Deterministic Seasonality in Dickey-Fuller Tests: Should we Care?*

Artur C. B. da Silva Lopes†
Instituto Superior de Economia
e Gestão (ISEG–UTL) and CEMAPRE

July 27, 2003

Abstract

This paper investigates the properties of Dickey-Fuller tests for seasonally unadjusted quarterly data when deterministic seasonality is present but it is neglected in the test regression. While for the random walk case the answer is straightforward, an extensive Monte Carlo study has to be performed for more realistic processes and testing strategies. The most important conclusion is that the common perception that deterministic seasonality has nothing to do with testing for the long-run properties of the data is incorrect. Further numerical evidence on the shortcomings of the general-to-specific *t-sig* lag selection method is also presented.

Keywords: unit root; Dickey-Fuller tests; similar tests; seasonality; Monte Carlo.

JEL Classification: C22, C52

*I am very grateful to João Nicolau, João Santos Silva and Uwe Hassler for helpful comments and suggestions. I am also grateful to seminar participants at ISEG, the Growth and Business Cycles in Theory and Practice 2002 Conference and the ESRC Econometric Study Group 2002 Conference, where previous versions of this paper were presented. Obviously, the usual disclaimer applies. Financial support from Fundação para a Ciência e Tecnologia (FCT) through Programa POCTI (2000/ECO/33778) is also gratefully acknowledged.

†Address: ISEG-UTL, Rua do Quelhas 6, 1200-781 Lisboa, Portugal. E-mail: asl@iseg.utl.pt.

1 Introduction

Dickey and Fuller (1979) [DF] tests are now a basic tool in the tool-kit box of most time series researchers. Following the recommendation of Ghysels (1990) and Ghysels and Perron (1993), for variables sampled at infra-annual frequencies DF tests should be applied to seasonally unadjusted (raw) data. Otherwise, their power can be even lower than usual. Although in this case the simple procedure proposed by Hylleberg et al. (1990) [HEGY] is also able to deliver evidence for the presence (absence) of seasonal unit roots, many researchers prefer using (A)DF[1] tests, the purpose of their analysis being concentrated only on the long-run (zero frequency) properties of the data. Moreover, as in some circumstances the HEGY test for the zero frequency unit root may have less power than the DF test, Franses (1996, p. 73) recommends that the latter should be used to complement the former: "... in practice, therefore, one may consider an additional step where there are no seasonal unit roots, i.e. a standard ADF test in a *regression that includes seasonal dummies*" (the italics is ours). Therefore, investigating the properties of DF tests for seasonally unadjusted data is a major concern.

On this regard, previous research — and particularly Ghysels et al. (1994) [GLN], Rodrigues and Osborn (1999) and Rodrigues (2000) — has focused exclusively on the effects of neglecting non-stationary stochastic seasonality. The major outcome of this work is that even when the data generation process (DGP) contains seasonal unit roots ADF tests can be validly used, provided that the test regression is sufficiently augmented with lags of the dependent variable to account for the presence of such non-stationary components. Otherwise, serious over-rejections of the unit root null arise. However, as the consequences of neglecting the presence of deterministic seasonality have not been addressed yet, the main purpose of this paper is precisely to fill that gap.

From a somewhat different perspective this paper addresses the issue of similarity of DF tests with respect to the parameters of the seasonal cycle. Clearly, the framework of the HEGY tests is more adequate for this purpose, the need to include the seasonal dummy variables in the test regression arising from the (seasonal) initial values. Contrasting with this approach, the presence of deterministic seasonality is

[1] Given the importance of the work of Said and Dickey (1984), a more precise acronym would be DFS.

hidden when the analysis relies on a first-order difference equation: the similarity problem seems to be present with regard only to the initial value and the value of the drift parameter. To illustrate this problem, the analysis is confined to the quarterly data case. However, it extends straightforwardly to other frequencies (e.g., monthly, weekly, etc.).

Actually, the motivation for this research arose from the observation that some practitioners do not include the usual set of seasonal intercepts when conducting ADF tests over seasonally observed time series[2], while at least some of them add those regressors to their HEGY regressions. Although we acknowledge that possibly this is not the most current practice, there does not seem to exist any research that has addressed this issue.

On the other hand, it is well known that the selection of the lag truncation parameter may affect inferences on the presence of unit roots, sometimes even dramatically [for a recent example see Murray and Nelson (2000)]. Thus, this paper also aims to investigate the finite sample behaviour of DF tests when the most popular procedure for lag selection is used jointly with a non-similar test regression. Anticipating the conclusions, some results might seem somewhat surprising, implying that as yet there is no universal, clear-cut recommendation for empirical research, yielding satisfactory size and power performance.

The remainder of this paper is organized as follows. Section 2 discusses the issue of (non-) similarity of DF tests when the data contain deterministic seasonality. As the approach is mainly analytical, a very simple DGP is used, i.e., a set of seasonal dummy variables superimposed on a random walk. In section 3 the analysis is further complicated through the consideration of more realistic DGP's and test regressions. The results of an extensive Monte Carlo study are presented to study the small sample behaviour of ADF tests when the *t-sig* general-to-specific procedure is used to select the lag truncation parameter. Section 4 presents the results of an empirical illustration where some tentative results are obtained only after a simple but somewhat detailed univariate analysis. The final section draws the most important conclusions and briefly discusses some routes for future research. A separate Appendix contains critical values that may be useful for empirical research.

[2]Some recent examples include Lim and McAleer (1999), Mártin-Alvarez et al. (1999), Metin and Muslu (1999) and Patterson (2000, pp. 295-9). In many other cases it is not clear whether the data contain only weak deterministic seasonality or have been seasonally adjusted. Another questionable example is provided in the work of Ng and Perron (2001), where GLS (local-to-unity) detrending of inflation rate series is performed using a constant only.

2 The (Non-) Similarity of DF Tests for Seasonal Time Series

More than 15 years ago, in an illuminating paper, Dickey et al. (1986) provided a straightforward answer to the question: Does the removal of seasonal means affect the limiting distribution of the DF test statistic(s)? As is well known, their answer was a clear: No! It does not. However, their result come up to be misinterpreted by many practitioners, who considered it as an indifference statement as to whether to include or not the set of seasonal dummy regressors in the test regression. Hence, this section begins formulating a simpler question: Does the non-removal of seasonal means affect the distribution of DF test statistics when the data contain deterministic seasonality? The answer is: Yes! It obviously does, as the test regression must at least account for all the deterministic components present in the DGP, and as follows straightforwardly from the work of Kiviet and Phillips (1992), *inter alia.* Hence, a further question must be posed: In what way?

Through this paper attention will be frequently focused on time series generated by the model

$$x_t = \Delta\, y_t = \mu + \sum_{i=1}^{4} \gamma_i\, D_{it} + u_t, \tag{1}$$

where y_t typically denotes a logged transformed series, μ is a drift parameter, $\sum_{i=1}^{4} \gamma_i = 0$, D_{it} $(i = 1, 2, 3, 4)$ represent the usual set of seasonal dummy variables and $\{u_t\}$ is a weakly stationary and invertible ARMA(p, q) process in the innovation sequence $\{\epsilon_t\} \sim \mathrm{iid}(0, \sigma^2)$, i.e., $\phi(L)\, u_t = \theta(L)\, \epsilon_t$, all the roots of $\phi(L)$ and $\theta(L)$ lying outside the unit circle. Notice that u_t may contain stochastic seasonality which we initially assume to be stationary.

This equation corresponds to one of the basic models typically considered in the literature on seasonality and it represents the standard model from which Miron and his co-authors[3] have derived their stylized facts about the seasonal cycle. And although the importance of deterministic seasonality seems to have been overstated by Miron, it appears that this model provides a good approximation to the behaviour of many macroeconomic time series, particularly for those corresponding to quantity

[3] See, e.g., Barski and Miron (1989) and Miron (1994, 1996). See also Hylleberg (1994), Franses et al. (1995) and Lopes (1999) for a critical appraisal of Miron's work. Actually, Miron usually adopts the parametrization $x_t = \Delta\, y_t = \sum_{i=1}^{4} \alpha_i\, D_{it} + u_t$, placing no restriction on the α_i parameters. However, the parametrization adopted here is more convenient, as it allows separating the parameters of the seasonal cycle from the overall drift (mean); cf. Ghysels and Osborn (2001, pp. 20-24).

variables. Furthermore, even when $\{u_t\}$ is seasonally non-stationary, with $\phi(L)$ containing factors such as $(1+L)$ and/or $(1+L^2)$, deterministic seasonality is also usually present at some extent [see Abeysinghe (1994) and Lopes (1999), *inter alia*].

To answer the previous question a simplified version of equation (1) was utilized for the DGP, namely

$$\Delta y_t = \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t, \quad \varepsilon_t \sim \text{nid}(0,1), \tag{2}$$

where the seasonal cycle depends on a single parameter (δ) and, without loss of generality, $\sigma_\varepsilon = 1$ (as shown below, the relevant magnitude is the standardized seasonal, $K_s = \delta/\sigma_\varepsilon$). Obviously, although providing a similar test with respect to the initial value (y_0), the $\text{DF}_{c(nd)}$ or $\tau_{c(nd)}$ test statistic obtained from the OLS regression $y_t = \alpha + \rho y_{t-1} + v_t$, or

$$\Delta y_t = \alpha + \phi\, y_{t-1} + v_t, \tag{3}$$

where $\text{DF}_{c(nd)} = \hat{\phi}/\hat{\sigma}_{\hat{\phi}}$, $\phi = 1 - \rho$, cannot provide a test for $H_0 : \rho = 1$ ($\phi = 0$) which is similar with regard to the nuisance parameter(s) reflecting the seasonal cycle. In order to achieve similarity, both exact and asymptotic, one must add the seasonal dummy regressors to (3) and estimate the regression

$$\Delta y_t = \sum_{i=1}^{4} \alpha_i\, D_{it} + \psi\, y_{t-1} + \omega_t. \tag{4}$$

Then, the corresponding $\text{DF}_{sd}(\equiv \tau_{sd} = \hat{\psi}/\hat{\sigma}_{\hat{\psi}})$ test statistic is also invariant to the value of y_0 because the unity vector lies in the space spanned by the columns of the exogenous regressors.[4] That is, this is a case where invariance of the unit root tests dispenses the addition of a redundant regressor. However, when $\mu \neq 0$ is added to (2), invariance clearly requires including the usual linear trend term in (4).

The asymptotic answer to the previous question is provided through the following result.

Proposition *Assume that the data generation process is given by equation (2), with $y_0 = 0$, but that inference on the existence of an unit root is based on equation (3) with the intercept term omitted. Then, as $T \to \infty$,*

$$\tau_{(nd)} \Rightarrow [1 + K_s^2]^{-1/2}\Big[\frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W^2(r)dr)^{1/2}} - (K_s^2/2)\Big(\int_0^1 W^2(r)dr\Big)^{-1/2}\Big)\Big],$$

[4] That is, using the obvious matrix notation $\Delta y = X\,\alpha + \psi y_{-1} + \omega$ for equation (4), it is clear that $M_x\, y_{-1}$, where $M_x = I - X'(XX)^{-1}X$, does not depend on the α parameters; cf. Kiviet and Phillips (1992).

where $\Rightarrow$ *denotes weak convergence in distribution,* $K_s = \delta/\sigma_\varepsilon$ *and* $W(r)$ *represents a standard Wiener process defined on* $[0,\,1]$.

This result follows straightforwardly from the proposition presented in Franses and Haldrup (1994) [FH], where the distribution of DF tests for time series contaminated by additive outliers (AO's) is analyzed. In fact, the relation between equations (2) and (3) is that the former implies the presence of peaks and troughs in all observations which can be viewed as AO's when the letter is considered. As the negative $(-\delta)$ and positive (δ) "AO's" occur with a "probability" $(\pi =)1/2$, the result stated above is a very simple corollary of the theorem proved by FH [5]. Moreover, when the test regression (3) contains the intercept term (and whether $y_0 = 0$ or not) the standard Wiener process is replaced by a demeaned (standard) Wiener process. In the case that the linear trend term is also included in (3), then a (demeaned and) detrended Brownian motion process arises.

Hence, it is clear that:

i) when $\delta = 0$ the limiting distribution is obviously the usual DF distribution;

ii) when $\delta \neq 0$ the limiting distribution contains the nuisance parameters reflecting the seasonal cycle and it is shifted to the left;

iii) moreover, this shift depends only on the standardized seasonal, $K_s = \delta/\sigma_\varepsilon$.

To gauge the adherence of this result to small samples, a Monte Carlo study was performed using TSP 4.5 [Hall and Cummins (1999)]. Table 1 reports some fractiles for the distributions of $\mathrm{DF}_{c(nd)}$ and DF_{sd} when the data are generated by equation (2) and regressions (3) and (4) are used to test for a unit root. The following features clearly emerge:

a) the numerical evidence closely agrees with the analytical based expectations, the shift of the distribution to the left being perceptible even when $K_s = 0.1$ and $T = 80$ only, and becoming rather dramatic as K_s grows;

b) using $T = 800$ to approximate the asymptotic distribution, except for the 0.01 and 0.99 fractiles cases, our results for DF_{sd} coincide with those of Fuller (1996, table 10.A.2, p. 642); however, in small samples the (adequate) inclusion of the

[5] Similarly, the limiting distribution of $T\hat{\phi}$ is also easily obtained from the proposition proved in FH. Notice also that for the non-differenced series only positive outliers occur with "probability" $(\pi =)$ 1/2. Actually, using FH's framework, equation (2) results from considering the DGP as $y_t = \delta\delta_t + u_t$, $u_t = u_{t-1} + \epsilon_t$, with $\delta_t = 1$ when $t \mod (2) = 0$ and $\delta_t = 0$ otherwise.

seasonal intercepts produces also a clear shift of the distribution to the right. While the reason why this effect occurs is obvious,[6] it is also clear that currently used small samples critical values are not strictly correct when one includes the seasonal dummies in the set of the deterministic regressors. For this very reason, in a separate appendix we provide the adequate critical values for the case where the trend term is added to equation (4) (see table A.1).

Table 1. Fractiles of the distribution of Dickey-Fuller test statistics based on 50 000 Monte Carlo replications. The DGP is $\Delta y_t = \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \epsilon_t,\ \epsilon_t \sim \text{nid}(0,1)$.

δ	T	0.01	0.05	0.10	0.50	0.90	0.95	0.99
					$\text{DF}_{c(nd)}$			
0.0	80	−3.52	−2.90	−2.59	−1.56	−0.42	−0.05	0.64
	160	−3.47	−2.88	−2.57	−1.56	−0.43	−0.07	0.62
	800	−3.46	−2.86	−2.58	−1.57	−0.45	−0.09	0.56
0.1	80	−3.53	−2.91	−2.60	−1.57	−0.43	−0.06	0.63
	160	−3.49	−2.89	−2.58	−1.56	−0.44	−0.08	0.60
	800	−3.47	−2.87	−2.59	−1.58	−0.45	−0.10	0.55
1.0	80	−4.76	−3.88	−3.43	−2.12	−0.99	−0.68	−0.11
	160	−4.66	−3.83	−3.40	−2.12	−1.01	−0.70	−0.13
	800	−4.57	−3.79	−3.39	−2.13	−1.03	−0.72	−0.17
5.0	80	−17.05	−13.80	−12.18	−7.43	−4.31	−3.72	−2.90
	160	−16.68	−13.59	−12.05	−7.40	−4.32	−3.72	−2.86
	800	−16.26	−13.40	−11.97	−7.41	−4.34	−3.76	−2.95
10.0	80	−33.62	−27.17	−23.97	−14.65	−8.51	−7.34	−5.79
	160	−32.86	−26.79	−23.74	−14.58	−8.53	−7.36	−5.74
	800	−32.07	−26.43	−23.61	−14.62	−8.55	−7.42	−5.83
					DF_{sd}			
any	80	−3.42	−2.82	−2.51	−1.51	−0.38	−0.02	0.67
	160	−3.43	−2.84	−2.53	−1.53	−0.41	−0.05	0.64
	800	−3.45	−2.86	−2.57	−1.57	−0.44	−0.08	0.56

Note: whereas $\text{DF}_{c(nd)}$ is obtained from regression (3), DF_{sd} results from regression (4)

Additionally, using unreported numerical results (available from the author), it is also legitimate to conclude that:

[6] Intuitively, including the seasonal dummies is equivalent to reducing the sample size, as they can be viewed as impulse dummies.

c) the asymptotic prediction is remarkably accurate even for samples as small as $T = 80$. In fact, for example, all the fractiles of the distribution of $\mathrm{DF}_{c(nd)}$ for the cases when $(\delta, \sigma_\varepsilon)$ is (1/2, 1/2) and (1, 0.2) are identical to those presented in Table 1 for $\delta = 1$ and 5 respectively.

d) As K_s grows, besides shifting to the left, the distribution of $\mathrm{DF}_{c(nd)}$ becomes also flatter (see also figure 1 in FH).

e) All these features are also present when the trend term is added to equations (3) and (4), the only effect being a slower rate of convergence of $\mathrm{DF}_{sd,t}$ to the asymptotic distribution. However, for $\mathrm{DF}_{ct(nd)}$ the small sample distributions still match very closely the asymptotically based forecasts; for example all the fractiles still coincide exactly for the cases mentioned in c) even when T is only 80.

3 Implications for Empirical Research

Having observed that neglecting deterministic seasonality has the same effect that unaccounted additive outliers occurring in all observations, the implications that follow for the properties of DF tests are obvious: spurious rejections of the unit root null will arise and the problem may become rather dramatic when the standardized seasonal is large. Intuitively, this is also simple to understand, as the unaccounted seasonal cycle produces the wrong impression that in every observation there exists a "shock" which has a purely transitory effect. When these "shocks" are large relatively to the standard deviation of the real shocks spurious evidence for stationarity will emerge very often.

3.1 The random walk case

For example, when the data is generated by equation (2) and equation (3) is used to test for the unit root then, using common nominal 5% critical values[7] for the case when $T = 80$: a) when $K_s = 0.1$ the estimated real size of the $\mathrm{DF}_{c(nd)}$ test statistic, based on 10 000 replications, is 5.12%; b) however, when $K_s = 1$, the unit root null will be rejected in about 20.45% of the times; c) for the case of strong seasonal patterns as those corresponding to $K_s = 5$ the estimated real size is 99.13%.

[7]Except when explicitly mentioned, all over this paper we have used the critical values derived from the response surface analysis of MacKinnon (1991) as these seem to be the most popular among practitioners.

As another example, we have considered the case when the seasonal cycle is given by $-\delta_1 = \delta_4 = 1$ and $\delta_2 = -\delta_3 = 0.5$. When $T = 160$ the estimated actual size of $\mathrm{DF}_{c(nd)}$ is 13.6%. Since, on the other hand, when $\delta = 0.75$ in (2) the estimated actual size is 12.76%, the simple DGP adopted here seems to provide a good approximation to more heterogeneous seasonal patterns provided that δ is close to $\sum_{i=1}^{4} |\delta_i|/4$.

Finally, for the random walk with drift case ($\mu \neq 0$), when $T = 80$ the estimated actual sizes for the cases $K_s = 0.1$, 1 and 5 are 5.07%, 32.15% and 100%, respectively. That is, for most macroeconomic time series the over-rejection problem is even more serious than for those cases where the concern is on (non-) stationary around a constant level.

3.2 More realistic settings I: size

Fortunately, DGP's such as the one of equation (2) are considered only in very special circumstances, implying also that testing for an autoregressive unit root is rarely based on equations (3) and (4). That is, a more realistic setting is the one provided by

$$\psi(L)\Delta y_t = \mu + \sum_{i=1}^{4}(-1)^i \,\delta\, D_{it} + \theta(L)\varepsilon_t, \tag{5}$$

where $\psi(L)$ may contain some root(s) on the unit circle but not equal to unity (i.e., only seasonal unit roots are allowed).

To cope with the nuisance parameters governing the additional autocorrelation, the "never-mind-deterministic-seasonality-practitioner" is assumed to base inferences on the regression

$$\Delta y_t = \alpha + \beta\, t + \phi\, y_{t-1} + \sum_{j=1}^{k} \lambda_j \Delta y_{t-j} + \epsilon_{nd,t}, \tag{6}$$

where k represents the lag truncation parameter, which we assume to be estimated using the general-to-specific (GS) *t-sig* modelling strategy, based on 5% (asymptotic) level tests, recursively performed on the λ_j parameters. Other lag length selection procedures could have been considered but, following the recommendations in Campbell and Perron (1991), Hall (1994) and Ng and Perron (1995), the *t-sig* procedure seems to be the most popular in empirical research. The corresponding $\tau_{ct(nd)}$ statistic $(\hat{\phi}/\hat{\sigma}_{\hat{\phi}})$ is denoted with $\mathrm{ADF}_{ct(nd)}$.

On the other hand, the investigation on the finite sample size performance uses as benchmark case the statistic $\mathrm{ADF}_{sd,t}$ ($\tau_{sd,t} = \hat{\psi}/\hat{\sigma}_{\hat{\psi}}$) produced by the "correct" regression

$$\Delta y_t = \sum_{i=1}^{4} \alpha_i D_{it} + \beta^* t + \psi\, y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + \epsilon_{sd,t}, \qquad (7)$$

where, using a somewhat loose notation, we also denote with k the lag truncation parameter. However, it should be clear that while this parameter is also selected using the GS, *t-sig* 5% procedure, there is no presumption that it equals the k of equation (6). Actually, the results of Taylor (2000) suggest that the k estimated using equation (6) will tend to exceed the one resulting from equation (7).

In both cases, it is also assumed that the researcher uses a "seasonally modified" deterministic rule procedure for setting the upper bound for k, k_{max}. Namely, $k_{max} = 4, 8$ and 12 were employed for $T = 48, 80$ and 160 respectively.[8] Following an almost universally adopted practice k_{min} is always set to zero.

The DGP's that we have considered for investigating the small sample size properties are the following:

$$\text{DGP1:} \quad \Delta y_t = \mu + \textstyle\sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \qquad (8)$$

$$\text{DGP2:} \quad \Delta y_t = \mu + \textstyle\sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t - \theta_4 \varepsilon_{t-4}, \qquad (9)$$

$$\text{DGP3:} \quad \Delta y_t = \mu + \textstyle\sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_4 \varepsilon_{t-4}, \qquad (10)$$

$$\text{DGP4:} \quad y_t = \mu + y_{t-4} + \textstyle\sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t, \qquad (11)$$

$$\text{DGP5:} \quad (1 + 0.9L)(1 + 0.4L^2)\Delta y_t = \mu + \textstyle\sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t. \qquad (12)$$

Given the seminal work of Schwert (1989) and the research that followed, equation (8) dispenses detailed comments [9]. Model (9) is the "seasonal twin" of DGP1 and DGP3 combines the features of DGP1 and DGP2, producing a gap in the autocorrelation function of the differenced series at lag 2.

Besides the nonseasonal unit root, DGP4 contains all the seasonal unit roots and therefore reflects the concerns of GLN, Franses (1996), Rodrigues and Osborn (1999) and Rodrigues (2000). Preliminary numerical evidence on this case has been reported

[8] These seem what we might call as resulting from a "consensual" or "popular" l_s procedure, where the l procedures presented by Schwert (1989) are adapted to the quarterly case. Two features must be noticed: a) the length of the autoregressions is not really a multiple of four, as they are equal to 5, 9 and 13, respectively; b) for the $T = 48$ and 80 cases the upper bounds for the augmenting lag lengths do not satisfy Said and Dickey (1984) condition that $k_{max}/T^{1/3} \to 0$ as $T \to \infty$. However, these features have a negligible impact on the numerical evidence.

[9] However, see also Pantula(1991). Interestingly enough, it should be pointed out that the three time series that Schwert refers as motivating examples are monthly time series not seasonally adjusted.

in Lopes (2002). On the other hand, DGP5 corresponds to a near semiannual unit root case, while the complex roots are distant from the unit circle. The motivation for considering this case arose from the observation that the seasonal unit root most commonly reported in empirical research is the root -1. Moreover, as Ghysels and Osborn (2001, p. 92) point out, a changing seasonal pattern is more likely to involve adjacent quarters then, say, reverting the roles of summer and winter. That is, the root -1 is *a priori* more plausible than the complex roots.

Several other ARIMA models were also considered as DGP's but the evidence that we got adds little to the one which is presented. Therefore it is omitted. The same argument applies to the zero drift DGP/no trend in regression case. Moreover, though we have considered $K_s = \delta = 0, 0.1, 1, 5$ and 10, only the cases $K_s = \delta = 0, 1$ and 5 are reported, the remaining cases also adding little evidence to the analysis. Hence, it must be noted that table 2 contains only the most important numerical evidence. As the main purpose is the evaluation of commonly used procedures, the 5% nominal critical values were again taken from MacKinnon (1991), both for the ADF$_{ct(nd)}$ and ADF$_{sd,t}$ statistics.

Hence, the main question seeking an answer is now the following: Does the GS, *t-sig* (5% level) method robustifies unit root inferences based on a non-similar test regression? Before observing table 2, where the answer is provided, one must take into consideration that:

i) the conjecture is that adding lags of the dependent variable to the test regression, while possibly leaving the asymptotic distribution unchanged, might alleviate the spurious stationary evidence problem in small samples as those regressors might approximate the effect of the omitted seasonal intercepts;

ii) recent research by Taylor (2000) has highlighted the shortcomings of the most common lag selection methods — and particularly of the GS *t-sig* procedure — when the test regression contains deterministic regressors, even when their presence is necessary to render the tests similar. That is, the inclusion of such regressors produces a systematic finite sample bias towards zero in the estimators of the autoregressive augmenting parameters, thereby leading to lag structures which are too much parsimonious [see also Ng and Perron (2001)]. In turn, the effect of this under-fitting is well known for the DGP's that we have considered: poor size properties in small samples.

Then, the most salient features concerning the small sample size behaviour emerging from table 2 are the following:

Table 2. Size estimates of $\mathrm{ADF}_{ct(nd)}$ and $\mathrm{ADF}_{sd,t}$ at the nominal 5% level using the GS *t-sig* 5% strategy(based on 10 000 replications)

T (k_{max})			48(4)		80(8)		160(12)	
δ	θ_1	θ_4	$\mathrm{ADF}_{ct(nd)}$	$\mathrm{ADF}_{sd,t}$	$\mathrm{ADF}_{ct(nd)}$	$\mathrm{ADF}_{sd,t}$	$\mathrm{ADF}_{ct(nd)}$	$\mathrm{ADF}_{sd,t}$
			DGP1: $\Delta y_t = \mu + \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$					
0	0.4	—	0.383	0.312	0.244	0.222	0.131	0.122
	0.8	—	0.882	0.860	0.707	0.683	0.465	0.446
1	0.4	—	0.146	0.312	0.102	0.222	0.076	0.122
	0.8	—	0.692	0.860	0.456	0.683	0.210	0.446
5	0.4	—	0.138	0.312	0.096	0.222	0.075	0.122
	0.8	—	0.501	0.860	0.280	0.683	0.151	0.446
			DGP2: $\Delta y_t = \mu + \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t - \theta_4 \varepsilon_{t-4}$					
0	—	0.4	0.215	0.181	0.180	0.152	0.116	0.101
	—	0.8	0.323	0.300	0.261	0.240	0.274	0.257
1	—	0.4	0.356	0.181	0.187	0.152	0.104	0.101
	—	0.8	0.614	0.300	0.456	0.240	0.474	0.257
5	—	0.4	0.406	0.181	0.164	0.152	0.101	0.101
	—	0.8	0.676	0.300	0.455	0.240	0.469	0.257
			DGP3: $\Delta y_t = \mu + \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_4 \varepsilon_{t-4}$					
0	0.4	0.4	0.582	0.518	0.648	0.594	0.617	0.585
	0.8	0.8	0.375	0.338	0.274	0.240	0.127	0.117
1	0.4	0.4	0.453	0.518	0.369	0.594	0.341	0.585
	0.8	0.8	0.335	0.338	0.187	0.240	0.116	0.117
5	0.4	0.4	0.558	0.518	0.330	0.594	0.326	0.585
	0.8	0.8	0.425	0.338	0.177	0.240	0.109	0.117
			DGP4: $y_t = \mu + y_{t-4} + \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t$					
0	—	—	0.045	0.076	0.061	0.057	0.064	0.061
1	—	—	0.053	0.041	0.078	0.055	0.072	0.061
5	—	—	0.011	0.040	0.075	0.056	0.060	0.060
	—	—	DGP5: $(1+0.9L)(1+0.4L^2)\Delta y_t = \mu + \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + \varepsilon_t$					
0	—	—	0.264	0.289	0.145	0.171	0.070	0.074
1	—	—	0.164	0.284	0.089	0.167	0.067	0.075
5	—	—	0.182	0.204	0.094	0.141	0.068	0.069

a) except for the seasonal unit roots case (DGP4), a clear and strong picture of serious size distortions immediately arises. However, it should be also clear that the reasons which lie behind this behaviour are very different for the two statistics.

b) As expected, the inflated rejection frequencies tend to decrease as T grows. This is a reflection of the consistency of the GS *t-sig* method, which provides little comfort for most practitioners.

c) The (non-)similarity of the ($\text{ADF}_{ct(nd)}$) $\text{ADF}_{sd,t}$ test statistic(s) is also clear but for the latter this is a two-edged-knife, particularly in the cases of DGP's 1, 3 and 5. In other words, though performing better than the $\text{ADF}_{ct(nd)}$ statistic in 56% of the cases of table 2, the $\text{ADF}_{sd,t}$ statistic is obviously less robust to the problem mentioned in ii) above and, as the sample size grows, the under-fitting problem vanishes more slowly than for this one [10].

d) Although alleviating somewhat the size distortion problem, the *t-sig* method is a poor remedy for the non-similarity of ADF tests neglecting deterministic seasonality. Actually, except for the case of DGP5, even when there is no such behaviour in the data, the $\text{ADF}_{sd,t}$ statistic is not so badly oversized as $\text{ADF}_{ct(nd)}$. Furthermore, for the (unreported) case of the random walk (e.g., DGP1 with $\theta_1 = 0$), size distortions are very small for the latter but they are relatively large for the $\text{ADF}_{ct(nd)}$ statistic, particularly when $T = 48$ and 80.

The exception mentioned in a) deserves some attention, the $t - sig$ method performing remarkably well in the case of DGP4. Following Rodrigues (2000), it should be noted that equation (11) can be written as

$$\Delta y_t = \mu + \varphi\, y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \phi_3 \Delta y_{t-3} + \sum_{i=1}^{4} (-1)^i \delta + \varepsilon_t,$$

where $\varphi = 0$ and $\phi_1 = \phi_2 = \phi_3 = -1$ and the corresponding regressors are non-stationary. Hence, their t-statistics, and particularly the one of ϕ_3, do not follow a standard distribution. Therefore, using the critical values taken from the standard normal distribution leads to a test of $H_0 : \phi_3 = 0$ whose power converges very quickly to one as T grows. So quickly, indeed, that even in relatively small samples that null is

[10] To further substantiate this argument, additional (unreported) numerical evidence is available concerning DGP5: average estimated lag lengths, estimated coefficients and power estimates for the t-test on the third augmenting lag coefficient for the regression with $k = 3$.

always rejected. That is, initiating the *t-sig* procedure with $k_{max} \geq 3$ invariably leads one to stop at least when $k = 3$, allowing to capture the presence of the non-stationary regressors and thereby ensuring the good size behaviour of the tests.

3.3 More realistic settings II: power

Obviously, a study of the power performance of both statistics is also helpful for applied researchers. Hence, the question is now the following: Is the poor remedy provided by complementing the non-similar test regression with the GS *t-sig* method cheap, as it should be? The answer is: No, clearly not!, as the numerical evidence presented below shows.

Before proceeding, one explanation must be provided. While it is quite obvious that power must be adjusted for size in the case of $\text{ADF}_{ct(nd)}$, it is not so clear that the same correction should be applied over $\text{ADF}_{sd,t}$. However, Taylor (2000) has already pointed out this problem. That is, though Hall (1994) and Ng and Perron (1995) showed that most lag selection data based methods do not asymptotically affect the distribution of ADF statistics, the case changes completely for finite samples. As Taylor (2000) emphasizes, published critical values assume that k is fixed and do not take into account neither the values of k_{max} and k_{min} nor the significance level used in the GS *t-sig* method. Therefore, we have generated the finite sample critical values for some cases and, since these are useful for empirical research, we report them in the Appendix (table A.2 [11]).

For the power performance analysis the following DGP's were considered:

$$\text{DGP6:} \quad y_t = \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + 0.01\, t + \rho\, y_{t-1} + \varepsilon_t, \tag{13}$$

$$\text{DGP7:} \quad y_t = \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + 0.01\, t + \phi_4\, y_{t-4} + \varepsilon_t, \tag{14}$$

$$\text{DGP8:} \quad (1-\phi_1 L)(1+0.9L)(1+0.4L^2)y_t = \sum_{i=1}^{4}(-1)^i \delta\, D_{it} + 0.01\, t + \varepsilon_t. \tag{15}$$

While DGP6 and DGP7 are "classical", DGP8 seems to be the most empirically relevant for our purposes. That is, there is again a near-semi-annual unit root (-1) but the complex (annual) roots lie far from the non-stationary region.

Then, what emerges from table 3 is that the remedy, although poor, is indeed very expensive: the estimated power of $\text{ADF}_{ct(nd)}$ is higher than the one of $\text{ADF}_{sd,t}$ in only 11.1% of the cases and in most of these the gain is insignificant. Unsurprisingly, almost all of these cases occur when $K_s = \delta = 0$. Contrasting with this, the estimated gains

[11] As can be seen, our critical values are much different from those obtained from MacKinnon (1991) and differ also from those in Cheung and Lai (1995), both of which assuming a fixed k.

Table 3. Size-adjusted and estimated power (for 5% level tests) of $ADF_{ct(nd)}$ and $ADF_{sd,t}$ using the GS *t-sig* (5%) lag selection method (based on 10 000 replications)

$[T](k_{max})$	$\rho(\phi_4)[\phi_1]$					
δ	$ADF_{ct(nd)}$	$ADF_{sd,t}$	$ADF_{ct(nd)}$	$ADF_{sd,t}$	$ADF_{ct(nd)}$	$ADF_{sd,t}$
DGP6: $y_t = \sum_{i=1}^{4}(-1)^i\delta\, D_{it} + 0.01\,t + \rho\, y_{t-1} + \varepsilon_t$						
	0.95		0.90		0.85	
[80](8)						
0	0.069	0.073	0.123	0.124	0.211	0.209
1	0.070	0.072	0.108	0.123	0.161	0.209
5	0.067	0.071	0.104	0.123	0.149	0.210
[160](12)						
0	0.124	0.122	0.344	0.339	0.615	0.607
1	0.105	0.123	0.215	0.339	0.332	0.608
5	0.102	0.124	0.208	0.341	0.321	0.608
DGP7: $y_t = \sum_{i=1}^{4}(-1)^i\delta\, D_{it} + 0.01\,t + \phi_4\, y_{t-4} + \varepsilon_t$						
	0.90		0.80		0.70	
[80](8)						
0	0.041	0.048	0.055	0.064	0.080	0.094
1	0.035	0.048	0.043	0.064	0.063	0.093
5	0.045	0.047	0.067	0.062	0.068	0.089
[160](12)						
0	0.061	0.068	0.124	0.134	0.246	0.261
1	0.056	0.066	0.115	0.133	0.223	0.262
5	0.056	0.067	0.143	0.134	0.238	0.259
DGP8: $(1-\phi_1 L)(1+0.9L)(1+0.4L^2)y_t = \sum_{i=1}^{4}(-1)^i\delta\, D_{it} + 0.01\,t + \varepsilon_t$						
	0.95		0.90		0.85	
[80](8)						
0	0.146	0.198	0.239	0.320	0.360	0.464
1	0.071	0.197	0.118	0.317	0.193	0.461
5	0.077	0.164	0.133	0.262	0.218	0.396
[160](12)						
0	0.117	0.137	0.314	0.341	0.585	0.604
1	0.102	0.137	0.274	0.340	0.509	0.604
5	0.104	0.131	0.283	0.334	0.540	0.602

in power resulting from accounting for deterministic seasonality are not only much more pervasive but, above all, they are much more significant. For example, for the case of DGP8 with $K_s = \delta = 1$ the power gains are, in relative terms, always above 100% and, in two cases where power is more difficult to obtain, above 150%. That is, accounting for the "seasonal AO's" really pays in power terms. Obviously, the reverse side of the coin of the previous subsection bears the liability for this: the non-removal of the "AO's" through the seasonal intercepts tends to produce more liberal lag lengths which, in turn, have the usual implication in terms of power performance.

4 Empirical Illustration

To illustrate empirically the previous analysis a simple example is provided concerning some Portuguese economic time series (see Table 4) [12]. Since the only purpose is to illustrate that analysis, we neglect the possibilities of double unit roots, outliers (besides the "seasonal" ones), structural breaks (including seasonal mean shifts), heterocedasticity, non-stationary stochastic seasonality and non-linearities. The logic is also very simple: a) when the two estimated lag truncation parameters are close we prefer using the p-value computed for the $\text{ADF}_{sd,t}$ statistic; b) otherwise, when the k's are somewhat dissimilar, a more thoughtful but simple investigation is performed using ARIMA modelling.

While for most of the series the evidence for a unit root is about the same, whether or not one considers deterministic seasonality, for three of them interesting divergencies occur. Moreover, the discrepancies could be even larger if asymptotic p-values, that do not take into consideration the presence of data based lag augmentation, were used.

The series for private and public consumption, GFCF, exports, imports and inflation are in the first group. However, a tendency for the procedure that accounts for deterministic seasonality to produce evidence more supportive of the unit root hypothesis is observed. The remaining three series seem to illustrate the analysis of subsection 3.2., i.e., it appears that they provide examples of the spurious rejection situation.

Since the selected lag length is the same, the case of the production index for the electricity industry is the most straightforward: a dramatic decision reversal occurs when the practitioner is sticked to the popular 5% level rule.

[12] All the series were collected using the publications from I.N.E. (Instituto Nacional de Estatística). They are obviously available from the author on request.

Table 4. Empirical results for some Portuguese economic time series

	sample	k_{max}	$\text{ADF}_{sd[t]}$ (k) $[p]$	$\text{ADF}_{c[t](nd)}$ (k) $[p]$
GDP	77:1–98:4	8	−2.02 (1) [0.59]	−3.29 (6) [0.06]
Priv. Cons.	77:1–98:4	8	−2.49 (4) [0.35]	−2.67 (4) [0.25]
Pub. Cons.	77:1–98:4	8	−2.20 (5) [0.50]	−2.22 (5) [0.51]
GFCF	77:1–98:4	8	−2.16 (8) [0.52]	−2.23 (8) [0.48]
Exports	60:1–98:4	12	−2.92 (7) [0.18]	−2.92 (7) [0.15]
Imports	60:1–98:4	12	−2.89 (12) [0.19]	−2.94 (12) [0.13]
Inflation	74:2–00:4	12	−1.79 (3) [0.38]	−1.87 (12) [0.37]
IPI–Total	74:1–95:4	8	−2.36 (4) [0.41]	−3.09 (8) [0.10]
IPI–Electr.	68:1–98:4	12	−3.44 (12) [0.07]	−3.47 (12) [0.04]

Notes: 1) "IPI" means industrial production index; 2) with the exception of the inflation rate, all the series were previously logarithmized; 3) the p-values for the $\text{ADF}_{sd[t]}$ statistics were estimated using Monte Carlo simulations based on 50 000 replications; 4) the p-values for the $\text{ADF}_{c[t](nd)}$ statistics were estimated using a routine built in TSP 4.5 based on Cheun and Lai(1995).

In a certain sense, the other two cases are even much stronger: neglecting deterministic seasonality in the GDP and IPI–total series drastically reduces the amount of evidence supporting the presence of a unit root. Moreover, a closer inspection reveals that in both cases the $\text{ADF}_{ct(nd)}$ statistics are (presumably) producing spurious evidence for trend stationarity. In fact, provided deterministic terms are properly considered, Box-Jenkins analysis clearly supports the shorter autoregressions associated with the $\text{ADF}_{sd,t}$ statistics. Simply allowing that longer autoregressions cope with the similarity problem is not enough in these cases.

5 Concluding Remarks

The answer to the title question is now clear: We certainly should care! The reasons are also quite obvious. We should not neglect deterministic seasonality because: a) otherwise tests will not be invariant to the parameters of the seasonal cycle; b) as a consequence, in the case of the simplest I(1) process, i.e., the random walk, the implications for the size properties of the tests may be disastrous, and the general-to-specific *t-sig* lag selection method is a poor remedy for the problem in this case and in more empirically relevant settings; c) moreover, although the remedy is poor, it is very expensive too because size-adjusted power may be much lower than in the benchmark case of the similar test.

Therefore, the main recommendation for empirical work is a straightforward extension of the one provided by Ghysels, Lee and Noh (1994, p. 432) concerning tests for seasonal unit roots: the inclusion of the seasonal dummies in the test regressions "... appears to be a prudent decision in empirical applications in order to perform tests for" both the nonseasonal and the "seasonal unit roots". This is because the common perception that deterministic seasonality has nothing to do with testing for the long-run properties of the data is incorrect. Not accounting for its presence leads to non-similar Dickey-Fuller test statistics, plagued with problems of spurious evidence for stationarity and a rather poor power behaviour.

This paper has also left some routes open for further research. In particular, concerning lag selection methods, our numerical evidence simply confirms and extends the one presented by Taylor (2000) on the size properties of Dickey-Fuller tests: in small samples and when deterministic regressors are required for similarity, the GS $t - sig$ method may perform very poorly. The following alternative methods seem to deserve attention:

a) the two-stage procedure suggested in Taylor (2000), where in the first stage the test regression is estimated omitting the deterministic regressors and selecting the lag truncation order using a data-based procedure. In the second stage the estimated k is imposed on the similar test regression.

b) Adapting the ADF^{GLS} tests proposed by Elliot, Rothenberg and Stock (1996) to the case of seasonally observed variables, possibly using the modified information criteria suggested by Ng and Perron (2001) to select the lag truncation parameter.

Concerning a), preliminary numerical evidence where we have omitted only the seasonal intercepts indicates that although alleviating the size distortion problem in some cases, in many other situations, and particularly when $T = 48$ and 80 only, significant over-rejections still subsist, and the procedure may behave worse than the ADF$_{sd,t}$ statistics. Moreover, exact similarity with respect to the parameters of the seasonal cycle is not strictly achieved. Suggestion b) seems to be more promising, both in terms of power and size performance. However, the values of the parameters for the local-to-unity GLS detrending need to be determined for the case of seasonally observed data.

6 Appendix

Table A.1. Fractiles of the distribution of Dickey-Fuller test statistic $\tau_{sd,t}$ (DF$_{sd,t}$) based on 50 000 Monte Carlo replications

T	0.01	0.05	0.10	0.50	0.90	0.95	0.99
48	−3.94	−3.32	−3.02	−2.04	−1.11	−0.80	−0.19
100	−3.94	−3.36	−3.07	−2.11	−1.17	−0.87	−0.25
160	−3.94	−3.37	−3.10	−2.14	−1.19	−0.89	−0.27
400	−3.95	−3.39	−3.11	−2.16	−1.24	−0.93	−0.31
800	−3.95	−3.40	−3.12	−2.17	−1.24	−0.92	−0.29
2000	−3.99	−3.42	−3.13	−2.18	−1.25	−0.94	−0.30

Table A.2. Finite sample critical values for the ADF$_{sd}$ and ADF$_{sd,t}$ statistics using the GS *t-sig*, 5% level method (based on 50 000 replications)

T $(k_{max}; k_{min})$	48 (4;0)		80 (8;0)		160 (12;0)	
	ADF$_{sd}$	ADF$_{sd,t}$	ADF$_{sd}$	ADF$_{sd,t}$	ADF$_{sd}$	ADF$_{sd,t}$
1%	−3.64	−4.32	−3.60	−4.32	−3.57	−4.19
5%	−2.93	−3.60	−2.96	−3.61	−2.92	−3.55
10%	−2.59	−3.24	−2.63	−3.26	−2.61	−3.24

Note: T denotes the available sample size (and not the regression length).

References

[1] Abeysinghe, T. (1994). Deterministic seasonal models and spurious regressions, *Journal of Econometrics*, 61, pp. 259-72.

[2] Barski, R. B. and Miron, J. A. (1989). The seasonal cycle and the business cycle, *Journal of Political Economy*, 97, pp. 503-34.

[3] Campbell, J. Y. and Perron, P. (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots, *NBER Macroeconomics Annual*, pp. 141-200.

[4] Cheung, Y.-W. and Lai, K. S. (1995). Lag order and critical values of the augmented Dickey-Fuller test, *Journal of Business and Economic Statistics*, 13, pp. 277-80.

[5] Dickey, D. A., Bell, W. R. and Miller, R. B. (1986). Unit roots in time series models: tests and implications, *American Statistician*, 40, pp. 12-26.

[6] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, pp. 427-31.

[7] Elliot, G., Rothenberg, T. J. and Stock, J. H. (1996). Efficient tests for an autoregressive unit root, *Econometrica*, 64, pp. 813-36.

[8] Franses, P. H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*, Cambridge University Press.

[9] Franses, P. H. and Haldrup, N. (1994). The effects of additive outliers on tests for unit roots and cointegration, *Journal of Business and Economic Statistics*, 12, pp. 471-8.

[10] Franses, P. H., Hylleberg, S. and Lee, H. S. (1995). Spurious deterministic seasonality, *Economics Letters*, 48, pp. 249-56.

[11] Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed., John Wiley & Sons.

[12] Ghysels, E. (1990). Unit-root tests and the statistical pitfalls of seasonal adjustment: the case of U. S. postwar real gross national product, *Journal of Business and Economic Statistics*, 8, pp. 145-52.

[13] Ghysels, E., Lee, H. S. and Noh, J. (1994). Testing for unit roots in seasonal time series, *Journal of Econometrics*, 62, pp. 415-442.

[14] Ghysels, E. and Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series*, Cambridge University Press.

[15] Ghysels, E. and Perron, P. (1993). The effect of seasonal adjustment filters on tests for a unit root, *Journal of Econometrics*, 55, pp. 57-98.

[16] Hall, A. (1994). Testing for a unit root in time series with pretest data-based model selection, *Journal of Business and Economic Statistics*, 12, pp. 461-70.

[17] Hall, B. H. and Cummins, C. (1999). *TSP Version 4.5 Reference Manual* and *User's Guide*, TSP International.

[18] Hylleberg, S. (1994). Modelling seasonal variation, *in Nonstationary Time Series Analysis and Cointegration*, Hargreaves, C. P. (ed.), pp. 153-78, Oxford University Press.

[19] Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S. (1990). Seasonal integration and cointegration, *Journal of Econometrics*, 49, pp. 215-238.

[20] Kiviet, J. F. and Phillips, G. D. A. (1992). Exact similar tests for unit roots and cointegration, *Oxford Bulletin of Economics and Statistics*, 54, pp. 349-67.

[21] Lim, C. and McAleer, M. (2000). A seasonal analysis of Asian tourist arrivals to Australia, *Applied Economics*, 32, pp. 499-509.

[22] Lopes, A. C. B. da S. (1999). Spurious deterministic seasonality and autocorrelation corrections with quarterly data: further Monte Carlo results, *Empirical Economics*, 24, pp. 341-59.

[23] Lopes, A. C. B. da S. (2002). The order of integration for quarterly macroeconomic time series: a simple testing strategy, forthcoming in *Empirical Economics.*

[24] MacKinnon, J. (1991). Critical values for cointegration tests, *in Long-Run Economics Relationships*, Engle, R. F. and Granger, C. W. J. (eds.), Oxford University Press, pp. 267-76.

[25] Martin-Álvarez, F. J., Cano-Fernández, V. J. and Cáceres-Hernández, J. J. (1999). The introduction of seasonal unit roots and cointegration to test index aggregation optimality: an application to a Spanish farm price index, *Empirical Economics*, 24, pp. 403-14.

[26] Metin, K. and Muslu, I. (1999). Money demand, the Cagan model, testing rational expectation vs adaptive expectations: the case of Turkey, *Empirical Economics*, 24, pp. 415-26.

[27] Miron, J. A. (1994). The economics of seasonal cycles, *in* Sims, C. A. (ed.), *Advances in Econometrics*, Sixth World Congress, vol. I, pp. 213-255.

[28] Miron, J. A. (1996). *The Economics of Seasonal Cycles*, MIT Press.

[29] Murray, C. J. and Nelson, C. R. (2000). The uncertain trend in U.S. GDP, *Journal of Monetary Economics*, 46, pp. 79-95.

[30] Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association*, 90, pp. 268-81.

[31] Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power, *Econometrica*, 69, pp. 1519-54.

[32] Pantula, S. G. (1991). Asymptotic distributions of unit-root tests when the process is nearly stationary, *Journal of Business and Economic Statistics*, 9, pp. 63-71.

[33] Patterson, K. (2000). *An Introduction to Applied Econometrics: a Time Series Approach*, St. Martin's Press.

[34] Rodrigues, P. M. M. (2000). A note on the application of the DF test to seasonal data, *Statistics and Probability Letters*, 47, pp. 171-5.

[35] Rodrigues, P. M. M. and Osborn, D. R. (1999). Performance of seasonal unit root tests for monthly data, *Journal of Applied Statistics*, 26, pp. 985-1004.

[36] Said, E. S. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, pp. 599-607.

[37] Schwert, G. W. (1989). Tests for unit roots: a Monte Carlo investigation, *Journal of Business and Economic Statistics*, 7, pp. 147-59.

[38] Taylor, A. M. R. (2000). The finite sample effects of deterministic variables on conventional methods of lag-selection in unit root tests, *Oxford Bulletin of Economics and Statistics*, 62, pp. 293-304.