REVISTA DE ECONOMÍA DEL ROSARIO, 4 (DICIEMBRE 2001) 117-142

PERSISTENCE IN MACROECONOMIC TIME SERIES: IS IT A MODEL INVARIANT PROPERTY?*

Guglielmo Maria Caporale**
Centre for Monetary and Financial Economics
South Bank University London
g.m.caporale@sbu.ac.uk

Nikitas Pittis
Department of Banking and Financial Management
University of Piraeus

ABSTRACT

This paper argues that persistence is not an invariant feature of a time series, but depends on the context in which the series is used: as the parameters of any dynamic model are defined relative to a particular information set, any change in the set of conditioning variables might affect the resulting estimates. We define persistence of a variable as the rate at which its autocorrelation function decays to zero, and show that inference about persistence of a variable is invariant to the addition of other conditioning variables only if those variables do not Granger-cause the variable of interest. Furthermore, we establish that measured persistence is a function of the model selected in a more fundamental way in the case of unstable systems. These findings suggest that, unless more restrictions derived from economic theory are imposed, issues such as the effectiveness of stabilisation policies cannot be settled empirically, and the debate between Keynesian and RBC theorists will remain inconclusive.

Keywords: Persistence, autocorrelation function, conditioning information set, probabilistic structure, dynamic models, Granger causality, cointegration.
JEL classification: C22, C32, E32.

---

* We are grateful to Stephen Hall, Aris Spanos and Thanasis Stengos for useful comments and suggestions. The usual disclaimer applies.

** Corresponding author, Professor Guglielmo Maria Caporale, South Bank University, Business School, 103 Borough Road, London SE1 OAA, United Kingdom, Tel: +44 (0) 20 7815 7012, Fax: +44 (0) 20 7815 8226, E-mail: g.m.caporale@sbu.ac.uk

I. INTRODUCTION

Persistence, i.e. the extent to which events today have an effect on the whole future history of a stochastic process, is a central issue in macroeconomic theory and policy. For example, in their seminal paper Nelson and Plosser (1982) argued that the presence of unit roots meant that shocks were persistent, and hence that the data were consistent with Real Business Cycle (RBC) models, in which most shocks to GNP were technology shocks. Campbell and Mankiw (1987a, 1987b), on the other hand, suggested that an ARMA(2,2) model provided the best description of the data for US real GDP, and hence that this is generated by a difference-stationary (DS) (or unit root) process. They also concluded that the long-run response of US GDP to a unit shock, given by the cumulative response function $A(1)$, is greater than 1, which implies that there is no trend-reversion.[1]

De Long and Summers (1988) claimed that stabilisation policies were more effective in the post-war period, when a larger fraction of the variance of US GNP could be explained by a stochastic trend. Numerous studies attributed the high degree of persistence exhibited by GDP to supply factors (see e.g. King et. al. (1991), Shapiro and Watson (1988), and Blanchard and Quah (1989)), although West (1988) showed that persistence is also consistent with Keynesian models of business cycles.[2] Similarly, a lot of effort was devoted to estimating the degree of persistence of unemployment and to pinning down its causes (see e.g., Blanchard and Summers (1986), and Alogoskoufis and Manning (1988)).

Various statistics have been proposed to capture the persistence of macroeconomic time series. Cochrane (1991a) argued that, because any time series with a unit root can be decomposed into a stationary series and a random walk, and the latter can have arbitrarily small variance, persistence should be measured as the ratio of the variance of the change in the random walk component to the variance of the actual change (see Cochrane (1988)). Furthermore, unit root tests are not informative about persistence (see Cochrane (1991b)). Firstly, the argument that series which are more likely to reject unit root tests are also those with "less persistent" shocks has no theoretical justification. Secondly, the persistence of univariate prediction error shocks can be very different from that of multivariate prediction error shocks, and also of the "true" underlying shocks.[3]

---

1 Christiano and Eichenbaum (1989), however, pointed out that this inference was very sensitive to the choice of ARMA specification from a set of models which had equally good fit.

2 For a general discussion of the econometric and policiy issues involved, see Campbell and Perron (1991), and McCallum (1993).

3 Lippi and Reichlin (1992) pointed out that another measure of persistence often used in empirical studies (see e.g. Clark (1987), or Watson (1986)), which is based on standard unobserved components models (UCARIMA) developed in Beveridge and Nelson (1981), is necessarily less than one as a mathematical consequence of the structure of these models.

In a different context, Cavaglia (1992) demonstrated how a measure of persistence may be obtained through the use of Kalman filtering. Finally, an alternative test statistic, known as rescaled range statistic (R/S), was first introduced by Hurst (1951) and then refined by Mandelbrot (1972, 1975), and Lo (1991), whose modified rescaled range statistic converges to a well-defined random variable under the null hypothesis of short-term dependence, and can distinguish between short-run and long-run dependence.

All the studies considered so far derived measures of persistence in the context of univariate models, i.e. using a particular conditioning information set. The question arises, however, whether estimates of persistence are invariant to model selection: would inference stay the same if a multivariate framework was adopted? Cochrane and Sbordone (1988), for example, provided a measure of persistence for GNP and stock prices which makes use of multivariate information. However, their statistic relies on strong non-testable identifying restrictions. Lupi (1993) also suggested that measures of persistence are not invariant to the information set, and that in a general probabilistic framework they are inadequate to capture persistence in terms of non-mixing properties.

Evans and Reichlin (1994) go one step further: they show that a widely used measure of persistence, i.e. the Beveridge-Nelson (BN) decomposition into trend and cycle, is non-increasing in the number of conditioning variables, and it is strictly decreasing if the additional conditioning variables Granger-cause the variable of interest, say y_t - a larger information set implies that more of y_t is forecastable and ascribed to the cyclical component, therefore resulting in a lower measure of persistence. Cochrane (1994) illustrates the empirical importance of this insight: the addition of Granger causal variables dramatically alters measures of transitory components in US GNP and stock prices. The same result might account for the fact that estimates of the cyclical component in aggregate output derived from multivariate systems using sectoral data exhibit larger variance than in univariate models of output (see e.g. Lee, Pesaran and Pierse (1992)).

This paper also examines the issue of whether persistence of a macroeconomic time series is a model invariant property. However, it differs from the existing literature in three respects. Firstly, the argument underlying our analysis is that any dynamic model can be interpreted as a statistical parameterisation of the probabilistic structure of the variable of interest, based on an implicit conditioning information set (see Spanos (1995b)). Therefore changes in the set of conditioning variables are likely to affect statistical inference.[4] Secondly, we define persistence as the memory of a process, with the latter being the rate at which the autocorrelation function of y_t decays to zero - this is more common in the theoretical literature on

[4] For instance, Spanos (1990) showed that it is not generally the case that the unit root found in the autoregressive (AR) representation of the series will persist in the context of a vector autoregressive (VAR) representation, and that the invariance conditions amount to Granger noncausality restrictions.

time series (see Priestley (1981)). Thirdly, we show that whether persistence is a model invariant property depends not only on causality restrictions, but also on the stability of the system - measured persistence of a variable y_t is a function of the model selected in a more fundamental way in the case of unstable systems.

The layout of the paper is the following. Section II defines formally the concept of persistence, and addresses the question whether model selection affects the estimated degree of persistence of a given variable under the assumption that the system is dynamically stable. Section III analyses the implications of dynamic instability of a system for the invariance conditions of measured persistence. Section IV illustrates the theoretical results using two empirical examples. Section V draws some conclusions.

II. MEASURING PERSISTENCE IN STABLE SYSTEMS

Let us consider the process $\{y_t, t \in T\}$, and let us also assume that it exhibits normality, Markovness and stationarity. These assumptions about its probabilistic structure imply that $\{y_t, t \in T\}$ can be described by the following formulation:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} \sim N\left\{\begin{bmatrix} \mu_y \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_y(0) & \sigma_y(1) \\ \sigma_y(1) & \sigma_y(0) \end{bmatrix}\right\} \tag{1}$$

Assuming that the relevant conditioning information set is given by $\sigma(y_{t-1})$, then the regression function $E(y_t / \sigma(y_{t-1}))$ gives rise to the usual homoscedastic AR(1) representation of y_t:

$$y_t = a_0 + ay_{t-1} + u_{1t} \tag{2}$$

Where u_{1t} is a white noise process. The parameters a y a_0 are related to the moments of the joint distribution $f(y_t, y_{t-1}; \vartheta)$ through the following relations:

$$a = \frac{\sigma_y(1)}{\sigma_y(0)} < 1 \tag{3}$$

$$a_0 = \mu_y(1 - a) \tag{3a}$$

Let us now define *persistence* as the effect of a 1 percent innovation on the long-run level of a series, say y_t. This definition implies that persistence can be seen as the *memory* of y_t, which is the rate at which the autocovariance (or, alternatively, the autocorrelation) function decays to zero (see Priestley (1981)). Intuitively, this is because the memory of a series indicates the rate at which a dynamic system returns to its initial state, which could be any steady state, after being perturbed by a shock.[5] If the memory of the process dies out as time passes by, then persistence is small and eventually becomes zero, whereas in the case of constant memory, i.e. when the system never returns to its initial state after a shock, persistence is constant. In other words, the faster the rate at which the autocorrelation function vanishes, the smaller the degree of persistence.

The assumptions of Markovness and stationarity made about the probabilistic structure of the process under consideration together imply that the autocovariance function of (2) decays at an exponential rate given by:

$$Cov(y_t, y_{t+\tau}) = c_y(\tau) = \sigma_y(0)a^{|\tau|} \tag{4}$$

or, in terms of the autocorrelation function:

$$Corr(y_t, y_{t+\tau}) = \rho_y(\tau) = a^{|\tau|} \tag{4a}$$

with $|a| < 1$

Clearly, the smaller the absolute value of *a* is, the faster is the rate at which the autocorrelation function approaches zero. This leads to the selection of *a* as a natural measure of persistence for the system (2). It should also be noted that in this case the autocorrelations decay to zero so fast that they are summable:

$$\sum_{\tau=0}^{\infty} \rho_y(\tau) < \infty \tag{4b}$$

It can be shown that (4b) is a sufficient condition for ergodicity (see Priestley (1981)).

It must be noted that $\{y_t\}$ is by construction a second-order stationary process, as indicated by the covariance structure given by (1). If, on the other hand, $\{y_t\}$ were a process which satisfies the difference equation (2), with u_t being a white noise process, $|a| < 1$ and $y_0 = 0$, then $\{y_t\}$ would be second-order stationary, but only asymptotically. This means that it would not attain stationarity until it had "forgotten" its initial starting value (see Priestley (1981)).

Next, let us consider the case where $\{y_t, t \in T\}$ is correlated with another process $\{x_t, t \in T\}$, which implies that we should consider the vector stochastic process $\{Z_t, t \in T\}$, where $Z_t = (y_t, x_t)'$, instead of the two independent processes $\{y_t\}$

[5] In the literature, persistence is normally defined as the value towards which the impulse response function converges in the case of I(1) variables, and as the area under the impulse response function (which would be infinity if the variable was non-stationary) for an I(0) series. In both cases, this corresponds to the sum of the coefficients of the moving average representation of the process. Our definition is basically equivalent.

and $\{x_t\}$. Let us impose the same probabilistic structure on $\{Z_t, t \in T\}$:

$$\begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma(0) & \Sigma(1) \\ \Sigma(1)' & \Sigma(0) \end{bmatrix} \right\} \tag{5}$$

where

$$\mu = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \Sigma(0) = \begin{bmatrix} \sigma_y(0) & \sigma_{yx}(0) \\ \sigma_{xy}(0) & \sigma_x(0) \end{bmatrix}, \Sigma(1) = \begin{bmatrix} \sigma_y(1) & \sigma_{yx}(1) \\ \sigma_{xy}(1) & \sigma_x(1) \end{bmatrix} \tag{5a}$$

where in general $\Sigma(1) \neq \Sigma(1)'$, since $\sigma_{yx}(1) \neq \sigma_{xy}(1)$. In the stationary environment being considered, $\Sigma(0)$ is a positive definite matrix.

The individual elements of $\Sigma(0)$ and $\Sigma(1)$ are:

$$\sigma_y(0) = Var(y_t)$$
$$\sigma_{yx}(0) = Cov(y_t, x_t)$$
$$\sigma_{xy}(0) = Cov(x_t, y_t)$$
$$\sigma_x(0) = Var(x_t)$$
$$\sigma_y(1) = Cov(y_t, y_{t-1})$$
$$\sigma_{yx}(1) = Cov(y_t, x_{t-1})$$
$$\sigma_{xy}(1) = Cov(x_t, y_{t-1})$$
$$\sigma_x(1) = Cov(x_t, x_{t-1})$$

It is important at this stage to note that, depending on the conditioning one chooses, one may have different dynamic models, namely VAR, or DLR (Dynamic Linear Regression) models. If the conditioning information set is chosen to be $\sigma(Z_{t-1})$, the relevant regression function becomes:

$$E(Z_t / (Z_{t-1})) = \boldsymbol{a_0} + AZ_{t-1} \tag{6}$$

which in turn gives rise to the usual VAR model:

$$Z_t = a_0 + AZ_{t-1} + U_t \tag{7}$$

$$\text{where } a_0 = [I - A]\mu = \begin{bmatrix} a_y \\ a_x \end{bmatrix} \tag{7a}$$

$$\text{and } A = \Sigma(1)\Sigma(0)^{-1} \tag{7b}$$

where $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ (7c)

and U_t is a two-dimensional white noise process.

The elements of A are related to the moments of the joint distribution $f(Z_t, Z_{t-1}; \vartheta)$ through the following relations:

$$a_{11} = \frac{\sigma_y(1)\sigma_x(0) - \sigma_{yx}(1)\sigma_{xy}(0)}{|\Sigma(0)|} \tag{8a}$$

$$a_{12} = \frac{\sigma_{yx}(1)\sigma_y(0) - \sigma_y(1)\sigma_{yx}(0)}{|\Sigma(0)|} \tag{8b}$$

$$a_{21} = \frac{\sigma_{xy}(1)\sigma_y(0) - \sigma_y(1)\sigma_{yx}(0)}{|\Sigma(0)|} \tag{8c}$$

$$a_{22} = \frac{\sigma_x(1)\sigma_y(0) - \sigma_{xy}(1)\sigma_{xy}(0)}{|\Sigma(0)|} \tag{8d}$$

or more conveniently:

$$a_{11} = \frac{\sigma_y(1) - a_{12}\sigma_{xy}(0)}{\sigma_y(0)} \tag{8e}$$

$$a_{12} = \frac{\sigma_{yx}(1) - a_{11}\sigma_{yx}(0)}{\sigma_x(0)} \tag{8f}$$

$$a_{21} = \frac{\sigma_{xy}(1) - a_{22}\sigma_{xy}(0)}{\sigma_y(0)} \tag{8g}$$

$$a_{22} = \frac{\sigma_x(1) - a_{21}\sigma_{yx}(0)}{\sigma_x(0)} \tag{8h}$$

As for the conditional covariance matrix, we have:

$$Cov(Z_t / Z_{t-1}) = \Sigma(0) - A\Sigma(1)' = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \tag{8j}$$

where $\omega_{11} = Var(y_t / Z_{t-1})$, $\omega_{12} = Cov(y_t, x_t / Z_{t-1})$, $\omega_{21} = Cov(x_t, y_t / Z_{t-1})$, and $\omega_{22} = Var(x_t / Z_{t-1})$.

On the other hand, if the relevant information set is chosen to be $I = \{(y_{t-1}), X_t = x_t, X_{t-1} = x_{t-1}\}$, then the relevant regression function is given by

$$E(y_t / \sigma(y_{t-1}), X_t = x_t, X_{t-1} = x_{t-1}) = c_1 + a_1 y_{t-1} + b_1 x_t + b_2 x_{t-1} \tag{9}$$

which gives rise to the Dynamic Linear Regression model for y_t:

$$y_t = c_1 + a_1 y_{t-1} + b_1 x_t + b_2 x_{t-1} + e_{1t} \tag{10}$$

with the statistical parameters of interest $b = [b_1, a_1, b_2]'$ being related to the moments of the joint distribution $f(Z_t, Z_{t-1}; \vartheta)$ through the following relationships:

$$b = \sigma'_{yx} \Sigma_{22}^{-1}$$

where

$$\sigma'_{yx} = [\sigma_{yx}(0), \sigma_y(1), \sigma_{yx}(1)]$$

and

$$\Sigma_{22} = \begin{bmatrix} \sigma_x(0) & \sigma_{xy}(1) & \sigma_x(1) \\ \sigma_{xy}(1) & \sigma_y(0) & \sigma_{yx}(0) \\ \sigma_x(1) & \sigma_{xy}(0) & \sigma_x(0) \end{bmatrix}$$

More conveniently, one can express the parameters of interest in the DLR as functions of parameters of interest in the VAR in the following way:

$$a_1 = a_{11} - b_1 a_{21} \tag{11a}$$

$$b_1 = [\omega_{12} / \omega_{22}] \tag{11b}$$

$$b_2 = a_{12} - b_1 a_{22} \tag{11c}$$

where a_{11}, a_{21}, a_{12}, a_{22}, ω_{12} and are defined as previously.

Again the probabilistic structure of the system results in an exponentially decreasing autocovariance function for y_t similar to (4). This time, though, persistence of the system is measured by a_1 as defined in (11). In general $a \neq a_1$, and therefore measured persistence of y_t depends on the selected dynamic system.

It must be noted that persistence in (10) is also determined by the rate at which the autocovariance or the autocorrelation function vanishes. However, the autocovariance function of (10) is defined in terms of the joint distribution of y_t, y_{t+k} conditional on $X_t = x_t, X_{t-1} = x_{t-1}$. More specifically, the autocovariance function of (2) is defined as:

$$Cov(y_t, y_{t+\tau}) = E\{(y_t - E(y_t))(y_{t+\tau} - E(y_{t+\tau}))\} =$$
$$\iint (y_t - E(y_t))(y_{t+\tau} - E(y_{t+\tau})) f(y_t, y_{t+\tau}) dy_t dy_{t+\tau} \tag{12a}$$

whereas the conditional autocovariance function of (10) is defined as:

$$Cov(y_t, y_{t+\tau} / x_t, x_{t-1}, x_{t+\tau}, x_{t+\tau-1}) = Cov(y_t, y_{t+\tau} / x_t, x_{t-1})$$
$$E\{(y_t - E(y_t / x_t, x_{t-1}))(y_{t+\tau} - E(y_{t+\tau} / x_t, x_{t-1}))\} =$$
$$\iint (y_t - E(y_t / x_t, x_{t-1}))(y_{t+\tau} - E(y_{t+\tau} / x_t, x_{t-1})) f(y_t, y_{t+\tau} / x_t, x_{t-1}) dy_t dy_{t+\tau} \tag{12b}$$

This in turn implies that the rate at which the autocovariance function of (10) vanishes, as measured by a_1, will now be a function of the moments of the joint distribution $f(Z_t, Z_{t-1})$, whereas is a function of the moments of $f(y_t, y_{t-1})$.

Let us now derive conditions under which the two measures of persistence will be the same:

Proposition 1: Persistence of y_t will be invariant to models (2) and (10), if and only if there is no Granger causality between y_t and x_t in any direction. [6]

Proof

(If part) Assume that $a = a_{11}$. This implies, given (8a), that $a_{12} = 0$ ($\sigma_{xy}(0) \neq 0$ in general). Next, if $a_{11} = a_1$, then it follows from (11a) that $a_{21} = 0$, since b_1 in general is different from zero. Therefore, if $a = a_{11} = a_1$, then $a_{12} = a_{21} = 0$.

(Only if part). Assume that $a_{12} = 0$; then (8a) implies that $a = a_{11}$. Also if $a_{21} = 0$, then $a_1 = a_{11}$. Therefore, if $a_{12} = a_{21} = 0$, then $a = a_{11} = a_1$.

Alternatively we can express the noncausality restrictions $a_{12} = 0$ and $a_{21} = 0$ in terms of the moments of the joint distribution $f(Z_t, Z_{t-1})$ as follows:

$$a_{12} = 0 \Leftrightarrow \sigma_{yx}(1)\sigma_y(0) = \sigma_y(1)\sigma_{yx}(0) \tag{GNC 1}$$

$$a_{21} = 0 \Leftrightarrow \sigma_{xy}(1)\sigma_x(0) = \sigma_x(1)\sigma_{xy}(0) \tag{GNC 2}$$

---

6 Note that in the context of a VAR(1) system (and hence of its reparameterisation as a DLR model), Granger causality is the same as *long-run causality*. In the more general case with a lag polynomial of order *n*, though, Granger causality cannot distinguish between short- and long-run causality: a variable *y* may have transitory effects on another variable *x*, but the long-run behaviour of *x* can still be invariant to the behaviour of *y* (see Hall and Wickens (1993) for more details).

These restrictions, together with the relationships given by (10a), imply the following parameters for model (10):

$$a_1 = \frac{\sigma_y(1)}{\sigma_y(0)}$$

$$b_1 = \frac{\sigma_x(0)\sigma_{xy}(1)[\sigma_x(1)\sigma_y(1) - \sigma_x(0)\sigma_y(0)]}{\sigma^3{}_x(1)\sigma_y(0) - \sigma^2{}_x(0)\sigma_x(1)\sigma_y(0)}$$

$$b_2 = \frac{\sigma_{xy}(1)[\sigma_x(1)\sigma_y(1) - \sigma_x(0)\sigma_y(0)]}{\sigma_y(0)[\sigma^2{}_x(0) - \sigma_x^2(1)]}$$

where clearly $a_1 = a$. It is important to note that $a_1 = a$ if and only if $a_{12} = a_{21} = 0$. For example, if one assumes absence of causality in only one direction, e.g. $a_{12} = 0$, then the parameters of model (10) will be totally different from those in (7). In particular a_1 will be equal to:

$$a_1 = \frac{A}{B}$$

where

$$\begin{aligned} A = {} & \sigma_x(0)\sigma_{xy}(1)\sigma_x(1)\sigma_y(0)\sigma_{yx}(0) - \sigma_x^2(1)\sigma_y(0)\sigma_{yx}^2(0) - \sigma_x(0)\sigma^2{}_{xy}(1)\sigma_x(1)\sigma_y(1) \\ & - \sigma_x^2(0)\sigma_x(1)\sigma_y(0)\sigma_y(1) + \sigma_x^2(1)\sigma_y(0)\sigma_y(1) + \sigma^2{}_x(0)\sigma_{xy}(1)\sigma_{yx}(0)\sigma_y(1) \end{aligned}$$

$$\begin{aligned} B = {} & \sigma_x(0)\sigma_y(0)[\sigma_x(0)\sigma_{xy}^2(1) - \sigma_x^2(0)\sigma_y(0) + \sigma_x^2(1)\sigma_y(0) - 2\sigma_{xy}(1)\sigma_x(1)\sigma_{yx}(0) \\ & + \sigma_x(0)\sigma_{yx}^2(0)] \end{aligned}$$

The following points are noteworthy:

1. The notion of non-causality which is relevant here is that of Granger non-causality, according to which x does not cause y if past values of x have no usefulness in predicting y. More formally, $f(y_t / Z_{t-1}) = f(y_t / y_{t-1})$, which under normality is equivalent to saying $a_{12} = 0$.

2. Granger non-causality should be distinguished from the case in which instantaneous non-causality is also assumed. The latter case implies $f(y_t / x_t, Z_{t-1}) = f(y_t / y_{t-1})$, and this condition does not hold if only $a_{12} = 0$ is assumed. This alternative notion of non-causality should be expressed in terms of the parameters of $f(y_t / x_t, Z_{t-1})$ as: $b_1 = b_2 = 0$. If this condition holds, then, from (11), $a_{12} = 0$, but the opposite is not true in general.

3. The preceding discussion makes clear that $a_{12} = 0$ can be compatible with $b_2 \neq 0$. In fact b_2 can be regarded as the difference between two components. The first is the direct effect of x_{t-1} on y_t when the contemporaneous dependence between y_t and x_t, which is measured by a_{12}, is left unmodelled, and the second is the indirect effect of x_{t-1} on y_t through the contemporaneous dependence between and y_t, and x_t intertemporal dependence between x_t and x_{t-1}, which is measured by $b_1 a_{22}$.

GENERAL RESULT FOR STABLE SYSTEMS: If a variable y_t, which is in a steady state, is hit by a shock, its persistence in alternative stable linear systems such as (2) or (10) decreases at an exponential rate (system invariant feature of persistence). However, the exponential rate, in general, is not the same in the two systems (system dependent feature of persistence). Persistence will be model invariant if and only if there is no Granger causality in any direction among the variables involved. Needless to say, this result refers to the effect of the shock for a finite period of time (τ). If $\tau \Rightarrow \infty$ persistence in both systems will be the same and equal to zero.

It must be noted that so far $\{y_t, t \in T\}$, and $\{Z_t, t \in T\}$ have been assumed to be stationary around the fixed means μ_y and μ respectively. Alternatively, one could assume stationary around some deterministic trend polynomials, without changing the covariance structure. For instance, one could assume that:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} \sim N\left\{\begin{bmatrix} m(t) \\ m(t-1) \end{bmatrix}, \begin{bmatrix} \sigma_y(0) & \sigma_y(1) \\ \sigma_y(1) & \sigma_y(0) \end{bmatrix}\right\} \tag{12}$$

The only changes in the regression model would be the addition of a deterministic trend polynomial of the same order, whereas the relationship of the autoregressive coefficient with the moments of the distribution would remain unaltered.

III. PERSISTENCE IN UNSTABLE SYSTEMS

Our analysis so far has only been concerned with stable systems, i.e. the case when $a_1 < 1$ in (2), both eigenvalues in (7) are less than one, and $a_1 < 1$ in (10). Let us now also consider the unit roots case, and hence let us assume that the process $\{y_t, t \in T\}$ is a general Wiener process:

$$\begin{bmatrix} y_t \\ y_s \end{bmatrix} \sim N\begin{bmatrix} m_y(t) \\ m_y(s) \end{bmatrix}, \begin{bmatrix} \sigma_y(0)t & \sigma_y(|t-s|)min(t,s) \\ \sigma_y(|t-s|)min(t,s) & \sigma_y(0)s \end{bmatrix} \tag{13}$$

where $m(t)$ is a trend polynomial of order n, and $\sigma(|t-s|)$ is a constant, which is a function of the distance between the two points in time. By making the further assumption of Markovness, we obtain:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} \sim N\left\{\begin{bmatrix} m_y(t) \\ m_y(t-1) \end{bmatrix}, \begin{bmatrix} \sigma_y(0)t & \sigma_y(1)(t-1) \\ \sigma_y(1)(t-1) & \sigma_y(0)(t-1) \end{bmatrix}\right\} \tag{14}$$

Note that a general Wiener-Markov process as defined in (14) does not have independent and stationary increments unless i) $m_y(t)$ is a linear function of t and ii) $\sigma_y(\tau)=\sigma_y(0)$ for every τ. The proof of the first part of this statement is straightforward, since for any polynomial $m_y(t)$ of order n, $E[y_t]$ will be a trend polynomial of order n-1. The second part can be proved by examining the autocovariance function of y_t:

a)
$$\begin{aligned} Var[\Delta y_t] &= Var(y_t)+Var(y_{t-1})-2Cov(y_t,y_{t-1}) \\ &= 2\sigma_y(0)t-2\sigma_y(1)-\sigma_y(0)+2\sigma_y(1) \end{aligned} \tag{14a}$$

Clearly $Var(y_t)$ is a function of t, unless $\sigma_y(0)=\sigma_y(1)$, in which case the variance of the increments becomes $\sigma_y(0)$.

b)
$$\begin{aligned} Cov[\Delta y_t, \Delta y_{t-\tau}] &= 2\sigma(\tau+1)t-\sigma(\tau+2)t-\sigma(\tau)t-2\sigma(\tau+1)\tau \\ &\quad +\sigma(\tau+2)\tau\sigma(\tau)\tau+\sigma(\tau+1)+\sigma(\tau)-2(\tau+1) \end{aligned} \tag{14b}$$

It is apparent that $Cov[\Delta y_t, \Delta y_{t-\tau}]$ is a function of both t and τ unless $\sigma(\tau)=\sigma(0)$ for every τ, in which case the covariance of the increments becomes zero.

The regression and skedastic functions then become:

$$E(y_t / y_{t-1}) = m_y(t) - \frac{\sigma_y(1)}{\sigma_y(0)} m_y(t-1) + \frac{\sigma_y(1)}{\sigma_y(0)} y_{t-1} \tag{15}$$

$$Var(y_t / y_{t-1}) = \sigma_y(0)t - \frac{[\sigma_y(1)(t-1)]^2}{\sigma_y(0)(t-1)} \tag{15b}$$

or

$$Var(y_t / y_{t-1}) = \frac{\sigma_y(1)}{\sigma_y(0)} + \frac{[\sigma_y^2(0)-\sigma_y^2(1)]}{\sigma_y(0)} = a + a_2 t \tag{15c}$$

Note that in the general Wiener case the autoregressive coefficient is still time-independent; the conditional variance, however, is a function of t.

If the polynomial $m(t)$ is linear in t, and under the unit root restriction $\sigma_y(1)=\sigma_y(0)$, the above regression function gives rise to the usual random walk with drift model:

$$y_t = \mu_y + y_{t-1} \tag{16}$$

Whereas if *m(t)* is a quadratic function of *t*, one obtains the random walk model with a drift and linear trend:

$$y_t = \mu_y + \mu_{y1}t + y_{t-1} \tag{17}$$

with the conditional variance in both cases being equal to:

$$Var(y_t / y_{t-1}) = \sigma(0)$$

which is free of *t*.

The covariance function of the unstable model (16) now decreases at a *linear-like (not exponential)* rate. This can be shown as follows: In the case of a Wiener-Markov process with the unit root restrictions, the covariance function $Cov(y_t, y_{t+\tau})$ is not a function of the distance $|\tau|$, but rather of the actual date *t*. Assuming that $\tau > 0$, we have:

$$c(\tau) = Cov(y_t, y_{t+\tau}) = \sigma_y(0)min(t, t+\tau) = \sigma_y(0)t \tag{18}$$

or

$$\rho(\tau) = Corr(y_t, y_{t+\tau}) = \frac{\sigma_y(0)t}{[\sigma_y(0)t\sigma_y(0)(t+\tau)]^{1/2}} = \left(\frac{t}{t+\tau}\right)^{1/2} \tag{18b}$$

Equation (18) represent the case known in the literature as "constant memory". Indeed, regardless of the distance between y_t and $y_{t+\tau}$, their autocovariance remains constant. On the other hand, the autocorrelation function given by (18b) suggests that the correlation between y_t and $y_{t+\tau}$ decreases as τ increases at a linear-like rate. This results from standardising the covariance between y_t and $y_{t+\tau}$ using the square root of the product of the corresponding variances. In such a case, the autocorrelation function appears to vanish as $\tau \Rightarrow \infty$, implying *pseudo* asymptotic independence.

The difference from the stable case in terms of the autocorrelation function is that the autocorrelations decrease at a linear-like and not exponential rate. This results in correlations decaying to zero so slowly that they are not summable:

$$\sum_{\tau=-\infty}^{\infty} \rho(\tau) = \infty \tag{18c}$$

In terms of the autocovariance function, stable systems exhibit an exponentially decreasing memory, whereas unstable systems are characterised by constant memory. Persistence in unstable systems, therefore, remains constant even for $\tau \Rightarrow \infty$.

Let us also assume that $\{x_t, t \in T\}$ has a unit root in its autoregressive representation, and in particular that it is a random walk with drift:

$$x_t = \mu_x + x_{t-1} \tag{19}$$

Where, as previously, the unit root restriction takes the form: $\sigma_x(1) = \sigma_x(0)$. If we now assume that y_t and x_t are correlated, then we should consider the vector stochastic process $Z_t = (y_t, x_t)'$, which is assumed to have the same probabilistic structure as before (i.e. it is a Wiener, Markov vector stochastic process):

$$\begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} \sim N \left\{ \begin{bmatrix} m(t) \\ m(t-1) \end{bmatrix} \begin{bmatrix} \sum(0)t & \sum(1)(t-1) \\ \sum(1)'(t-1) & \sum(0)(t-1) \end{bmatrix} \right\} \tag{20}$$

In a non-stationary environment, ***Σ****(0)* will not always be positive definite - it might also be positive semi-definite, in which case ***Σ****(0)*$^{-1}$ does not exist. (This case is analysed in subsection A). For the time being we assume that ***Σ****(0)* is invertible, i.e. ***Σ****(0)*$^{-1}$ exist, and therefore the conditional mean and conditional covariance, of Z_t are given by:

$$\begin{aligned} E(Z_t / Z_{t-1}) &= m(t) - \sum(1)\sum(0)^{-1} m(t-1) + \sum(1)\sum(0)^{-1} Z_{t-1} \\ &= m(t) - Am(t-1) + AZ_{t-1} \end{aligned} \tag{20a}$$

$$\begin{aligned} Cov(Z_t / Z_{t-1}) &= \sum(0)t - \sum(1)\sum(0)^{-1}\sum(1)'(t-1) \\ &= \sum(0)t - A\sum(1)'(t-1) \end{aligned} \tag{20b}$$

where ***m****(t)* is a trend polynomial of known order, and ***Σ****(0)* and ***Σ****(1)* are defined as in (5a). Having chosen $\sigma(Z_{t-1})$ to be the relevant conditioning information set, we get, as in the case of stable systems, a VAR model:

$$Z_t = a_0(t) + AZ_{t-1} + V_t \tag{21}$$

where now the vector of constants is a function of time, and the conditional covariance matrix $Cov(Z_t / Z_{t-1}) = \Sigma(0)t - A\Sigma(1)'(t-1) = \Omega(t)$ is also a function of time. More specifically:

$$a_0(t) = m(t) - Am(t-1) \tag{22}$$

and the individual elements of $\Omega(t)$ are the following:

$$\omega_{11}(t)=\sigma_{y}(0)t-\left[a_{11}\sigma_{y}(1)+a_{12}\sigma_{yx}(1)\right](t-1) \tag{23a}$$

$$\omega_{12}(t)=\sigma_{yx}(0)t-\left[a_{11}\sigma_{yx}(1)+a_{12}\sigma_{x}(1)\right](t-1) \tag{23b}$$

$$\omega_{21}(t)=\sigma_{xy}(0)t-\left[a_{21}\sigma_{y}(1)+a_{22}\sigma_{yx}(1)\right](t-1) \tag{23c}$$

$$\omega_{22}(t)=\sigma_{x}(0)t-\left[a_{21}\sigma_{yx}(1)+a_{22}\sigma_{x}(1)\right](t-1) \tag{23d}$$

On the contrary, the matrix A is time independent, being given by:

$$A=\sum(1)\sum(0)^{-1} \tag{24}$$

and therefore A is defined, as in the case of stable systems, by (7c). In this general case the VAR is not operational since the elements of the conditional covariance matrix are functions of time.

Alternatively, if we choose $D=\{\sigma(y_{t-1}), X_t=x_t, X_{t-1}=x_{t-1}\}$ as the relevant conditioning information set, we obtain again a reparameterisation of the VAR which is the usual **DLR** model:

$$y_t = c_1(t) + a_1(t)y_{t-1} + b_1(t)x_t + b_2(t)x_{t-1} + v_{1t} \tag{25}$$

where:

$$a_1(t) = a_{11} - b_1(t)a_{21} \tag{26a}$$

$$b_1(t) = [\omega_{12}(t) / \omega_{22}(t)] \tag{26b}$$

$$b_1(t) = a_{12} - b_1(t)a_{22} \tag{26c}$$

and

$$Var(y_t/Z_{t-1}, X_t = x_t) = s_{11}(t) - [s_{12}(t)\, s_{21}(t)/s_{22}(t)] \tag{26d}$$

and a_{11}, a_{21}, a_{22}, $\omega_{ij}(t)$ are defined as before. In this general case DLR is not an operational model since its parameters are functions of time.

Under what condition does (25) become operational? It can be seen that the source of parameter-time dependence in the DLR is the time dependence of $Cov(Z_t/Z_{t-1})$. Therefore, the conditions under which the DLR becomes operational must be identical to the conditions that make $Cov(Z_t/Z_{t-1})$ independent of t.

So far we have not restricted the time-homogeneity structure of $\{Z_t\}$ by imposing the unit root restrictions for $\{y_t\}$ and $\{x_t\}$ which, as we said, take the form: $\sigma_y(1)=\sigma_y(0)$. and $\sigma_x(1)=\sigma_x(0)$.. By imposing these restrictions we obtain:

$$\sum(1)=\begin{bmatrix}\sigma_y(0) & \sigma_{yx}(1)\\ \sigma_{xy}(1) & \sigma_x(0)\end{bmatrix}$$

Under these restrictions, however, $Cov\ (Z_{tt}/Z_{t-1})$ is still time dependent. This in turn implies that, if y_t and x_t are $I(1)$, then both VAR and DLR models are in general, non-operational.

Let us look at $Cov\ (Z_t/Z_{t-1})$. more closely. From (20b) it is straightforward to see that $Cov\,(Z_t\,/\,Z_{t-1})=\left[\Sigma(0)-A\,\Sigma(1)'\right]t+A\,\Sigma(1)'$. This enables us to argue that necessary and sufficient conditions for $Cov\ (Z_{tt}/Z_{t-1})$ to be time-independent take the form:

$$\left[\sum(0)-A\sum(1)'\right]=0 \tag{26b}$$

Proposition 2: Necessary and sufficient conditions for (26b) to be true are:

$$\begin{aligned}&\sigma_y(1)=\sigma_y(0)\\ &\sigma_x(1)=\sigma_x(0)\\ &\sigma_{yx}(0)=\sigma_{yx}(1)=\sigma_{xy}(1)\end{aligned} \tag{26c}$$

which is equivalent to Granger non-causality between y_t and x_t, in the sense that $a_{21} = a_{12} = 0$ (see GNC1-2). The proof of this proposition is trivial.

It can also be shown that under (26c) and the univariate unit root conditions $\sigma_y(1)=\sigma_y(0)$ and $\sigma_x(1)=\sigma_x(0)$, the matrix A becomes equal to the identity matrix. This in turn implies the absence of Granger causality in any direction between y_t and x_t. Obviously, in such case both eigenvalues of A are equal to unity: $\lambda_1=\lambda_2=1$.

The DLR model has now become operational since

$$Cov\,(z_t\,/\,Z_{t-1})=A\sum(1)'=\begin{bmatrix}\omega_{11}' & \omega_{12}'\\ \omega_{21}' & \omega_{22}'\end{bmatrix}$$

with parameters:

$$a_1(t) = a_{11} = 1 \tag{26e}$$

$$b_1 = \left[\omega'_{12} / \omega'_{22}\right] = \left[\sigma_{yx}(0) / \sigma_x(0)\right] \tag{26f}$$

$$b_2 = -b_1 a_{22} = -b_1 \tag{26g}$$

and

$$Var\,(y_t / Z_{t-1}, X_t = x_t) = \left[\sigma_y(0) - \sigma^2_{yx}(0)\right] / \sigma_x(0) \tag{26h}$$

Equations (26e) to (26h) make it clear that even in the case of no Granger causality in any direction between y and x, the AR representation of y_t is not equivalent to its DLR representation, unless $\sigma_{yx}(0) = 0$, i.e. unless there is no contemporaneous dependence between y and x. Persistence, however, will be the same, since equation (26e) enables us to argue:

Proposition 3: persistence will be invariant to models (16) or (25) if and only if condition (26c) holds, or, in other words, if there is no Granger causality among the variables involved.

In the presence of Granger causality, however, and under the invertibility of $\boldsymbol{\Sigma}(0)$, persistence in the context of (25) is not definable since a_1 is a function of t. In this case, we can say that the persistence of the series is not constant over time, in the sense that the effect of a 1 percent change in the innovation at time $t = s_1$, is different from that of an increase in innovation at $t = s_2$.

A. Singular versus non singular $\boldsymbol{\Sigma}(0)$

So far the analysis has been carried out under the maintained hypothesis that $\boldsymbol{\Sigma}(0)$ is positive definite matrix, which ensures the existence of $\boldsymbol{\Sigma}(0)^{-1}$. Positive definiteness of $\boldsymbol{\Sigma}(0)$ is equivalent to saying that the correlation coefficient between y_t and x_t is different from one, i.e. $\rho_{yx} \neq 1$, or that $\left[\sigma_y(0)\sigma_x(0)\right]^{1/2} > \sigma_{yx}(0)$. This condition is ensured if and only if y_t and x_t are linearly independent in an exact sense, i.e. if there is no scalar $k \neq 0$ such that $Pr(y_t = kx_t) = 1$. In the stationary environment given by (5), this is always true. In the non-stationary case described by (20), this can also be true if y_t and x_t do not share a common stochastic trend (the case analysed so far). If, however, the stochastic behaviour of y_t and x_t is governed by a common trend, then $\rho_{yx} \to 1$ very fast. In such a case $\boldsymbol{\Sigma}(0)$, being positive semi-definite, will be singular. This is the case commonly referred to as **cointegration** between y_t and x_t, where the first two moments of the conditional distribution $f(Z_t/Z_{t-1})$ are not given by (20a) and (20b), but rather:

$$E(Z_t / Z_{t-1}) = m(t) - \sum(1)\sum(0)^{-1} m(t-1) + \sum(1)\sum(0)^{-1} Z_{t-1} \tag{27a}$$

$$Cov\,(Z_t / Z_{t-1}) = \sum(0)t - \sum(1)\sum(0)^{-1}\sum(1)'(t-1) \tag{27b}$$

where $\boldsymbol{\Sigma}(0)^{-1}$ is the generalized inverse of $\boldsymbol{\Sigma}(0)$, i.e. it is such that:.

$$\sum(0)\sum(0)^{-1}\sum(0) = \sum(0)$$

This case has been examined by Spanos (1989), who proved that under the unit root restrictions $\sigma_y(1) = \sigma_y(0)$, $\sigma_x(1) = \sigma_x(0)$, and singularity of $\boldsymbol{\Sigma}(0)$, the conditional covariance matrix becomes:

$$Cov\,(Z_t / Z_{t-1}) = \sum(0) \tag{27c}$$

which is free of t, implying that DLR model described by (25) becomes operational. The statistical parameters of interest in this model are related to those of the VAR, under cointegratig restrictions, through the following relationships:

$$a_1 = a_{11} - b_1 a_{21} = (a_{11} + a_{22}^{-1}) \tag{28a}$$

$$b_1 = \left[\sigma_y(0) / \sigma_x(0)\right]^{1/2} \tag{28b}$$

$$b_2 = b_1\,(1 - a_{11} - a_{22}) \tag{28c}$$

The cointegration case is very interesting because under cointegration the coefficient becomes less than unity, which results in a stable DLR model. This occurs because the cointegration restrictions, as described by $|\boldsymbol{\Sigma}| = 0$, do not allow a_{12} and a_{21} to be simultaneously zero, which ensures the existence of Granger causality in at least one direction. This in turn implies, through (28a) and (8a) and given fact that b_1 is now positive, that a_1 will be less than one.

Another way to show this is the following. First, consider the relation between the eigenvalues of A and the elements of A:

$$\lambda_1, \lambda_2 = \frac{a_{11} + a_{22} \pm \left[(a_{11} + a_{22})^2 - 4\,det(A)\right]^{1/2}}{2} \tag{29}$$

in the absence of cointegration $\lambda_1 = \lambda_2 = 1$, and therefore the condition for no cointegration takes the form:

$$(a_{11} + a_{22}) - (a_{11}a_{22} - a_{12}a_{21}) = 1 \tag{30}$$

if $a_{12} = a_{21} = 0$, then the form (8) it follows that $a_{11} = a_{22} = 1$, and hence (30) holds. This means that a sufficient condition for (30) to hold, i.e. for no cointegration, is Granger non-causality in any direction. Therefore, a necessary condition for cointegration amounts to Granger causality in at least one direction. If this is the case, i.e. if either $a_{12} \neq 0$ or $a_{21} \neq 0$ or both, then (8) implies that either $a_{11} < 0$ or $a_{22} < 0$ or both. This in turn implies that:

$$a_1 = a_{11} + a_{22}^{-1} < 1 \tag{31}$$

Here persistence of y_t depends on the model selected in **a more fundamental way than in stable systems**. That is, although persistence of y_t is constant in the context of model (16), it decreases at an exponential rate in the context of model (25) with cointegrating restrictions.

Before we present some empirical results, it is worth discussing briefly how often in practice one encounters a situation where Granger causality in at least one direction coincides with no cointegration (non-singular $\boldsymbol{\Sigma}(0)$), thus yielding non-operational models. Lets us assume that $a_{12} \neq 0$ and $a_{21} = 0$. This clearly invalidates condition (26c), and generally results in a non-operational VAR and DLR models. The matrix A then takes the form:

$$A = \begin{bmatrix} 1 - a_{12}\dfrac{\sigma_{xy}(0)}{\sigma_y(0)} & a_{12} \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad tr(A) = 2 - a_{12}\frac{\sigma_{xy}(0)}{\sigma_y(0)} = \lambda_1 + \lambda_2 = \lambda_1 + 1 \tag{32}$$

No-cointegration can be though of a situation in which both λ_1 and λ_2 are greater than or equal to one. If one of the eigenvalues is less than one, then the system is cointegrated. This implies that, under no cointegration, the sum of the eigenvalues must be greater or equal to 2, which means that $\lambda_1 \geq 1$, or equivalently, than $a_{11} > 1$. For this to be the case a_{12} has to be such that $a_{12}\,[\sigma_{yx}(0)/\sigma_y(0)] < 0$. On the other hand, if $a_{12}\,[\sigma_{yx}(0)/\sigma_y(0)] > 0$, then $a_{11} = \lambda_1 < 1$, and the system is cointegrated, which implies that $\boldsymbol{\Sigma}(0)$ is singular and both VAR and the DLR models become operational. However, cases where $a_{11} > 1$ are not very often encountered in practice, which means that evidence of causality usually indicates that there is cointegration.

The preceding analysis suggests that a test of the null hypothesis of no-cointegration can take the form $H_0 : a_1 = 1$ against the one-sided cointegration alternative $H_1 : a_1 < 1$. The t-test, however, does not follow standard asymptotics under the null. Critical values can be obtained by Monte-Carlo methods as in the univariate case (see Dickey and Fuller (1981)). However, such and undertaking is beyond the scope of the present paper – we use instead the DF critical values for the univariate case as a guide for the corresponding ones in the multivariate case.

Finally, some discussions on the restrictiveness of the above analytical framework is required. First, the Wiener-Markov framework is mainly motivated by the empirical literature on unit roots and cointegration where one unit root and normality are widely used. Even in the cases where normality in the errors is not explicitly assumed, the linear conditional mean and the homoscedastic conditional variance brings the whole framework very close to normality. This is due to the characterisation result provided by Nimmo-Smith (1979) and Spanos (1995a), according to which if both conditional means $E(y_t/x_t)$ and $E(x_t/y_t)$ are linear in y_t and x_t respectively, and only one conditional variance $Var(y_t/x_t)$ is homoscedastic, then the joint distribution $f(y_t, x_t)$ is normal.

Another restrictive feature of the above framework is its "bivariate" structure. This has been employed only for simplicity, and the results are also valid in a more general multivariate framework. To see this more clearly, one can assume that the vector Z_t in the stochastic process $\{Z_t, t \in T\}$ is equal to $Z_t = (y_t, X_t)$ where $X_t = (X_{1t}, X_{2t}, ..., X_{kt})$. Then $\mathbf{\Sigma}(0)$ and $\mathbf{\Sigma}(1)$ in (20) become:

$$\sum(0) = \begin{bmatrix} \sigma_y(0) & \sigma_{yx}(0) \\ \sigma_{xy}(0) & \sigma_x(0) \end{bmatrix} \qquad \sum(1) = \begin{bmatrix} \sigma_y(1) & \sigma_{yx}(1) \\ \sigma_{xy}(1) & \sigma_x(1) \end{bmatrix} \tag{32}$$

where $\sigma_{yx}(\tau)$ and $\mathbf{\Sigma}_x(\tau)$, $\tau = 0,1$ are a $(1xk)$ vector and (kxk) matrix respectively.

IV. TWO EMPIRICAL EXAMPLES

This section reconsiders the vexed question of persistence of GDP (see, e.g., Nelson and Plosser (1982) and Campbell and Mankiw (1987a, 1987b) in the light of theoretical results derived above. Let us assume that y_t denotes US GDP. A parametric measure of persistence can be derived by estimating a in (2) to obtain:[7]

$$\underset{(0.154)}{y_t = 0.150} + \underset{(0.023)}{0.980 y_{t-1}} \tag{33}$$

$LM(1)=0.35 \qquad LM(4)=0.27$
$Q(8)=0.09 \qquad BJ=0.43$

Where LM(i) is a Lagrange multiplier residual autocorrelation F-test of order i, Q(8) is a Ljung-Box portmanteau test of order eight, and BJ is a Bera-Jarque normality test (p-values are reported). The estimated model seems reasonably well specified statistically, since there is no evidence for higher order linear dynamics. On the basis of this model (i.e. random walk with a drift), one would infer that the persistence of output is constant.

[7] All data used in this section are quarterly and were obtained from the OECD Main Economic Indicators. For both the US and Japan, income is measured as GDP, and is deflated by the GDP deflator, the monetary variable is M1 (also deflated by the GDP deflator), and the interest rate is the three-month Treasury bill rate. The sample period is 1979Q1-1993Q4.

Next, let us investigate the possibility that money and/or interest rates are correlated with output (see, e.g., Sims (1972)). In order to do this, we first estimate a DLR regression with money and lagged money as regressors:

$$y_t = \underset{(0.698)}{1.798} + \underset{(0.163)}{0.587}\, y_{t-1} + \underset{(0.166)}{0.272}\, m_t - \underset{(0.182)}{0.125}\, m_{t-1} \tag{34}$$

$LM(1)=0.08$ $\quad LM(4)=0.29$
$Q(8)=0.11$ $\quad BJ=0.76$

The first thing to notice is that the coefficient of y_{t-1}, i.e. a_1, is now much lower, pointing towards a stable system with an exponentially decaying memory. This result implies that a_{12} (and perhaps a_{21}) is different from zero, i.e. that there exists Granger causality in at least one direction. The fact that a_1 is much less than one implies that Granger causality in the form will result in cointegration between y_t and x_t. Standard Johansen tests, however, fail to reject the null of no-cointegration (TS=7.56, λ–max =7.20). This conflicting evidence can be explained in terms of the low power of Johansen test, or interpreted as an indication that the estimated coefficient $\hat{a}_1 = 0.587$ is in fact not significantly different from one. An alternative test for the null of no cointegration against the cointegration alternative can take the form $H_0 : a_1 = 1$ against $H_1 : a_1 < 1$. A t-test for this hypothesis takes the value of 2.53, which is less than the corresponding DF critical value, suggesting that the null should not be rejected.

In view of the above conflicting evidence, we proceed to include the interest rate in the conditioning information set, thus estimating the following DLR:

$$y_t = \underset{(0.565)}{1.894} + \underset{(0.123)}{0.564}y_{t-1} + \underset{(0.119)}{0.298}m_t - \underset{(0.131)}{0.132}m_{t-1} + \underset{(0.013)}{0.029}i_t - \underset{(0.012)}{0.057}i_{t-1} \tag{35}$$

$LM(1)=0.07$ $\quad LM(4)=0.24$
$Q(8)=0.46$ $\quad BJ=0.56$

Once again the coefficient of lagged output seems to be much lower than one, providing evidence that we are dealing with a stable system. This time, however, y_t, m_t and i_t appear to be cointegrated (TS=32.89, λ–max =22.38). Moreover, the t-test for the hypothesis $a_1 = 1$ takes the value 3.544, which is greater than the corresponding DF critical value.

This example clearly demonstrates how inference about the persistence of output depends on the dynamic model chosen. Had the analysis been based on (33), one would have concluded that output follows a random walk with drift, and therefore that it is highly persistent. If, on the other hand, one estimates (35) persistence appears to be decaying at an exponential rate.

Another example will be useful to demonstrate the result that persistence is invariant to model selection if there is no Granger causality among the variables included in the alternative models. Let us consider the persistence of Japanese GDP. As before, by estimating (2) we obtain:

$$y_t = \underset{(0.147)}{0.634} + \underset{(0.013)}{0.951}\, y_{t-1} \qquad (36)$$

$LM(1)=0.11$ $\quad LM(4)=0.31$
$Q(8)=0.12$ $\quad BJ=0.46$

Which once more gives support to the random walk model with drift. If we consider the issue of persistence in the context of a DLR model where money is present we get:

$$y_t = \underset{(0.416)}{-0.314} + \underset{(0.074)}{0.999}y_{t-1} - \underset{(0.119)}{0.119}m_t + \underset{(0.102)}{0.124}m_{t-1} \qquad (37)$$

$LM(1)=011$ $\quad LM(4)=0.93$
$Q(8)=0.47$ $\quad BJ=0.15$

It is noteworthy that the coefficient of y_{t-1} is not statistically different from one, with the rest of the coefficients being characterised by relatively large standard errors. At first sight this seems to be the case where there is no cointegration between y_t and m_t, and also no Granger causality seems to be present. This is necessary and sufficient for the unit root found in the AR representation to carry forward to the DLR representation of y_t. It is also interesting that the estimate of $b_2 = -0.119$ is very close to $-b_1 = -0.124$, which is additional evidence consistent with Granger non-causality restrictions (see equation (26g)). One can conclude that model (37) is operational in the sense that the coefficient in (37) will be time independent, and (37) can be expressed and properly estimated in first differences. Therefore, persistence in Japanese output seems to be invariant to models (33) and (37).

V. CONCLUSIONS

This study has revisited an issue which has attracted a lot of attention in recent years (see, e.g. Cochrane (1988), and Campbell and Mankiw (1987a, 1987b)), i.e. persistence in macroeconomic time series. The main argument of the present paper is that persistence is not an invariant feature of a time series, but depends on the context in which the series is used. As the parameters of a dynamic model are defined relative to a particular information set, any change in the set of conditioning variables might affect the resulting estimates (see Spanos (1995b)). If one selects an alternative statistical parameterisation of the probabilistic structure of the variable of interest, inference about a statistical property such as persistence generally turns out not to be model invariant.

Having defined persistence as the memory of a series, we have shown that both Granger-causality restrictions and the stability properties of the system affect measured persistence. More specifically, in the case of stable systems both univariate

AR models and Dynamic Linear Regression (DLR) models (which simply a reparameterisation of VAR specifications) exhibit an exponentially decaying autocorrelation function, but the exponential rate is the same if and only if there is no Granger causality in either direction between the variables of the system. Conversely, in the case of unstable systems, the autocorrelation function decays to zero at a linear-like rate in the context of the univariate model, but it is exponentially decaying within a DLR model with cointegrating restrictions.

These conclusions are important for econometric practice, and suggest that economic theory should be used carefully to guide the choice of the conditioning set, or else asking the question "how persistent is GDP?" (or some other economic variable) might not be very informative about the underlying Data Generation Process (DGP) and the sources of economic fluctuations. Unless more restrictions derived from economic theory are imposed, issues such as the effectiveness of stabilisation policies cannot be settled empirically, and the debate between Keynesian and RBC theorists will remain inconclusive.

REFERENCES

Alogoskoufis, G.S., Manning A. (1988). "On the persistence of unemployment". *Economic Policy* 7, 428-469.

Beveridge, S., Nelson, C.R. (1981). "A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle". *Journal of Monetary Economics* 7, 151-154.

Blanchard, O.J., Quah, D. (1989). "The dynamic effects of aggregate demand and supply disturbances". *American Economic Review* 79, 655-673.

Blanchard, O.J., Summers, L.H. (1986). "Hysteresis and the European unemployment problem". *NBER Macroeconomics Annual.* MIT Press: Cambridge MA.

Campbell, J.Y., Mankiw, N.G. (1987a). "Are output fluctuations transitory?". *Quarterly Journal of Economics* 102, 857-880.

Campbell, J.Y., Mankiw, N.G. (1987b). "Permanent and transitory components in macroeconomic fluctuations". *American Economic Review Papers and Proceedings* 77, 111-117.

Campbell, J.Y., Perron, P. (1991). "Pitfalls and opportunities: What macroeconomists should know about unit roots", *NBER Macroeconomics Annual.* MIT Press: Cambridge MA; 141-201.

Cavaglia, S. (1992). "The persistence of real interest differentials - A Kalman filtering approach". *Journal of Monetary Economics* 29, 429-443.

Christiano, L.J. Eichenbaum, M. (1989). "Unit roots in GNP: Do we know and do we care?". *Carnegie-Rochester Conference Series on Public Policy* 32, 7-62.

Clark, P.K. (1987). "The cyclical component of US economic activity". *Quarterly Journal of Economics* 102, 797-814.

Cochrane, J.H. (1988). "How big is the random walk in GNP?". *Journal of Political Economy* 96, 893-920.

Cochrane, J.H. (1991a). "A critique of the application of unit root tests". *Journal of Economics Dynamics and Control* 15, 275-284.

Cochrane, J.H. (1991b). "Comment on 'Pitfalls and opportunities: What macroeconomists should know about unit roots' ", *NBER Macroeconomics Annual*. MIT Press: Cambridge MA; 201-210.

Cochrane, J.H. (1994). "Permanent and transitory components of GNP and stock prices". *Quarterly Journal of Economics* 109, 241-265.

Cochrane, J.H., Sbordone, A.M. (1988). "Multivariate estimates of the permanent component of GNP and stock prices". *Journal of Economics Dynamics and Control* 12, 255-296.

De Long, J.B., Summers, L.H. (1988). "On the existence and interpretation of a 'unit root' in US GNP". NBER Working Paper 2716.

Dickey, D.A., Fuller, W.A. (1981). "The likelihood ratio statistic for autoregressive time series with a unit root". *Econometrica* 49, 1057-1072.

Evans, G., Reichlin, L. (1994). "Information, forecasts, and measurement of the business cycle". *Journal of Monetary Economics* 33, 233-254.

Granger, C.W.J., Terasvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford University Press: Oxford.

Hall, S.G., Wickens, M.R. (1993). "Causality in integrated systems". Discussion Paper No. 27-93, Centre for Economic Forecasting, London Business School.

Hurst, H. (1951). "Long term storage capacity of reservoirs". *Transactions of the American Society of Civil Engineers* 116, 770-799.

King, R., Plosser, C., Stock, J.H., Watson, M.W. (1991). "Stochastic trends and economic fluctuations". *American Economic Review* 81, 819-840.

Lee, K.C., Pesaran, M.H., Pierse, R.G. (1992). "Persistence of shocks and its sources in a multisectoral model of UK output growth". *Economic Journal* 102, 342-356.

Lippi, M. Reichlin, L. (1992). "On persistence of shocks to economic variables - A common misconception". *Journal of Monetary Economics* 29, 87-93.

Lo, A.W. (1991). "Long-term memory in stock prices". *Econometrica* 59, 1279-1313.

Lupi, C. (1993). "Persistence in macroeconomic time series: Some general results". Paper presented at the Econometric Society European Meeting, Uppsala University, Uppsala, Sweden.

Mandelbrot, B. (1972). "Statistical methodology for non-periodic cycles: From the covariance to R/S analysis". *Annals of Economic and Social Measurement* 1, 259-290.

Mandelbrot, B. (1975). "Limit theorems on the self-normalized range for weakly and strongly dependent processes". *Wahrscheinlichkeitstheorie verw Gebiete* 31, 271-285.

McCallum, B.T. (1993). "Unit roots in macroeconomic time series: Some critical issues". NBER Working Paper 4368.

Nelson, C., Plosser, C. (1982). "Trends and random walks in macroeconomic time series: Some evidence and implications". *Journal of Monetary Economics* 10, 139-162.

Nimmo-Smith, I. (1979). "Linear regression and sphericity". *Biometrika* 66. 390-392.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press: London.

Shapiro, M.D., Watson, M.W. (1988). "Sources of business cycle fluctuations". *NBER Macroeconomics Annual*. MIT Press: Cambridge MA; 111-148.

Sims, C.A. (1972). "Money, income and causality". *American Economic Review* 62, 540-552.

Spanos, A. (1989). On cointegration and unit roots: Statistical parameterisation and operational models. Mimeo, Virginia Polytechnic Institute and State University.

Spanos, A. (1990). "Unit roots and their dependence on the conditioning information set". In: Fomby, T., and G. Rhodes. *Advances in Econometrics: Cointegration, Spurious Regressions, and Unit Roots*. JAI Press Inc. Greenwich, Conneticut 8,271-292.

Spanos, A. (1995a). "On normality and the linear regression model". *Econometric Reviews* 14, 195-203.

Spanos, A. (1995b). "On theory testing in econometrics: Modeling with nonexperimental data". *Journal of Econometrics* 67, 189-226.

Watson, M.W. (1986). "Univariate detrending methods with stochastic trends". *Journal of Monetary Economics* 18, 49-75.

West, K.D. (1988). "On the interpretation of the near-random-walk behavior in GNP". *American Economic Review* 78, 202-211.