Phi| > 1$ one may easily check lack of asymptotic independence : $\lim_{k \to \infty} E(Z_{t+k} Z_t) \neq 0$ and lack of asymptotic boundedness : $\lim_{t \to \infty} P(|Z_t| \leq M_\epsilon) < \epsilon$ for any $0 < \epsilon, M_\epsilon < \infty$, from which follows lack of moment existence. Some ambiguity remains for models with roots $|P_j| \equiv 1$ on the unit circle ; in this case long-memory occurs but first order moments exist, e.g. $\sup_t E(Z_t) = Z_0$. This is not the only peculiarity of unit-root processes ; other special features will be met in statistical inference. However, from a stochastic coefficient viewpoint the probability of having $d > 0$ roots exactly on the unit circle is zero, hence processes of practical interest may simply be classified as stable or unstable.

It might be argued that in case of unstable models, decompositions like (2.3) do not make sense since they do not converge in probability. However, mathematical devices exist for building convergent expansions which may be used to define the parametric expressions of covariance functions and spectral densities. The basic idea is that of "inverting" unstable roots ; for simplicity, let be $|P_j| > 1$ for all j and write (2.1)-(2.2) as

$$z_t = \sum_{i=0}^{d} \Phi_i F^{(d-i)s} Z_{t-ds} = \prod_{i=1}^{d} (F^s - P_i) Z_{t-ds}$$

where $F = B^{-1}$ is the forward operator : $F^s Z_t = Z_{t+s}$. Hence the inverse of the above becomes

$$Z_{t-ds} = \prod_{i=1}^{d} (F^s - P_i)^{-1} z_t = \prod_{i=1}^{d} \left[-P_i^{-1}(1 - P_i^{-1} F^s)^{-1} \right] \prod_{j=1}^{p} (1 - \rho_j B)^{-1} a_t$$

4

Here, each factor $(1 - P_i^{-1} F')^{-1}$ can be expanded in a convergent power series of F, and each term $(1 - \rho_j B)$ can be expanded in a power series of B. The global result is a two-sided moving average decomposition $Z_t = \sum_{k=-\infty}^{+\infty} \psi_k a_{t-k}$ which converges in probability for all $t \in I$. The formal consequence of this representation is that current values of unstable processes Z_t are correlated with future innovations a_{t+k}; this seems to deny the possibility of making good forecasts from past data. In reality, the above result only means that unstable processes *formally admit* two-sided decompositions which, however, may have no substantial meaning in a context of physically realizable processes (where $t \geq 0$).

Simulations - Plots of realizations of simulated processes are useful for better understanding probabilistic properties of unstable models as well as to introduce heuristic elements of pattern identification. Figure 1 provide realizations of : (*a*) a random walk $Z_t = Z_{t-1} + a_t$; (*b*) a weakly unstable process $Z_t = 1.001 Z_{t-1} + a_t$ and a random walk plus drift $Z_t = .1 + Z_{t-1} + a_t$; (*c*) a "doubly" unstable process $(1 - 1.002 B)(1 - 1.001 B) Z_t = a_t$; (*d*) a periodic unstable process $(1 - 1.001 B)(1 - 1.001 B^{12}) Z_t = a_t$. In all simulations the disturbance sequence was independent gaussian $a_t \sim IN(0, 1)$. Important remarks that can be drawn from Figure 1 are : (1) random walks tend to be stationary in mean and only "explode" in presence of a deterministic drift ; (2) trends generated by random walks and unstable models are qualitative different : the first is linear and the other is exponential ; (3) series with two or more unstable roots tend to be smooth like deterministic functions. The reason is that in the decomposition $Z_t = \hat{Z}_t(1) + a_t$ the deterministic component

Figure 1 (a,b,c,d) - *Realizations of unstable processes.*
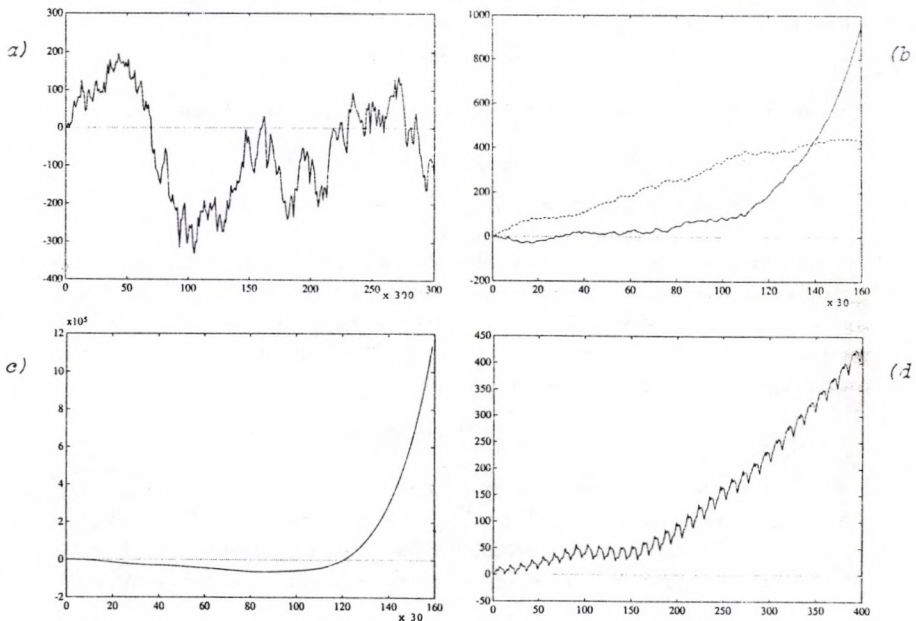
$\hat{Z}_t(1) = E(Z_t \mid Z_{t-1}, Z_{t-2} \ldots)$ rapidly diverges while a_t remains bounded in probability.

Combining equations (2.2) we obtain the *multiplicative* model $\phi_p(B)\Phi_d(B^s)Z_t = a_t$ with $a_t \sim \text{IID}$. Apart from the utility in terms of classification —since it separates ordinary and periodic, stable and unstable linear factors— the advantage of multiplicative filters over the additive ones is essentially a question of parsimony. This is apparent in the presence of periodicities since the two parameter model $(1 - \phi B)(1 - \Phi B^s)Z_t = a_t$ generates an AR(3) process. Special features are also involved in second order moments since the MA model $z_t = (1 - \theta B)(1 - \Theta B^s)a_t$ has 4 autocovariance coefficients different from zero, namely $\gamma(1), \gamma(s)$ and $\gamma(s-1) = \gamma(s+1)$, instead of 2 as in the additive case. Hence, multiplicative models have a richer internal dependence structure which is suitable for representing and forecasting complex stochastic processes. The price to be paid concerns the procedures of identification and estimation ; for example, even in absence of MA factors, estimation requires iterative algorithms. Standard differentiation applied to model (2.1) provides the gradient for the nonlinear least squares (NLS) estimator

$$-\frac{\partial a_t}{\partial \Phi_i} = \Phi(B^s)^{-1} B^{is} a_t \, , \qquad -\frac{\partial a_t}{\partial \phi_j} = \phi(B)^{-1} B^j a_t$$

but only in the second case this will coincide with the "observable" quantity z_{t-j} involved in the sequential OLS estimator.

Sequential strategies may be used, without loss of optimality, in the forecasting phase. Indeed, the forecast function at origin t for lead time h must satisfy the difference equation

$$\phi_p(B)\Phi_d(B^s)\hat{Z}_t(h) = 0 \, , \qquad \hat{Z}_t(h) = E[Z_{t+h} \mid Z_t, Z_{t-1}, \ldots]$$

and if $\{z_t\}$ has an ARMA representation, the sequential solution becomes

$$\hat{z}_t(h) = \sum_{j=1}^{h-1} \phi_j \hat{z}_t(h-i) + \sum_{j=h}^{p} \phi_j z_{t-j} + \sum_{j=h}^{q} \theta_j a_{t-j}$$

$$\hat{Z}_t(h) = \sum_{i=1}^{h-1} \Phi_i \hat{Z}_t(h-is) + \sum_{i=h}^{d} \Phi_i Z_{t-is} + \hat{z}_t(h)$$

As introduced in the abstract, data transformations $Y_t = g(Z_t)$ may cause serious problems for the performance of predictors. In particular, the optimal predictor (in mean square sense) of original series does not coincide with the inverse transformation of the predictor of the transformed series : $\hat{Z}_t(h) \neq g^{-1}[\hat{Y}_t(h)]$, $h > 0$. This result follows immediately from the extension to conditional expectation of the well known Jensen inequality :

$$E[g(Z_{t+h}) \mid Z_t, Z_{t-1}, \ldots] \neq g[E(Z_{t+h} \mid Z_t, Z_{t-1}, \ldots)]$$

where the sign $>, <$ of the inequality depends on the convexity or concavity of $g(\cdot)$. This conclusion applies to every kind of transformation : differencing as well as Box-Cox in which $\hat{Z}_t(h) \neq [\lambda \hat{Y}(h) + 1]^{1/\lambda}$. More specifically, the a-priori assumption of unit-roots for unstable processes and the log-transformation make the predictors biased downward.

2.2 Application - We now illustrate the above concepts with an application to the airline data-set of Box & Jenkins (1976, p.305). The series, displayed in Figure 2, regards monthly observations of passenger totals in U.S. international air travel from 1949 to 1960; the sample size is $N = 144$ and the period lag is $s = 12$. The model fitted by Box-Jenkins relies heavily on the multiplicative structure and a-priori transformations

$$(1-B)(1-B^{12})\log(Z_t) = \left(1 - \underset{(23.)}{.377}\, B\right)\left(1 - \underset{(42.)}{.587}\, B^{12}\right)\hat{a}_t\, , \quad Q_N = \Sigma_{t=14}^{144}\hat{a}_t^2 = .182 \qquad (2.4)$$

where the statistics in parentheses are t-ratios and Q_N is the sum of squared prediction errors. In this framework the log-transformation not only serves to stabilize the variance of $\{Z_t\}$ and to improve the efficiency of estimates, but is essential for the assumption of unit roots. In fact, the model fitted on original series gave lower parameter values

$$(1-B)(1-B^{12})Z_t = \left(1 - \underset{(56.)}{.310}B\right)\left(1 - \underset{(21.)}{.113}\, B^{12}\right)\hat{a}_t\, , \quad Q_N = 17,752$$

Questions raised in the previous discussion regard : (*i*) consistency of the unit roots assumption with data ; (*ii*) feasibility and efficacy of unstable roots modeling ; (*iii*) relationships between trend and seasonal components ; (*iv*) effect of data transformations on forecasts. The first step to take to answer these questions is the estimation of the model identified by Box-Jenkins without the unit-roots constraint :

$$\left(1 - \underset{(140.)}{.948}\, B\right)\left(1 - \underset{(1540.)}{1.021}\, B^{12}\right)\log(Z_t) = \left(1 - \underset{(20.)}{.361}\, B\right)\left(1 - \underset{(41.)}{.582}\, B^{12}\right)a_t\, , \quad Q_N = .178 \qquad (2.5a)$$

$$\left(1 - \underset{(289.)}{.915}\, B\right)\left(1 - \underset{(1574.)}{1.118}\, B^{12}\right)Z_t = \left(1 - \underset{(59.)}{.399}\, B\right)\left(1 - \underset{(72.)}{.489}\, B^{12}\right)a_t\, , \quad Q_N = 13,982 \qquad (2.5b)$$

Here, values of parameters in the two models seem quite homogeneous ; however, only in the case without log-transformation is the coefficient Φ far from unity and does it substantially improve the fit over previous estimates. The introduction of a deterministic drift does not modify this performance in any way

Figure 2 - *Plot of airline data-set* $[Z_t \underline{\quad} ; \log(Z_t) \text{----}]$.

7
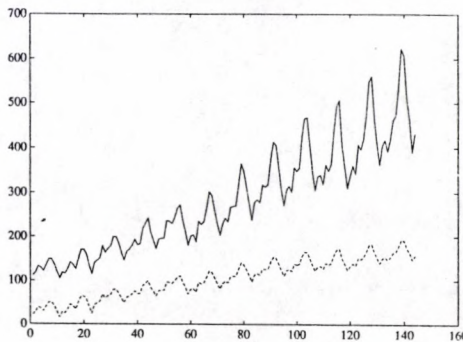
$$\left(1 - \underset{(256.)}{.901}\, B\right)\left(1 - \underset{(1461.)}{1.111}\, B^{12}\right)Z_t = \underset{(8.5)}{.275} + \left(1 - \underset{(53.)}{.382}\, B\right)\left(1 - \underset{(67.)}{.470}\, B^{12}\right)a_t\,, \qquad Q_N = 13,965$$

in fact, as seen in Figure 1b , unstable roots alone may generate trends .

Another estimation experiment regards the sequential structure of multiplicative models. The results confirm the crucial importance of unstable periodic roots

$$\left(1 - \underset{(5200.)}{1.114}\, B^{12}\right)Z_t = z_t\,, \qquad\qquad\qquad Q_z = 35,920 \qquad\qquad (2.6a)$$

$$\left(1 - \underset{(313.)}{.913}\, B\right)z_t = \left(1 - \underset{(64.)}{.393}\, B\right)\left(1 - \underset{(81.)}{.474}\, B^{12}\right)a_t\,, \qquad Q_a = 14,003 \qquad (2.6b)$$

In this case the introduction of a drift in the first equation has significant effects on the size of the roots, but the global fitting performance slightly worsens

$$\left(1 - \underset{(1806.)}{1.073}\, B^{12}\right)Z_t - \underset{(75.)}{12.666} = z_t\,, \qquad\qquad Q_z = 32,940$$

$$\left(1 - \underset{(196.)}{.837}\, B\right)z_t = \left(1 - \underset{(40.)}{.297}\, B\right)\left(1 - \underset{(38.)}{.361}\, B^{12}\right)a_t\,, \qquad Q_a = 14,302$$

At this point the following remarks can be made : (*1*) With or without log-transformation, the presence of roots significantly greater (Φ) and lower (ϕ) than unity is detected in the airline data-set. Without log-transformation the relaxation of the unit-root hypothesis improves the fitting performance by about 20% . (*2*) Periodic unstable models of the type $Z_t = \Phi Z_{t-s} + z_t\,, |\Phi| > 1$ are capable of simultaneously capturing seasonal and trend components. This is concretely shown by Figure 3 which compares the differenced series $(Z_t - Z_{t-12})$ with $(Z_t - 1.114 Z_{t-12})$, where the latter is nearly stationary. The conclusion is that seasonal and trend components are neither independent nor separable. (*3*) The high *t*-statistics in the previous estimations are due to the peculiar asymptotic properties of least squares estimates in presence of unstable roots. As we shall see in next sub-section standard limit theory does not hold and *t*-ratios are biased indicators.

Figure 3 - *Plot of filtered series* $[\,(Z_t - 1.114 Z_{t-12}) \underline{\quad\quad} ; (Z_t - Z_{t-12}) \text{----}\,]$.
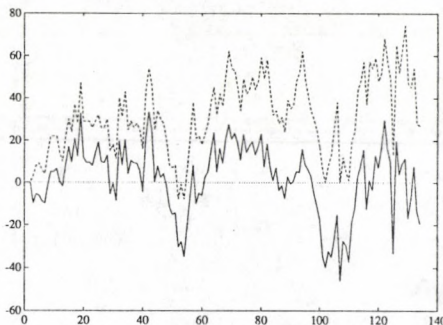
From previous estimations on log-transformed data it is not possible to discriminate precisely between models with unit or unstable roots; moreover, the values of the statistic Q_N cannot be compared with those of the models estimated on original data. Thus, the sole objective way to compare the performances of the various models is that of making forecasts of real data and computing empirical prediction errors. Given the origin t and the steps ahead $1 < h < 12$, the predictor of multiplicative models of the type $(1 - \phi B)(1 - \Phi B^{12})\log(Z_t) = (1 - \theta B)(1 - \Theta B^{12})a_t$, is given by

$$\hat{Z}_t^*(h) = \exp[\phi \log(\hat{Z}_t(h-1)) + \Phi \log(Z_{t+h-12}) - \phi \Phi \log(Z_{t+h-13}) - \Theta a_{t+h-12} + \theta \Theta a_{t+h-13}]$$

One of the hypothesis to be checked is whether this predictor is better than that which assumes $\phi = \Phi = 1$, but worse than $\hat{Z}_t(h)$ which works directly on original data. As a consequence of the conditional Jensen inequality it should be expected that $| \hat{Z}_t^*(h) - Z_{t+h} | > | \hat{Z}_t(h) - Z_{t+h} |$ for each t, h; absolute prediction errors (ApE) thus provide building blocks for model comparison. In order to mitigate the dependence of individual ApE_s from the particular forecast origin, it is necessary to average over t, obtaining the index

$$\text{MApE}_n(h) = \frac{1}{n} \sum_{\tau=1}^{n} | Z_{t+\tau+h} - \hat{Z}_{t+\tau}(h) | \tag{2.7}$$

where n is the sample size of the mean and τ shifts the forecast origin. In the airline application we have made forecasts over the last two years taking $t = 1958.12$, $n = 12$ and $h, \tau = 1, 2, \ldots 12$; more specifically, forecast origin was changed 12 times, and each time 12 forecasts were computed.

A plot of MApE statistics generated by models (2.4)-(2.5) is given in Figure 4; the best model is (2.5b) which does not transform data, whereas the worst one is (2.4) proposed by Box & Jenkins. Intermediate performance is provided by model (2.5a), with unstable roots but log-transformed data. Forecasting performance of unstable model (2.5b) can be further improved by respecifying the ARMA model of the series $\{z_t\}$ generated by (2.6a).

Figure 4 - *Plot of statistics (2.7) for models* $[(2.4) \_\_\_ ; (2.5a) \_._ ; (2.5b) \_\_\_]$.
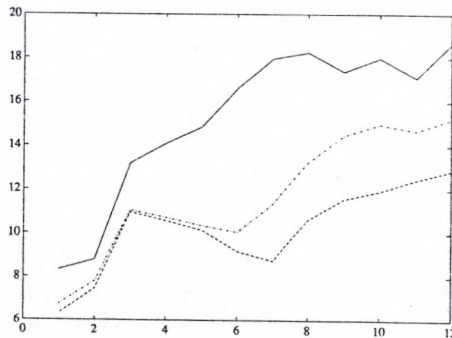
However, the task of the above exercise was the evaluation of statistical effects of arbitrary data transformations, for a *given* model structure. As we may see in Figure 4 these effects are considerable in long run forecasts since prediction errors increase, on average, even by more than 100%. This is the case, for example, of $\text{MApE}_{12}(7)$.

2.3 Inference - A large part of the original statistical literature concerned with the asymptotic properties of parameter estimators of unstable AR-processes, focused on first order models such as

$$Z_t = \Phi Z_{t-s} + a_t, \qquad a_t \sim \text{IID}(0, \sigma^2 < \infty), \quad t = s, s+1 \dots$$

where $1 \leq |\Phi| < \infty$ and $0 \leq |Z_0| < \infty$. If Z_0 is fixed or stochastically defined as $Z_0 = a_0$, the distribution of the process $\{Z_t\}$ is entirely determined by that of the disturbances $\{a_t\}$; in the gaussian case least squares and maximum likelihood methods provide the same estimator

$$\hat{\Phi}_N = \left(\sum_{t=s}^N Z_t Z_{t-s} \right) \left(\sum_{t=s}^N Z_{t-s}^2 \right)^{-1}, \qquad \hat{\sigma}_N^2 = \frac{1}{N-s} \left(\sum_{t=s}^N Z_t - \hat{\Phi}_N Z_{t-s} \right)^2$$

The first problem is to establish its convergence : if $|\Phi| < 1$ the process $\{Z_t\}$ is covariance stationary (i.e. it has second order moments, asymptotically independent of Z_0) and weak consistency follows from the ergodic and related theorems. Specifically, $\{Z_t Z_{t-s}\}$, $\{Z_{t-s}^2\}$ are mean stationary, asymptotically independent processes and their sample means converge in probability to the autocovariances $\gamma(s), \gamma(0)$. The first difficulty that one encounters in extending this analysis to the unstable case $|\Phi| \geq 1$ is that the coefficients $\gamma(k)$ are infinite for each lag k.

For $s = 1$, Rubin (1950) and Rao (1961) showed that the consistency property can still be obtained by redefining the normalizing function of the estimator components. Thus, instead of multiplying numerator and denominator of $\hat{\Phi}_N$ by $1/N$, if one divides by a function that diverges exponentially, such as

$$g(N)^2 = \{ |\Phi|^N / (\Phi^2 - 1) \}^2$$

it is possible to establish that $\hat{\Phi}_N \xrightarrow{P} (\Phi \zeta^2 / \zeta^2)$ where ζ is a well defined random variable. In particular, ζ is absolutely continuous, so that $P(\zeta = 0) = 0$, and has the same moments as $\{a_t\}$, that is $E(\zeta) = 0$ and $E(\zeta^2) = \sigma^2$. The key element of the proof consists in showing that the sequence $\zeta_N = (\Phi^2 - 1)^{1/2} \sum_{t=1}^N \Phi^{-t} a_t$ converges in probability to the random variable ζ; next, using $Z_t = \sum_{i=1}^t \Phi^i a_{t-i}$ one may obtain that $g(N)^{-2} \sum_{t=1}^N Z_{t-1}^2 \xrightarrow{P} \zeta$ and similarly $g(N)^{-2} \sum_{t=1}^N Z_t Z_{t-1} \xrightarrow{P} \Phi \zeta^2$ from $Z_t = \Phi Z_{t-1} + a_t$.

The problem of finding the asymptotic distribution of the bounded random functional $g(N)(\hat{\Phi}_N - \Phi)$ is much more involved and was tackled, in a nearly comprehensive form, by White (1958, 1959). Transparent results were achieved only for $Z_0 = 0$, $\Phi \neq 1$ and a condition of gaussian disturbances $a_t \sim \text{IN}(0, \sigma^2)$; in this case White (1958) proved that $g(N)(\hat{\Phi}_N - \Phi)$ converges in distribution to the standard Cauchy density

10

$$\frac{|\Phi|^N}{(\Phi^2-1)}(\hat{\Phi}_N-\Phi) \xrightarrow{D} C(y;0,1)=\frac{1}{\pi(1+y^2)}$$

As is known, Cauchy random variables have no moments and standardization refers to parameters of location and shape; in this case the density coincides with a t-Student with one degree of freedom.

Another important result obtained by White (1958) was to show that for $|\Phi|=1$ and $g(N)=N/\sqrt{2}$, the asymptotic distribution of $g(N)(\hat{\Phi}_N-\Phi)$ is non-degenerate and invariant to the choice of $Z_0 \neq 0$. In particular, it is related to functionals of a standard Brownian motion $\{W(t):0 \leq t \leq 1\}$

$$N(\hat{\Phi}_N-1) \xrightarrow{D} f(w)=\frac{1}{2}[W^2(1)-1]/\int_0^1 W^2(t)\,dt \qquad (2.8)$$

By contrast, in the general case $|\Phi|>1$ the inversion of the asymptotic moment generating function leads to non-standard probability models which depend on the values of $Z_0 \neq 0$ and $\Phi \neq 1$:

$$f(x;\theta)=\frac{\exp(-\theta)}{\sqrt{\pi}(1+x^2)}\sum_{k=0}^{\infty}\left(\frac{\theta}{1+x^2}\right)^k\frac{1}{\Gamma(k+1/2)}, \qquad \theta=Z_0(\Phi^2-1)\frac{1}{2}$$

Another distinctive feature of result (2.8) is its independence from the distribution of disturbances, the sole requirement is that $\{a_t\}$ be an IID sequence with finite moments. From a practical standpoint these invariance properties are very useful since they enable a unique tabulation of the non-standard density.

The need to recover the limiting-normality result established by Mann-Wald for stable models with $a_t \sim$ IID, led White (1959) to modify the standardizing function. Setting $g(N)=(\sum_{t=1}^{N}Z_{t-1}^2/\hat{\sigma}_N^2)^{1/2}$ he showed that

$$[\sum_{t=1}^{N}Z_{t-1}^2/\hat{\sigma}_N^2]^{1/2}(\hat{\Phi}_N-\Phi) \xrightarrow{D} N(0,1) \qquad (2.9)$$

which holds for every $|\Phi| \neq 1$ and $|Z_0| \geq 0$, but not for $|\Phi|=1$. An important limitation of (2.9) is that it only holds for gaussian disturbances $a_t \sim IN(0,\sigma^2)$; nevertheless, in this context it easily enables one to implement likelihood ratio tests. In fact, given the system of hypotheses $H_0:\Phi=\Phi_0$, with $\Phi_0 \neq 1$, versus $H_1:\Phi \neq \Phi_0$, the LR-statistic becomes

$$\lambda_N=\left[1-\frac{(\hat{\Phi}_N-\Phi_0)^2\sum_{t=1}^{N}Z_{t-1}^2}{\sum_{t=1}^{N}(Z_t-\Phi_0 Z_{t-1})^2}\right]^{N/2}$$

and, as usual, $[-2\log(\lambda_N)] \xrightarrow{D} \chi^2(1)$ as $N \to \infty$.

Original results of Rubin (1950) and White (1958) were further developed and extended by Anderson (1959) and Rao (1961) toward vector AR(1) and general AR(p) processes. Important refinements for the scalar AR(1) case regarded unit roots; e.g. Anderson showed that under the White's assumptions $a_t \sim IN$, $Z_0=0$ the limiting distribution of (2.8) is not symmetric. He also showed that when $|\Phi|>1$ the condition that disturbances be gaus-

11

sian is essential for proving (2.9), whereas it is not necessary in the stable case. At this point it is useful to summarize the various normalizing functions $g(N)$

$$g(N) = \begin{cases} [N/(1-\Phi^2)]^{1/2} & \text{if} \quad 0 < |\Phi| < 1 \\ N/\sqrt{2} & \text{if} \quad |\Phi| = 1 \\ |\Phi|^N/(\Phi^2-1) & \text{if} \quad 1 < |\Phi| < \infty \end{cases}$$

these functions determine the speed of convergence of the least squares estimator $\hat{\Phi}_N$ on the parameter space $-\infty < \Phi < +\infty$. Thus, despite the fact that unstable processes do not have moments they allow conventional estimators to be more efficient in mean square sense. In particular, since convergence rate of $\hat{\Phi}_N$ grows exponentially with modulus $|\Phi|$, we have a form of *super-consistency* in weak sense (i.e. not with probability one). This largely explains the high *t*-statistics of the estimates of unstable parameters encountered in the airline application. As we have seen, however, the price to be paid for this property is that asymptotic distributions of bounded functions of $(\hat{\Phi}_N - \Phi)$ are non-standard and need ad-hoc tabulations.

Simulations - A transparent way to numerically illustrate the super-consistency property is that of using least squares estimators in recursive form (RLS). The on-line implementation processes data one observation at time and provides corresponding parameter estimates; in this way it is possible to have graphical evidence of the speed of convergence. Formulas exist which simplify calculations, in particular they update estimates on the basis of previous values and avoid direct matrix inversion. If $\hat{\Phi}_N$, Z_{t-1} are $p \times 1$ vectors we have

$$\hat{\Phi}_t = \hat{\Phi}_{t-1} + \hat{\Gamma}_t Z_{t-1} [Z_t - \hat{\Phi}_{t-1}' Z_{t-1}] \qquad (2.10a)$$

$$\hat{\Gamma}_t = \hat{\Gamma}_{t-1} - \frac{\hat{\Gamma}_{t-1} Z_{t-1} Z_{t-1}' \hat{\Gamma}_{t-1}}{1 + Z_{t-1}' \hat{\Gamma}_{t-1} Z_{t-1}} \qquad (2.10b)$$

Simulations have been conducted on the model $Z_t = \Phi Z_{t-1} + a_t$ with $a_t \sim IN(0,1)$, $Z_0 \equiv 0$ under various designs of Φ. Recursive estimator (2.10) has always been initialized with

Figure 5 (a,b) - *Convergence of RLS estimates in* AR(1) *models under various designs of* Φ.

12
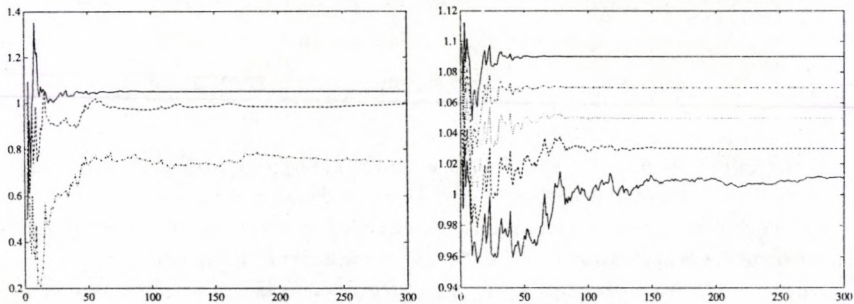
the values $\hat{\Gamma}_0 = 1$, $\hat{\Phi}_0 = .5$. Figure 5a plots series $\{\hat{\Phi}_t\}$ corresponding to $\Phi = .75, 1, 1.05$; super-consistency can, however, be better appreciated in Figure 5b that reports recursive estimates corresponding to $\Phi = 1.01, 1.03, 1.05, 1.07, 1.09$.

By the '50$_s$ much of the asymptotic theory of the first order AR process had been worked out. The limiting distribution had been characterized, in closed form, except, at most, for the unit-root case . Rubin's weak consistency result (which uniformly holds on the parameter space) and the fact that $|\Phi| = 1$ is a set of measure zero, explain the lack of interest in integrated processes. However, the extensive use of differencing in applied time series analysis and the subsequent interpretation of ARIMA processes as persistent shock models have provided new motivations. Recent developments include Rao (1978) who has extended White's intuition (2.9) by applying a theorem of Cramer to obtain an explicit expression for the limiting density of $(\hat{\Phi}_N - 1)N/\sqrt{2}$. Rao's expression is not very useful since it involves complicated special functions ; by contrast, functionals of the standard Brownian motion such as (2.8) may easily be simulated and tabulated.

An original and useful field of research regards the study of near-integrated processes, i.e. models with stable roots that are close to unity : $\Phi \approx 1$. In this case, although the asymptotic normality theory holds, it may provide a very poor approximation in small and moderate samples. To provide a solid statistical framework for nearly unstable processes the parameter Φ must be replaced by $\Phi_N = 1 - \delta/N$, where δ is a small constant. The dynamic representation then becomes $Z_t^N = \Phi_N Z_{t-1}^N + a_t$ with $Z_0^N = 0$ for all N. In this context Chan & Wei (1987) show, under mild regularity conditions, that

$$\left(\sum_{t=1}^{N} Z_{t-1}^2 \right)^{1/2} (\hat{\Phi}_N - \Phi_N) \xrightarrow{D} h(\gamma) = \frac{\int_0^1 (1+\gamma t)^{-1} W(t)\, dW(t)}{\left[\int_0^1 (1+\gamma t)^{-2} W^2(t)\, dt \right]^{1/2}}$$

where $\gamma = \exp(2\delta) - 1$ and $\{W(t) : 0 \le t \le 1\}$ is a standard Brownian motion. The important point is that $h(\delta)$ is a continuous family of distributions which may be simulated in order to obtain more accurate finite sample confidence intervals.

Identification - We conclude the section by investigating the impact of estimators' properties on the procedures for identifying unstable models. Assuming a multiplicative structure of the type (2.1)-(2.2) it seems natural to face the identification problem in a sequential way, i.e. identification of the order d of the unstable autoregressive filter; next, identification of the orders p, q of the ARMA filter of the stabilized series. Since values of unstable roots are not known a-priori, this strategy necessarily requires parametric estimation of linear models in the first step. In this case, given uncertainties about the statistical properties of the LS-estimator in presence of several unstable roots, the application of the sequential approach in the identification of the order $d > 1$ may be problematic. Indeed, such a strategy would require the LS-estimation at step $j \le d$ of the partial model $\hat{Z}_t^{(j-1)} = P_j \hat{Z}_{t-s}^{(j-1)} + Z_t^{(j)}$; but for $j < d$ the residual process $\{Z_t^{(j)}\}$ is still unstable so that statistical properties defined in the previous sub-section may not hold for $\hat{P}_{jN}, \hat{Z}_t^{(j)}$.

An approximate method for the selection of d, that may avoid these drawbacks, consists of assuming $|P_j| \approx 1$ for all $j = 1, \dots d$ and then applying standard techniques for integrated processes (see e.g. Cressie, 1988). However, as seen in the airline application, this solution tends to overestimate d, and so leads to misspecification of the ARMA model of $\{z_t\}$. We now propose an identification procedure that retains sequentiality but respects the statistical theory of unstable processes.

Step 1 - Given the value of periodicity $s > 1$ (known from sample survey information), estimate the models

$$Z_t = \Phi_1 Z_{t-s} + Z_t^{(s)}, \qquad Z_t = \phi_1 Z_{t-1} + Z_t^{(1)}$$

if $|\hat{\Phi}_{1N}| > 1$, $|\hat{\phi}_{1N}| > 1$ and $\{\hat{Z}_t^{(s)}\}$, $\{\hat{Z}_t^{(1)}\}$ are stationary, then set $d = 1$ and pass to Step 3. If both processes $\hat{Z}_t^{(\cdot)}$ are stable, choose the model for which $Q_N = \sum_t [\hat{Z}_t^{(\cdot)}]^2$ is lower.

Step 2 - In the negative case, fit the second order models

$$Z_t = \Phi_1 Z_{t-s} + \Phi_2 Z_{t-2s} + Z_t^{(2s)}, \qquad Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + Z_t^{(2)}$$

$$Z_t = \phi_1 Z_{t-1} + \Phi_1 Z_{t-s} - \phi_1 \Phi_1 Z_{t-1-s} + Z_t^{(1,s)}$$

and compute the characteristic roots $[\hat{P}_1, \hat{P}_2, \hat{\rho}_1, \hat{\rho}_2]_N$. If one of the series $\{\hat{Z}_t^{(2)}, \hat{Z}_t^{(2s)}, \hat{Z}_t^{(1,s)}\}$ is stable and the roots of the corresponding model are both unstable, then set $d = 2$ and pass to Step 3. If one of its roots is stable go back to Step 1. In case of multiple choice, select the model which has the lower value of the statistic Q_N.

Step 3 - By extending Step 2 to higher order models, one may finally obtain a series $\hat{Z}_t^{(k,hs)} = \prod_{i=1}^{k}(1 - \hat{\rho}_{iN}B) \prod_{j=1}^{h}(1 - \hat{P}_{jN}B^s) Z_t$, with $(k + h) = d$ and $|\hat{\rho}_{iN}, \hat{P}_{jN}| \geq 1$, which is stable and minimum variance. Setting $\hat{Z}_t^{(d)} = \hat{Z}_t^{(k,hs)}$ this can be used to identify —by means of standard techniques such as correlation functions and information criteria— the ARMA filter of the stationary component $\{z\}$.

Some statistical remarks are necessary in order to justify the above method.

1) Assuming $d = 1$, estimates at step 1 are super-consistent even though the stable series $\{z_t\}$ is not white noise. This follows by the decomposition $\hat{\Phi}_{1N} = \Phi_1 + (\sum_t z_t Z_{t-s})(\sum_t Z_{t-s}^2)^{-1}$ where, despite the possible correlation between $\{z_t, Z_{t-s}\}$, the last term tends to zero. If $d > 1$ the properties of estimates at Step 1 are uncertain, but they should exclude paradoxes like $|\hat{\Phi}_{1N}| < 1$ or $\{\hat{Z}_t^{(s)}\}$ stationary, so that one may proceed to Step 2.

2) Properties of OLS estimator in presence of $d > 1$ unstable roots have not been investigated. However, consistency holds as a consequence of the equivalence between any AR(p) model and the vector AR(1) system $Z_t = P Z_{t-1} + a_t$ —where $Z_t' = [Z_t \dots Z_{t-p+1}]$, $P' = [\Phi : I_{p-1}]$ with $\Phi' = [\phi_1, \dots, \phi_p]$, $a_t' = [a_t, 0 \dots 0]$— and the convergence of the LS-estimator $\hat{P}_N$ in case of instability of the matrix P (see Anderson, 1959).

3) Extension of the Slutsky theorem enables us to prove the convergence of the roots' estimators $\hat{P}_{iN}$, $\hat{\rho}_{iN}$ at Step 2 and the convergence of the series $\hat{Z}_t^{(d)}$ at Step 3. More precisely, if the order d is rightly identified, from the relation $\hat{P}_N = P + o_p(1)$ it follows that $\hat{Z}_t^{(d)} = z_t + o_p(1)$, and if $d = 1$ we have super-consistency : $\hat{Z}_t^{(d)} = z_t + o_p[1/g(N)]$.

14

Obviously, presence of several unstable roots should be avoided since it complicates the identification problem ; there are, however, a number of practical reasons in favour of the hypothesis $d \equiv 1$. For example, the size of an unstable root may be arbitrarily large making unnecessary the presence of other unstable factors; as seen in the simulation of Figure 1 (c,d) realizations of processes with two unstable roots are smooth like deterministic functions and cannot represent empirical data; finally, as seen in the airline application a single unstable factor may simultaneously capture several kinds of nonstationary components, such as trends, cycles and periodic oscillations.

Simulations - In conclusion we investigate, by simulations, the behaviour of LS estimators applied to unstable models of order 2 : $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$ with $a_t \sim IN(0,1)$ and factors $(1 - \rho_2 B)(1 - \rho_1 B) Z_t = a_t$. Using the recursive algorithm (2.10) we investigate the convergence of $\hat{\phi}_{1N}, \hat{\phi}_{2N}$ in presence of one and two unstable roots, and the properties of the direct estimator of the first root ρ_1. Obviously, detection of super-consistency for $\hat{\rho}_{1N}$ would confirm the validity of the previous identification strategy.

1) The first simulation regards a model with parameters $\phi_1 = .3$, $\phi_2 = .8$ which has one unstable factor since $\rho_1 = 1.057$, $\rho_2 = -.757$. Figure 6a plots recursive estimates of AR parameters ϕ_1, ϕ_2, while Figure 6b shows estimates of the unstable root $\rho_1 = 1.057$ in four independent replications. As we may see, despite the presence of autocorrelated residuals $z_t = \rho_2 z_{t-1} + a_t$ only in the second case do we have clear evidence of super-consistency.

Figure 6 (a,b) - *Plot of RLS estimates of* $[\phi_1 \underline{\quad} ; \phi_2 \text{----}]$ *and* $[\rho_1 \underline{}, -, ...]$ *in simulation 1* .
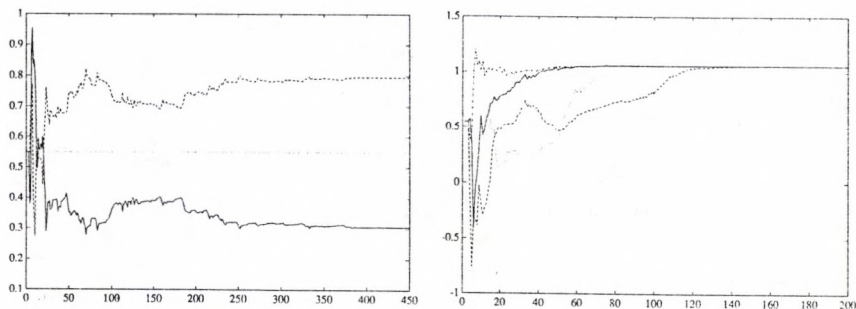
2) In the second simulation we have chosen $\phi_1 = 2.1$, $\phi_2 = -1.102$ which involve two unstable roots $\rho_1 = 1.075$, $\rho_2 = 1.025$. Figure 7a plots recursive estimates $\hat{\phi}_{1t}, -\hat{\phi}_{2t}$, and Figure 7b shows the path of $\hat{\rho}_{1t}$ in four independent replications. These results reproduce those of simulation 1, but with a speed of convergence that is twice as fast. It is certainly surprising that super-consistency of $\hat{\rho}_{1N}$ still holds (and sustantially improves) despite the fact that the corresponding residuals $Z_t^{(1)} = \rho_2 Z_{t-1}^{(1)} + a_t$ are unstable .

3) Finally, in order to clarify the effect of residual autocorrelation on super-consistency, we have estimated the model $Z_t = 1.05 Z_{t-1} + z_t$ with $z = a_t \sim IN(0,1)$ and $z_t = .75 z_{t-1} + a_t$. Recursive estimates, relative to two independent replications, are given in Figure 8 (a,b).

15

We may see that for $t > 20$ autocorrelation improves the convergence rate.

Figure 7 (a,b) - *Plot of RLS estimates of* $[\phi_1 \underline{\quad} ; -\phi_2 \text{---}]$ *and* $[\rho_1 \underline{\;}\text{--},...]$ *in simulation 2 .*
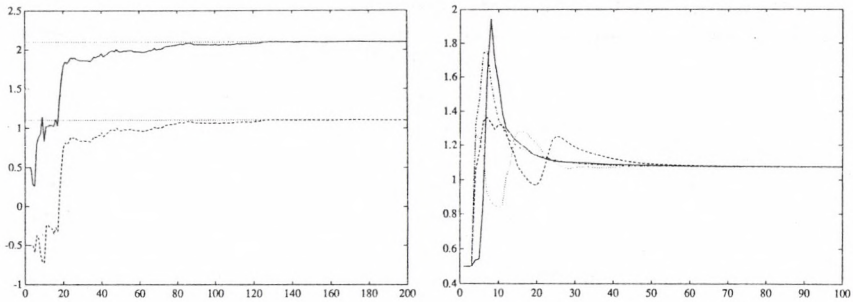
Figure 8 (a,b) - *Convergence of estimates* $\{\hat{\phi}_t\}$ *with* $[z_t \equiv a_t \underline{\quad} ; z_t \sim AR(1) \text{---}]$.
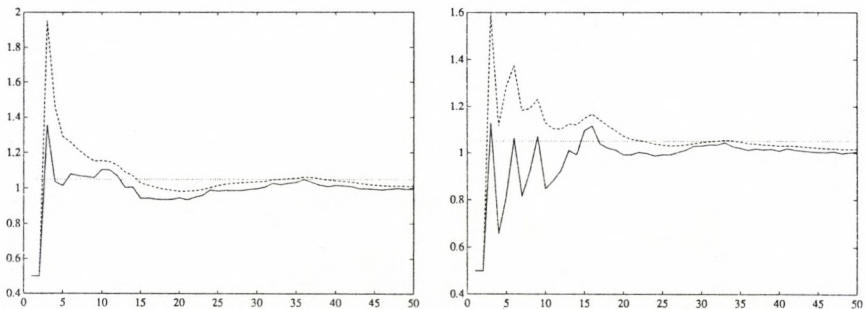
3 . Unstable Nonstationary Time Series

In this section we develop a general framework for modeling nonstationary time series with time-varying coefficients. The basic representation is linear ARMA but parameters are assumed to be general functions of past observations. Since these may wander indefinitely outside the stability regions of the constant parameter models, the treatment of the previous section is naturally extended. Basic tools for estimation are provided by recursive algorithms of type (2.10); however, more general methods exist and suitable optimization rule may be defined for their design.

3.1 Models - Given a stochastic process $z_t \sim \text{ARMA}(p,q)$ with input $a_t \sim \text{IN}(0, \sigma^2 < \infty)$ and constant drift μ, a general extension toward non-linearity and non-stationarity may be obtained by taking any bounded function $h(\cdot)$ of its regressors $\{z_{t-i}, a_{t-j}\}$:

16

$$z_t = h(1, z_{t-1}, \ldots, z_{t-p}, a_{t-1}, \ldots, a_{t-q}; \beta_t) + a_t, \qquad a_t \sim \text{IID}(0, \sigma_t^2) \qquad (3.1)$$

where $\{\beta_t, \sigma_t\} = O(1)$ are sequences of bounded deterministic parameters. The above representation cannot, however, be estimated even by modern techniques of non-parametric regression since the regressors $\{a_{t-j}\}$ are non-observable and the structure of $h(\cdot)$ changes over time. To establish a compromise between generality and tractability it is necessary to follow a *linear-functional* approach, by considering the parameters of the ARMA model as bounded functions of past events. By denoting the space of past realizations as $\Im_{t-1} = \{z_{t-1}, z_{t-2}, \ldots\}$, we may define

$$z_t = \mu(\Im_{t-1}) + \sum_{i=1}^{p} \phi_i(\Im_{t-1}) z_{t-i} + \sum_{j=1}^{q} \theta_j(\Im_{t-1}) a_{t-j} + a_t, \quad (a_t \mid \Im_{t-1}) \sim \text{IID}[0, \sigma^2(\Im_{t-1})] \quad (3.2)$$

A possible analytical relationship between expressions (3.1) and (3.2) follows from a Taylor expansion of $h(x_t; \beta_t)$, $x_t' = [1, z_{t-1} \ldots, \ldots, a_{t-q}]$ at a generic point x_τ, $\tau < t$

$$z_t = f(x_\tau) + \sum_{i=1}^{p} \phi_i(x_\tau)(z_{t-i} - z_{\tau-i}) + \sum_{j=1}^{q} \theta_j(x_\tau)(a_{t-j} - a_{\tau-j}) + a_t$$

where $\phi_i(\cdot)$, $\theta_j(\cdot)$ are elements of the vector of first derivatives $\partial h(x_\tau)/\partial x_\tau$. Apart from this loose connection, the representation (3.2) has an autonomous meaning; in particular we have the following characterizations :

i) model (3.2) is fundamentally nonlinear in the variables; however, conditional on the set of past events it becomes linear with time-varying coefficients. Assuming, for example, $(a_t \mid \Im_{t-1})$ normal, it follows that $(z_t \mid \Im_{t-1}) \sim \text{N}[\mu(\Im_{t-1}), \gamma(\Im_{t-1})]$ is a conditionally gaussian and non-stationary process (see Tjostheim, 1986a).

ii) The structure of (3.2) may be further generalized by including the influence of an exogenous process $\{e_t\}$ on the behaviour of parameters. In this case, if $\{a_t, e_t\}$ are orthogonal, the space of events becomes $\Im_t = \Im_t^a \oplus \Im_t^e$ where $\oplus$ is the direct sum operator. This situation occurs, for instance, in doubly stochastic systems in which parameters follows dynamics of the type $\beta_t = \beta_{t-1} + \delta a_{t-1} + e_t$ for every $\beta = \mu, \phi, \theta$.

Necessary conditions for the stochastic stability of models of type (3.2) obviously require that parameters $\beta_k(\Im_{t-1})$ be uniformly bounded in probability. Sufficient conditions can be obtained from the associated markovian representation

$$\begin{pmatrix} z_t \\ z_{t-1} \\ \cdots \\ z_{t-p+1} \end{pmatrix} = \begin{pmatrix} \mu_t \\ 0 \\ \cdots \\ 0 \end{pmatrix} \begin{pmatrix} \phi_{1t} & \cdots & \phi_{pt} \\ 1 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & ..1 & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \cdots \\ z_{t-p} \end{pmatrix} + \begin{pmatrix} \theta_{1t} & \cdots & \theta_{qt} \\ -1 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & ..-1 & 0 \end{pmatrix} \begin{pmatrix} a_{t-1} \\ a_{t-2} \\ \cdots \\ a_{t-q} \end{pmatrix} + \begin{pmatrix} a_t \\ a_{t-} \\ \cdots \\ a_{t-q+1} \end{pmatrix}$$

namely

$$z_t = \mu(\Im_{t-1}) + \Phi(\Im_{t-1}) z_{t-1} + \Theta(\Im_{t-1}) a_{t-1} + a_t$$

which solved for z_t, $t > 0$ with initial condition $z_0 = (\mu_0 + a_0)$ gives

17

$$z_t = \sum_{i=0}^{t} \prod_{j=1}^{t-i} \Phi_{i+j}(\mu_i + a_i) + \sum_{i=0}^{t-1} \prod_{j=2}^{t-i} \Phi_{i+j}\Theta_{i+1}a_i$$

Thus, if $\{a_t\}$ is uniformly bounded in probability : $\sup_t P(|a_t| \geq M_\varepsilon) \leq \varepsilon$ small, the vector process $\{z_t\}$ retains the same property if the stochastic matrices $\{\Phi_t\}$ are uniformly stable in probability. Letting $\lambda_k(\mathfrak{I}_{t-1}), k = 1 \ldots p$ be the eigenvalues of $\Phi(\mathfrak{I}_{t-1})$, the condition formally becomes $P\{|\lambda_k(\mathfrak{I}_{t-1})| \geq 1\} = 0$ for all $k, t > 0$. By defining S_p as the region of stability and stationarity of a standard AR(p) model, sufficient conditions for the stochastic stability of (3.2) can be restated as

$$P[\phi_i(\mathfrak{I}_{t-1}) \in S_p, i = 1, 2 \ldots p] = \begin{cases} \leq 1, & t \in T_i < \infty \\ \equiv 1, & \text{elsewhere} \end{cases}$$

that is the realizations of autoregressive coefficients ϕ_{it} may locally exit from the stability region S_p, only for finite periods of time T_i.

Although sufficient, these conditions are very demanding since they bind every realization of a vector of stochastic parameters. On the other hand, there is the need to deal with non-stationary processes $\{Z_t\}$ which are also unstable. This is possible by including in representation (3.2) linear factors which have roots wandering outside the unit circle :

$$\prod_{i=1}^{d} [1 - \rho_i(\mathfrak{I}_{t-1})B] Z_t = z_t, \qquad \overline{E}|\rho_i(\mathfrak{I}_{t-1})| = \lim_{T \to \infty}\left[\frac{1}{T}\sum_{t=1}^{T} E|\rho_i(\mathfrak{I}_{t-1})|\right] \geq 1 \qquad (3.3)$$

The fluctuations of unstable evolving roots $\rho_i(\mathfrak{I}_{t-1})$ can determine trends and cycles with complex transitory components, such as structural breaks and inverting slopes. With respect to unobserved component models of type (1.1), the advantage of representation (3.2)-(3.3) is that dynamics of parameters is not specified a-priori. In this way, it may capture situations in which stable roots become unstable (in probability) and vice versa.

The price to be paid for the adaptability of the above framework, is in forecasting. Given the functional dependence of parameters β_k from the process $\{z_t\}$, the definition of the exact expression of the predictor optimal in mean square sense is difficult. It is then necessary the resort to approximations which refer to existing algorithms

$$\hat{z}_t(h) = E[z_{t+h}|\mathfrak{I}_t] \approx \hat{\mu}_t(h-1) + \sum_{i=1}^{h-1} \hat{\phi}_{it}(h-1)\hat{z}_t(h-i) + \sum_{i=h}^{p} \hat{\phi}_{it}(h-1)\hat{z}_{t+h-i} + \sum_{j=h}^{q} \hat{\theta}_{jt}(h-1)\hat{a}_{t+h-j}$$

Consistently with adaptive forecasting methods, the predictors of parameters $\beta_{kt}(h-1) = E[\beta_k(\mathfrak{I}_{t+h-1})|\mathfrak{I}_t]$ may in turn be approximated by the last parameter estimates $\beta_k(t)$ available from recursive algorithms of type (2.10).

The situation may be substantially improved if, on the basis of the available recursive estimates $\beta(t)$, a model for $\{\beta_t\}$ is identified and used in forecasting. This approach belongs to multilayer (or hierarchical) systems in which the various levels are treated conditional on the set of information obtained in the lower layers. For a 3-layer system in which information spreads as in a waterfall, we have

18

$$z_t = \mu_t + \sum_{i=1}^{p_1} \phi_{it} z_{t-i} + \sum_{j=1}^{q_1} \theta_{jt} a_{t-j} + a_t$$

$$= x_t' \beta_t + a_t \qquad \text{(say)}$$

$$\beta_t = \eta_t + \sum_{i=1}^{p_2} \Phi_{it} \beta_{t-i} + \sum_{k=1}^{r_2} \Delta_{kt} x_{t-k-b_2} + \sum_{j=1}^{q_2} \Theta_{jt} e_{t-j} + e_t$$

$$= X_t' B_t + e_t \qquad \text{(say)}$$

$$B_t = N + \sum_{i=1}^{p_3} F_i B_{t-i} + \sum_{k=1}^{r_3} H_k X_{t-k-b_3} + \sum_{j=1}^{q_3} G_j U_{t-j} + U_t$$

with a suitable choice of the dimension of matrices. In the above system $\{a_t, e_t, U_t\}$ are scalar, vector and matrix white noises, which are pairwise orthogonal ; some of the roots of the stochastic polynomials $\phi_t(B^{-1})$, $\text{Det}[\Phi_t(B^{-1})]$ may wander outside the unit circle; finally, stationarity of regression coefficients is reached in the last layer.

Under regularity conditions, the selection of the orders (p_i, q_i, r_i) and of the lags $b_j, j = 2, 3$, may be obtained by typical off-line methods such as correlation functions and information criteria. An essential condition for correct identification is, however, that parameters have an average value different from zero : $0 < |\overline{E}[\beta_k(\Im_{t-1})]| < \infty$, where the operator $\overline{E}(\cdot)$ is defined in (3.3). The number of layers to consider depends on the possibility of obtaining a series of recursive estimates $\{\hat{B}_t\}$ which conforms to a model with constant parameters, although possibly unstable. Finally, the computation of the predictor proceeds in the opposite direction to that of building the system, that is starting from the last layer and using standard algorithms.

3.2 Estimation - Given a non-stationary ARMA model $z_t = \beta_t' x_t + a_t$, where $x_t' = [1, z_{t-1} \ldots a_{t-q}]$ and $\beta_t' = [\mu_t, \phi_{1t} \ldots \theta_{qt}]$, the general form taken by the recursive non-linear estimator $\beta(t)$ of parameters $\beta(\Im_t)$ is given by

$$\beta(t) = \beta(t-1) + \delta(t), \qquad \delta(t) = \hat{\Gamma}(t) \hat{\xi}(t) \bar{a}(t) \tag{3.4}$$

(see Ljung & Söderström, 1983 or Young, 1984), where $\hat{\Gamma}(t) = \hat{\Gamma}(t-1) - \hat{\Delta}(t)$ is the gain matrix, $\hat{\xi}(t) = [-\partial a_t / \partial \beta_t |_{\beta(t-1)}]$ is the gradient and $\bar{a}(t) = [z_t - \hat{x}(t)' \beta(t-1)]$ is the prediction error. It is worth noting that the gain matrix may coincide with the information (or covariance) matrix of the estimator and its updating mechanism is recursive like that of $\beta(t)$; the quantity $\hat{\Delta}(t)$ will be defined later. The one-step-ahead prediction error $\bar{a}(t)$ has a substantially different nature from the residual of regression $\hat{a}(t) = [z_t - \beta(t)' \hat{x}(t)]$: the first is used in loss functions and measures of validation, while the second updates the vector of regressors $\hat{x}(t) = [1, z_t \ldots \hat{a}(t-q)]$. Finally, in case of ARMA models the analytical expression of the gradient is $\xi_t = x_t / \theta_t(B)$ and may easily be computed on-line as $\hat{\xi}(t) = \hat{x}(t) - \sum_{j=1}^{q} \hat{\theta}_j(t-1) \hat{\xi}(t-j)$ or approximated with the vector of regressors $\hat{\xi}(t) \approx \hat{x}(t)$.

What fundamentally distinguishes the various algorithms is the way of updating the gain matrix. As a general remark, the necessary condition which enables (3.4) to track the parameter changes $\delta_t = [\beta_t - \beta_{t-1}]$ is that $\hat{\Gamma}(t)$ be positive definite and uniformly bounded in probability. Formally, for any $\varepsilon > 0$ there must exist positive numbers $m_\varepsilon, M_\varepsilon$ (which

depend only on ε) such that

$$0 < m_\varepsilon \le M_\varepsilon < \infty, \qquad P[\, m_\varepsilon I \le \hat{\Gamma}(t) \le M_\varepsilon I \,] \ge 1 - \varepsilon, \qquad t = 1, 2, 3 \ldots \tag{3.5}$$

In system identification engineering, several strategies have been implemented in order to satisfy condition (3.5) starting from a deterministic initial condition $mI < \hat{\Gamma}(0) < MI$. These strategies can classified into two groups, according to the fact that they intervene directly or indirectly on the updating mechanism $\hat{\Gamma}(t) = \hat{\Gamma}(t-1) - \hat{\Delta}(t)$.

The first group includes recursive nonlinear least squares (RNLS) algorithms in which observations are weighted with sequences of deterministic weights. If this operation is performed on past observations with exponentially decaying weights $\{\lambda_{t-k}\}$ we have exponential forgetting (EF), while if it is performed on new observations with positive weights $\{\mu_t\}$ we have directional forgetting (DF). It is possible to conceive a filter which executes both operations

$$\hat{\Gamma}(t)^{-1} = \lambda \cdot \hat{\Gamma}(t-1)^{-1} + \mu \cdot \xi(t)\xi(t)', \qquad 0 < \mu, \lambda < 1 \tag{3.6a}$$

$$\hat{\Gamma}(t) = \frac{1}{\lambda} \left[\hat{\Gamma}(t-1) - \frac{\hat{\Gamma}(t-1)\xi(t)\xi(t)'\hat{\Gamma}(t)}{\lambda/\mu + \xi(t)'\hat{\Gamma}(t-1)\xi(t)} \right] \tag{3.6b}$$

in this case the estimator (3.4) becomes $\beta(t) = \beta(t-1) + \mu \hat{\Gamma}(t)\zeta(t)\bar{a}(t)$. While the logic of discounting past observations is immediate, the meaning of the directional factor μ is that of preserving an ideal information matrix, by selecting new observations. However, as we shall see in the sequel, this task may effectively be accomplished only by a time varying sequence.

The second group of strategies intervenes on the gain matrix by adding fixed positive quantities in such a way to satisfy condition (3.5). The best known method is that inspired by the extended Kalman filter (EKF) where

$$\hat{\Gamma}(t) = \hat{\Gamma}(t-1) - \frac{\hat{\Gamma}(t-1)\xi(t)\xi(t)'\hat{\Gamma}(t)}{\sigma^2 + \xi(t)'\hat{\Gamma}(t-1)\xi(t)} + \Sigma, \qquad 0 < \sigma^2, \Sigma < \infty \tag{3.7}$$

which is consistent with the assumption of random walk parameters $\beta_t = \beta_{t-1} + e_t$ with gaussian noises $e_t \sim IN(0, \Sigma)$, $a_t \sim IN(0, \sigma^2)$. Note that the variance σ^2 in (3.7) plays the same role as does the factor μ in (3.6); moreover, both parameters are not essential in satisfying the tracking condition (3.5).

At this point a *general* recursive algorithm can simply be obtained by recomposing equations (3.6) and (3.7), i.e. equating $\lambda/\mu = \sigma^2$ and adding the matrix Σ to (3.6). However, the resulting algorithm suffers from redundancy and non-linearity of certain parameters. Simple algebraic adjustments lead to a more flexible and specialized structure which should have a greater adaptive capability, namely

$$\hat{\Gamma}(t) = \frac{1}{\lambda}\hat{\Gamma}(t-1) - \mu \left[\frac{\hat{\Gamma}(t-1)\xi(t)\xi(t)'\hat{\Gamma}(t)}{1 + \xi(t)'\hat{\Gamma}(t-1)\xi(t)} \right] + \ddot{\Gamma}, \qquad \ddot{\Gamma} = \text{diag}[\gamma_1, \ldots \gamma_{p+1+q}] \tag{3.8}$$

where the choice of $\hat{\Gamma} = \text{diag}$ tries to establish a compromise between flexibility ($\hat{\Gamma} = \Sigma$) and parsimony ($\hat{\Gamma} = \gamma I$). At the interpretative level, estimator (3.4)-(3.8) is consistent with an hypothesis of evolution of parameters given by the sum of stochastic components with different dynamics. By defining the set $\Im_t^i = \{z_i, \ldots, z_s\}$, we may write, in fact, $\beta(t) = f_\lambda(\Im_{t-1}^{i-\tau}) + f_\mu(\Im_{t-1}^{i-m}) + f_\gamma(\Im_t^i)$ where $f_\lambda(\cdot)$ is a function which depends on coefficient λ and $m = (1 + p + q)$ is the dimension of $\xi(t)$.

Useful extensions of the unified algorithm can be obtained from techniques recently introduced in signal processing, such as multistep filtering (see Benveniste, 1987) and on-line robustification. Both these modifications regard the structure of the estimation equation (3.4); in particular, they intervene on the increments $\delta(t)$. Multistep extension introduces dynamics by multiplying δ_t by a stable filter $\alpha(B)$; on-line robustification regularizes estimates by censoring anomalous prediction errors $\bar{a}(t)^*$. If dynamics are second order and robustification is based on the two-sigma rule, the estimation equation becomes

$$\beta(t) = \beta(t-1) + (1 + \alpha_1 B + \alpha_2 B^2) \cdot \hat{\Gamma}(t)\xi(t)\,[\bar{\mu}(t)\bar{a}(t)] \qquad (3.9a)$$

$$\bar{\mu}(t) = \begin{cases} 1 & \text{if} \quad |\bar{a}(t)| < 2\tilde{\sigma}(t-1) \\ 2\tilde{\sigma}(t-1)|\bar{a}(t)|^{-1} & \text{if} \quad |\bar{a}(t)| \geq 2\tilde{\sigma}(t-1) \end{cases} \qquad (3.9b)$$

$$\tilde{\sigma}(t)^2 = \lambda\,\tilde{\sigma}(t-1)^2 + (1-\lambda)\,[\bar{\mu}(t)\bar{a}(t)]^2 \qquad (3.9c)$$

In this framework, robustification is consistent with assumptions of smooth evolution of the regression coefficients and of the innovation variance $\tilde{\sigma}_t^2 = E(\bar{a}^2)$. Indeed, filter (3.9c) is a robust adaptive estimator of such a variance since written in a compact form it becomes $\tilde{\sigma}(t)^2 = (1-\lambda) \sum_{\tau=1}^{t} \lambda^{t-\tau} [\bar{\mu}(t)\bar{a}(t)]^2$ where $(1/\sum_{\tau=1}^{t} \lambda^{t-\tau}) \to (1-\lambda)$ as $t \to \infty$.

It is important to stress the meaning of the censoring factor $\bar{\mu}(t)$ since, as regards the RLS algorithm (3.6), it has a role similar to the directional forgetting coefficient μ. However, in the context of the covariance matrix (3.6b), the efficacy of the factor $\lambda/\bar{\mu}(t)$ in terms of regularization of recursive estimates and the fulfilment of condition (3.5) is not important since it is absorbed by other stochastic quantities. Moreover, the gain matrix (3.8) does not coincide with the covariance of estimates $\beta(t)$ and the role of coefficient μ in (3.8) is not the same as that in (3.6). These remarks lead us to conclude that if outliers are not present in observable data $\{Z_t\}$, but depend on parameter changes, then on-line robustification may be concerned with the estimation equation (3.4) alone.

Algorithm (3.8)-(3.9) involves $5 + p + q$ unknown coefficients, whose range of variation, except for $0 < \lambda < 1$, is somewhat wide. Until now, in system identification literature, only heuristic rules have been provided for their design; using experimental information their ideal ranges have been defined as $\lambda \in [.95, .99]$, $\mu \in [.1, .5]$, $\gamma \in [.0, .1]$ (see Ljung & Söderström, 1983). Given a sample realization $\Im_N = \{Z_1, \ldots, Z_N\}$, it is however appropriate to face the problem in terms of parametric estimation, by optimizing a suitable loss function. Squared prediction errors and recursive residuals provide natural building blocks, but it must be recalled their nature is substantially different. While the variance

of the *a-posteriori errors* $\hat{a}(t)$ may be reduced to zero by accordingly increasing $\mathrm{Det}\,\hat{\Gamma}(t)$ (i.e. $\hat{\Gamma}$), the same adjustment is never possible for the *a-priori errors* $\bar{a}(t)$. In practice, the two kinds of errors coincide in a situation of constant parameters, but in the case of non-stationarity only prediction errors retain a meaning in terms of goodness of fit. A sensible criterion to the estimation of the tracking coefficients of algorithm (3.8)-(3.9) is therefore

$$[\hat{\lambda},\hat{\mu},\hat{\gamma}_1,\hat{\alpha}_1,\hat{\alpha}_2;\hat{\gamma}_0,\beta_0,\hat{\sigma}_0]_N = \arg\min\left\{ Q_N = \sum_{t=p+1}^{N} [z_t - \hat{x}(t)'\beta(t-1)]^2 \right\} \tag{3.10}$$

this belongs to conditional least squares (CLS) strategies discussed by Klimko & Nelson (1978) and Tjostheim (1986b), where conditional refers to the set of past information : $\bar{a}_t = [z_t - E(z_t \mid \mathfrak{S}_{t-1})]$. Since starting values $\beta(0), \hat{\Gamma}(0), \tilde{\sigma}(0)$ have a significant role on the statistical performance of recursive algorithms, they have been included in the estimation problem (3.10) as unknown coefficients $\beta_0, \gamma_0 I, \sigma_0$. The direct estimation of initial values entirely avoids problems of transient behaviour and error propagation involved in heuristic solutions such as applying diffuse priors for $\hat{\Gamma}(0)$ as suggested by Harvey & Peters (1990). There should not be specific problems of parametric identifiability involved by $\beta_0, \gamma_0, \sigma_0$; however, to avoid problems in this sense rules of parsimonious parametrizations, such as $\hat{\Gamma}(0) = \hat{\Gamma} = \gamma I$, or constraints on $\beta_0' = [\mu_0, \phi_{10}, \dots, \theta_{q0}]$, must be followed.

3.3 Application - In this section we apply the adaptive framework discussed above to the models introduced in sub-section (2.2). The solution of the estimation problem (3.10) applied to the airline model in additive form : $Z_t = \eta + \phi Z_{t-1} + \Phi Z_{t-12} + \theta a_{t-1} + \Theta a_{t-12} + a_t$ provided disappointing results. Despite the statistical significance of many tracking coefficients, the value of the statistic $Q_N = 14,520$ remained superior to that of model (2.5b) with constant parameters. Thus, a time varying parameter modeling did not seem appropriate for the airline data-set, at least without differencing the series. In order to check this first impression we have focused on the first order model : $Z_t = \Phi_t Z_{t-12} + z_t$; estimation results are given in Table 1

Table 1 - *CLS estimates of the coefficents of algorithm (3.8)-(3.9) applied to model (2.6a).*

| Coeff.	Φ_0	γ_0	σ_0^2	λ	μ	γ_1	α_1	α_2	Q_N
Estim.	.986 (419.)	.350 (21.)	2.281 (179.)	.985 (930.)	.453 (87.)	.18E−4 (25.)	.471 (41.)	−.570 (48.)	16,149

In this case, the results are very encouraging since the statistical performance in terms of Q_N improves by about 50% over the constant parameter model (2.6a). This is a result of absolute merit since, unlike standard robustification methods, the statistic Q_N in (2.10) is not concerned with censored prediction errors $[\tilde{\mu}(t)\bar{a}(t)]^2$.

As a comment to the values of Table 1 we may note that tracking coefficients belong to their ideal ranges; however, even small changes by the more significant among them (such as λ) may drastically worsen Q_N. The contribution of coefficient γ_1, which accounts

22

for the component $\Sigma = \gamma_t I$ of the extended Kalman filter (2.7), is not very important. This means that the dynamics of parameter $\{\Phi_t\}$ are more complex than the random walk and updating of $\hat{\Gamma}(t)$ based on the addition of fixed quantities is too rigid. By contrast, a significant role is played by coefficient σ_0 which concerns the robustification mechanism; without robustification the value of Q_N becomes $18,130$.

Figure 8 shows the paths of basic components of algorithm (3.8)-(3.9) implemented with coefficients of Table 1 . We have : (a) parameters estimates $\hat{\Phi}(t)$; (b) gain "matrix" $\hat{\Gamma}(t)$; (c) censored errors $\tilde{\mu}(t)\bar{a}(t)$; (d) robustified increments $\hat{\delta}(t)$; (e) prediction errors and $\pm 2\tilde{\sigma}(t)$ bands ; (f) recursive residuals $\hat{a}(t)$ and "residuals" $\{\hat{z}_t\}$ of model (2.6a).

Figure 8 (a,b,c,d,e,f) - *Time paths of components of algorithm (3.8)-(3.9) generated with coefficients of Table 1 : (a) $\hat{\beta}(t)$; (b) $\hat{\Gamma}(t)$; (c) $\tilde{\mu}(t)\bar{a}(t)$; (d) $\hat{\delta}(t)$; (e) $\bar{a}(t), \pm 2\tilde{\sigma}(t)$; (f) $\hat{a}(t), \hat{z}_t$.*
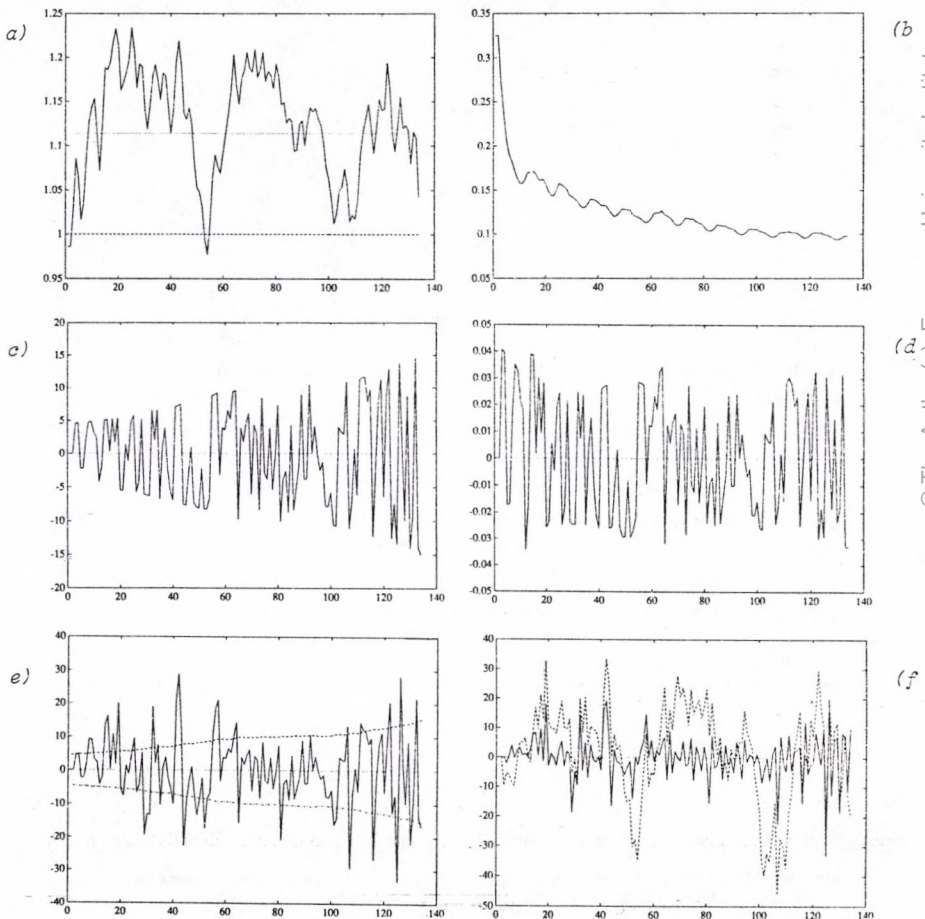
As a possible criticism to the above exercise, one may argue that a time-varying model $Z_t = \Phi_t Z_{t-12} + z_t$ which achieves the perfect fitting $z_t = 0$, $\forall t$ can be obtained simply by the parameter sequence $\hat{\Phi}_t^* = Z_t / Z_{t-12}$. In reality, this approach may be implemented only in first order models and, more importantly, it regards recursive residuals (i.e. a-posteriori errors) but not prediction errors. The value of statistic Q_N corresponding to this approach was indeed $\sum_t [Z_t - \hat{\Phi}_{t-1}^* Z_{t-12}] = 21,610$, significantly worse than that of Table 1. The reason for this "paradox" can be explained by Figure 9 which plots the series $\{\hat{\Phi}(t), \hat{\Phi}_t^*\}$: as we may see the pattern of the second is much more erratic. This confirms the importance, for the tracking and forecasting capabilities of adaptive algorithms, of regularization mechanisms such as robustification (3.9c).

Figure 9 - *Time paths of recursive estimates* $[\hat{\Phi}(t) \_\_\_ ; \hat{\Phi}_t^* ----]$.
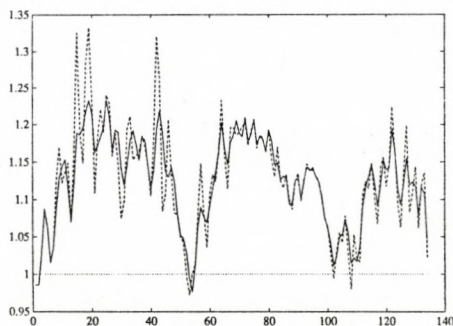
The attempt to improve results of Table 1 led us to consider more complex systems, such as

$$Z_t = \eta_t + \Phi_t Z_{t-12} + z_t , \qquad Z_t = \Phi_t Z_{t-12} + \Theta_t e_{t-12} + e_t \qquad (3.11)$$

Estimation results, obtained by minimizing (3.10) with the Davidson-Fletcher-Powel algorithm combined with the golden steplength search, are given in Table 2. Best performance in the estimation of the ARMA model (3.11) was provided by the pseudolinear implementation, which approximates the gradient as $\hat{\xi}(t) \approx \hat{x}(t) = [z_{t-12}, \hat{e}(t-12)]'$. The paths of the implied recursive estimates are shown in Figure 10 (a,b).

Table 2 - *CLS estimates of the coefficents of algorithm (3.8)-(3.9) applied to models (3.11).*

η_0 / Θ_0	Φ_0	γ_0	σ_0^2	λ	μ	γ_1	α_1	α_2	Q_N
1.522	.974	.250	2.25	.985	.421	.14E-4	.512	-.605	15,930
-.760	1.13	.021	1.92	.988	.581	.11E-3	-.244	.046	14,890

The improvements for the statistic Q_N in Table 2 as compared to Table 1 are not very significant and the constant parameter model (2.5a) still retains the best performance .

24

Figure 10 (a,b) - *Paths of recursive estimates of* $[\eta\_\_;\Phi\text{---}]$, $[\Phi\_\_;-\Theta\text{---}]$ *in models (3.11).*
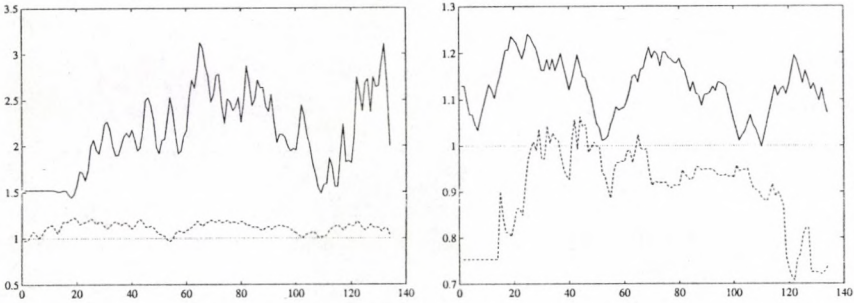
For this reason, we have chosen to model the "residual" series $\hat{z}_t = [Z_t - \hat{\Phi}(t-1)Z_{t-12}] = \bar{a}(t)$, displayed in Figure 8e, by an ARMA with constant parameters:

$$\hat{z}_t = \underset{(2.1)}{.183}\,\hat{z}_{t-9} - \underset{(1.9)}{.158}\,a_{t-4} - \underset{(5.9)}{.589} + a_t , \qquad Q_N = 11,893 \qquad (3.12)$$

The recursive estimation of this model has confirmed that the series $\{\hat{z}_t\}$ is nearly stationary in covariance, since it provides the values $Q_N = 11,524$ and $(\hat{\gamma}_0 = \hat{\gamma}_1) = 0.0$. However, the important point here is that the strategy of sequential filtering $Z_t \to z_t \to a_t$ seems essential for adaptive methods to monotonically improve the prediction statistic Q_N. Recall, in fact, that adaptive estimation of the joint model (2.5b) in additive form provided the disappointing value $Q_N = 14,520$.

In the context of multilayer modeling we have identified an ARMA model for the recursive estimates $\hat{\Phi}(t)$. Its structure is strongly influenced by the multistep filter $\alpha(B)$ included in the adaptive algorithm (3.9a) and by the MA(12) component left in the series $\hat{z}_t = \bar{a}(t)$:

$$\hat{\Phi}(t) = \underset{(43.)}{1.001}\,\hat{\Phi}(t-1) + \underset{(3.1)}{.291}\,e_{t-1} - \underset{(4.8)}{.438}\,e_{t-2} - \underset{(2.8)}{.267}\,e_{t-12} + e_t , \qquad Q_N = .068 \qquad (3.13)$$

Adaptive estimates of the above model without the MA(12) component are given in Table 3 (the tracking coefficients of the algorithm) and Figure 11 (the recursive estimates of parameters). These confirm that even the series $\hat{\Phi}(t)$ is nearly stationary and therefore the multilayer model building stops here.

Table 3 - *CLS estimates of the coefficents of algorithm (3.8)-(3.9) applied to model (3.13).*

Grad.	ϕ_0	θ_0^1	θ_0^2	γ_0	σ_0^2	λ	μ	γ_1	Q_N
ξ_t	1.008	.635	-.458	.413	.015	.944	.598	.007	.063
x_t	.998	.740	-.348	.154	.013	1.03	.241	.472	.061

Figure 11 - *Paths of recursive estimates of* $[\,\phi_1 \underline{\quad}\,;\theta_1 -\!\!-\,;\theta_2 ....\,]$ *in model (3.13).*
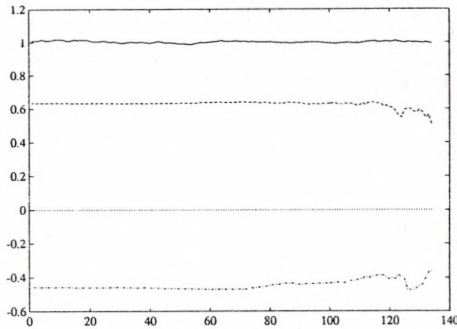

We conclude the application by evaluating the forecasting capability of the above framework. The sub-optimal adaptive predictor for series $\{Z_t\}$ is defined as

$$\hat{Z}_t^*(h) = \hat{\Phi}_t(h)\,\hat{Z}_t^*(h-12) + \hat{z}_t(h)\,, \qquad h = 1,2\ldots.$$

where $\hat{Z}_t^*(h-12) = Z_{t+h-12}$ for $h \le 12$ and $\hat{\Phi}_t(h)$, $\hat{z}_t(h)$ are standard predictors of the nearly stationary models (3.12)-(3.13). Figure 12 compares the MApE statistics (2.7) of the predictor $\hat{Z}_t^*(h)$ with those of the constant parameter models (2.4) (Box-Jenkins) and (2.5b) (unconstrained). The adaptive predictor uniformly provides the best result, but it should be recalled that this is obtained by a system which involves 15 parameters. On the other hand, some of these parameters do not seem essential, such as γ_i, α_j, and a good trade-off between efficacy and parsimony may be achieved. In this paper we have provided some examples of adaptive estimator modeling; the important conclusion is that such algorithms may run on time series containing unstable components. However, more radical modifications are still possible which may further improve the forecasting capability of dynamic models.

Figure 12 - *Plot of statistics (2.7) for models* $[(2.4) \underline{\quad}\,;(2.5b) -\!-\!-\,;(3.12,13) \text{-.-.-}]$.

26
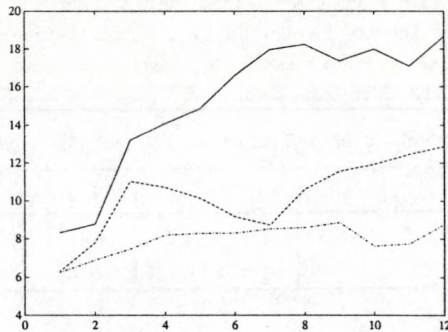
4 . References

ANDERSON T.W. (1959), On asymptotic distribution of estimates of parameters of stochastic difference equation. *Annals of Mathematical Statistics* **30**, 676-687.

BENVENISTE A. (1987), Design of one-step and multistep adaptive algorithms for the tracking of time varying systems. *Int. Jour. of Adaptive Control and Signals Processing* **1**, 3-29.

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

CRESSIE N. (1988), A graphical procedure for determining nonstationarity in a time series. *Jour. of American Statistical Association* **83**, 1108-1116.

CHAN N.H., WEI C.Z. (1987), Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* **15**, 1050-1063.

HARRISON P.J., STEPHENS C.F. (1976),Bayesian forecasting. *Jour. of Royal Statistic Society Ser.B* **38**, 205-248.

HARVEY A.C., PETERS S. (1990), Estimation procedures for structural time series models. *Jour. of Forecasting* **9**, 89-108.

KLIMKO L.A., NELSON P.I. (1978),On conditional least squares estimation for stochastic processes. *Annals of Statistics* **6**, 629-642.

LJUNG L., SöDERSTRöM T. (1983), *Theory and practice of recursive identification*. Cambridge (MA): MIT Press.

PARZEN E. (1982), ARARMA models for time series analysis and forecasting. *Jour. of Forecasting* **1**, 67-82.

PERRON P. (1989), The big crash, the oil price shock and the unit root hypothesis. *Econometrica* **57**, 1361-1401.

RAO M.M. (1961), Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Ann. of Mathematical Statistics* **32**, 195-218.

RAO M.M. (1978), Asymptotic distribution of an estimator of the boundary parameter af an unstable process. *Annals of Statistics* **6**, 185-190.

RUBIN H. (1950), Consistency of maximum likelihood estimates in the explosive case; in KOOPMANS T.C., *Statistical inference in dynamic economic models*, pp.356-364. New York: Wiley.

TJOSTHEIM D. (1986a), Some doubly stochastic time series models. *Jour.of Time Series Analysis* **7**, 51-72.

TJOSTHEIM D. (1986b), Estimation of nonlinear time series models. *Stochastic Processes and Applications* **21**,225-273.

WHITE J.S. (1958), The limiting distribution of of the serieal correlation coefficient in the explosive case I. *Ann.of Mathematical Statistics* **29**, 1188-1197.

WHITE J.S. (1959), The limiting distribution of of the serieal correlation coefficient in the explosive case II. *Ann. of Mathematical Statistics* **30**, 831-834.

WHITE H., DOMOWITZ I. (1984),Nonlinear regression with dependent observations. *Econometrica* **52**, 143-161.

YOUNG P.C. (1984), *Recursive estimation and time series analysis: an introduction*. New York: Springer-Verlag.

YOUNG P.C., NG C.N., LANE K., PARKER D. (1991), Recursive forecasting, smoothing and seasonal adjustent of non-stationary environmental data. *Jour. of Forecasting* **10**, 57-89.

EUI
WORKING
PAPERS

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

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

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

Please use order form overleaf

Publications of the European University Institute

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

From   Name . . . .
              Address. . . . .
              . . . .
              . . . .
              . . . .
              . . . .

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

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

No, Author   . . . .
Title: . . . .
No, Author   . . . .
Title: . . . .
No, Author   . . . .
Title: . . . .
No, Author   . . . .
Title: . . . .

Date   . . . .

            Signature   . . . .

Working Papers of the Department of Economics
Published since 1990

ECO No. 90/1
Tamer BASAR and Mark SALMON
Credibility and the Value of Information
Transmission in a Model of Monetary
Policy and Inflation

ECO No. 90/2
Horst UNGERER
The EMS – The First Ten Years
Policies – Developments – Evolution

ECO No. 90/3
Peter J. HAMMOND
Interpersonal Comparisons of Utility:
Why and how they are and should be
made

ECO No. 90/4
Peter J. HAMMOND
A Revelation Principle for (Boundedly)
Bayesian Rationalizable Strategies

ECO No. 90/5
Peter J. HAMMOND
Independence of Irrelevant Interpersonal
Comparisons

ECO No. 90/6
Hal R. VARIAN
A Solution to the Problem of
Externalities and Public Goods when
Agents are Well-Informed

ECO No. 90/7
Hal R. VARIAN
Sequential Provision of Public Goods

ECO No. 90/8
T. BRIANZA, L. PHLIPS and J.F.
RICHARD
Futures Markets, Speculation and
Monopoly Pricing

ECO No. 90/9
Anthony B. ATKINSON/ John
MICKLEWRIGHT
Unemployment Compensation and
Labour Market Transition: A Critical
Review

ECO No. 90/10
Peter J. HAMMOND
The Role of Information in Economics

ECO No. 90/11
Nicos M. CHRISTODOULAKIS
Debt Dynamics in a Small Open
Economy

ECO No. 90/12
Stephen C. SMITH
On the Economic Rationale for
Codetermination Law

ECO No. 90/13
Elettra AGLIARDI
Learning by Doing and Market Structures

ECO No. 90/14
Peter J. HAMMOND
Intertemporal Objectives

ECO No. 90/15
Andrew EVANS/Stephen MARTIN
Socially Acceptable Distortion of
Competition: EC Policy on State Aid

ECO No. 90/16
Stephen MARTIN
Fringe Size and Cartel Stability

ECO No. 90/17
John MICKLEWRIGHT
Why Do Less Than a Quarter of the
Unemployed in Britain Receive
Unemployment Insurance?

ECO No. 90/18
Mrudula A. PATEL
Optimal Life Cycle Saving With
Borrowing Constraints:
A Graphical Solution

ECO No. 90/19
Peter J. HAMMOND
Money Metric Measures of Individual
and Social Welfare Allowing for
Environmental Externalities

ECO No. 90/20
Louis PHLIPS/
Ronald M. HARSTAD
Oligopolistic Manipulation of Spot
Markets and the Timing of Futures
Market Speculation

* Working Paper out of print

ECO No. 90/21
Christian DUSTMANN
Earnings Adjustment of Temporary
Migrants

ECO No. 90/22
John MICKLEWRIGHT
The Reform of Unemployment
Compensation:
Choices for East and West

ECO No. 90/23
Joerg MAYER
U. S. Dollar and Deutschmark as
Reserve Assets

ECO No. 90/24
Sheila MARNIE
Labour Market Reform in the USSR:
Fact or Fiction?

ECO No. 90/25
Peter JENSEN/
Niels WESTERGÅRD-NIELSEN
Temporary Layoffs and the Duration of
Unemployment: An Empirical Analysis

ECO No. 90/26
Stephan L. KALB
Market-Led Approaches to European
Monetary Union in the Light of a Legal
Restrictions Theory of Money

ECO No. 90/27
Robert J. WALDMANN
Implausible Results or Implausible Data?
Anomalies in the Construction of Value
Added Data and Implications for Esti-
mates of Price-Cost Markups

ECO No. 90/28
Stephen MARTIN
Periodic Model Changes in Oligopoly

ECO No. 90/29
Nicos CHRISTODOULAKIS/
Martin WEALE
Imperfect Competition in an Open
Economy

✱✱✱

ECO No. 91/30
Steve ALPERN/Dennis J. SNOWER
Unemployment Through 'Learning From
Experience'

ECO No. 91/31
David M. PRESCOTT/Thanasis
STENGOS
Testing for Forecastible Nonlinear
Dependence in Weekly Gold Rates of
Return

ECO No. 91/32
Peter J. HAMMOND
Harsanyi's Utilitarian Theorem:
A Simpler Proof and Some Ethical
Connotations

ECO No. 91/33
Anthony B. ATKINSON/
John MICKLEWRIGHT
Economic Transformation in Eastern
Europe and the Distribution of Income*

ECO No. 91/34
Svend ALBAEK
On Nash and Stackelberg Equilibria
when Costs are Private Information

ECO No. 91/35
Stephen MARTIN
Private and Social Incentives
to Form R & D Joint Ventures

ECO No. 91/36
Louis PHLIPS
Manipulation of Crude Oil Futures

ECO No. 91/37
Xavier CALSAMIGLIA/Alan KIRMAN
A Unique Informationally Efficient and
Decentralized Mechanism With Fair
Outcomes

ECO No. 91/38
George S. ALOGOSKOUFIS/
Thanasis STENGOS
Testing for Nonlinear Dynamics in
Historical Unemployment Series

ECO No. 91/39
Peter J. HAMMOND
The Moral Status of Profits and Other
Rewards:
A Perspective From Modern Welfare
Economics

* Working Paper out of print

* Working Paper out of print

***

ECO No. 92/61
Sara CONNOLLY/John
MICKLEWRIGHT/Stephen NICKELL
The Occupational Success of Young Men
Who Left School at Sixteen

ECO No. 92/62
Pier Luigi SACCO
Noise Traders Permanence in Stock
Markets: A Tâtonnement Approach.
I: Informational Dynamics for the Two-
Dimensional Case

ECO No. 92/63
Robert J. WALDMANN
Asymmetric Oligopolies

ECO No. 92/64
Robert J. WALDMANN /Stephen
C. SMITH
A Partial Solution to the Financial Risk
and Perverse Response Problems of
Labour-Managed Firms: Industry-
Average Performance Bonds

ECO No. 92/65
Agustín MARAVALL/Víctor GÓMEZ
Signal Extraction in ARIMA Time Series
Program SEATS

ECO No. 92/66
Luigi BRIGHI
A Note on the Demand Theory of the
Weak Axioms

ECO No. 92/67
Nikolaos GEORGANTZIS
The Effect of Mergers on Potential
Competition under Economies or
Diseconomies of Joint Production

ECO No. 92/68
Robert J. WALDMANN/
J. Bradford DE LONG
Interpreting Procyclical Productivity:
Evidence from a Cross-Nation Cross-
Industry Panel

ECO No. 92/69
Christian DUSTMANN/John
MICKLEWRIGHT
Means-Tested Unemployment Benefit
and Family Labour Supply: A Dynamic
Analysis

ECO No. 92/70
Fabio CANOVA/Bruce E. HANSEN
Are Seasonal Patterns Constant Over
Time? A Test for Seasonal Stability

ECO No. 92/71
Alessandra PELLONI
Long-Run Consequences of Finite
Exchange Rate Bubbles

ECO No. 92/72
Jane MARRINAN
The Effects of Government Spending on
Saving and Investment in an Open
Economy

ECO No. 92/73
Fabio CANOVA and Jane MARRINAN
Profits, Risk and Uncertainty in Foreign
Exchange Markets

ECO No. 92/74
Louis PHLIPS
Basing Point Pricing, Competition and
Market Integration

ECO No. 92/75
Stephen MARTIN
Economic Efficiency and Concentration:
Are Mergers a Fitting Response?

ECO No. 92/76
Luisa ZANCHI
The Inter-Industry Wage Structure:
Empirical Evidence for Germany and a
Comparison With the U.S. and Sweden

ECO NO. 92/77
Agustín MARAVALL
Stochastic Linear Trends: Models and
Estimators

ECO No. 92/78
Fabio CANOVA
Three Tests for the Existence of Cycles
in Time Series

ECO No. 92/79
Peter J. HAMMOND/Jaime SEMPERE
Limits to the Potential Gains from Market
Integration and Other Supply-Side
Policies

ECO No. 92/80
Víctor GÓMEZ and Agustín
MARAVALL
Estimation, Prediction and Interpolation
for Nonstationary Series with the
Kalman Filter

ECO No. 92/81
Víctor GÓMEZ and Agustín
MARAVALL
Time Series Regression with ARIMA
Noise and Missing Observations
Program TRAM

ECO No. 92/82
J. Bradford DE LONG/ Marco BECHT
"Excess Volatility" and the German
Stock Market, 1876-1990

ECO No. 92/83
Alan KIRMAN/Louis PHLIPS
Exchange Rate Pass-Through and Market
Structure

ECO No. 92/84
Christian DUSTMANN
Migration, Savings and Uncertainty

ECO No. 92/85
J. Bradford DE LONG
Productivity Growth and Machinery
Investment: A Long-Run Look, 1870-
1980

ECO NO. 92/86
Robert B. BARSKY and J. Bradford
DE LONG
Why Does the Stock Market Fluctuate?

ECO No. 92/87
Anthony B. ATKINSON/John
MICKLEWRIGHT
The Distribution of Income in Eastern
Europe

ECO No.92/88
Agustín MARAVALL/Alexandre
MATHIS
Encompassing Unvariate Models in
Multivariate Time Series: A Case Study

ECO No. 92/89
Peter J. HAMMOND
Aspects of Rationalizable Behaviour

ECO 92/90
Alan P. KIRMAN/Robert
J. WALDMANN
I Quit

ECO No. 92/91
Tilman EHRBECK
Rejecting Rational Expectations in Panel
Data: Some New Evidence

ECO No. 92/92
Djordje Suvakovic OLGIN
Simulating Codetermination in a
Cooperative Economy

ECO No. 92/93
Djordje Suvakovic OLGIN
On Rational Wage Maximisers

ECO No. 92/94
Christian DUSTMANN
Do We Stay or Not? Return Intentions of
Temporary Migrants

ECO No. 92/95
Djordje Suvakovic OLGIN
A Case for a Well-Defined Negative
Marxian Exploitation

ECO No. 92/96
Sarah J. JARVIS/John
MICKLEWRIGHT
The Targeting of Family Allowance in
Hungary

ECO No. 92/97
Agustín MARAVALL/Daniel PEÑA
Missing Observations and Additive
Outliers in Time Series Models

ECO No. 92/98
Marco BECHT
Theory and Estimation of Individual and
Social Welfare Measures: A Critical
Survey

ECO No. 92/99
Louis PHLIPS and Ireneo M'guel
MORAS
The AKZO Decision: A Case of
Predatory Pricing?

ECO No. 92/100
Stephen MARTIN
Oligopoly Limit Pricing With Firm-
Specific Cost Uncertainty

* Working Paper out of print

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

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

✳✳✳

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