

Sazonalidade em Séries Temporais Económicas
uma introdução e duas contribuições

Artur C. B. da Silva Lopes
CEMAPRE e ISEG–UTL

Junho de 2007
Instituto Superior de Economia e Gestão
Universidade Técnica de Lisboa

Nota Preliminar

O presente texto serviu como base para a apresentação da lição de síntese do autor, efectuada muito recentemente. Agradecem-se os comentários e as sugestões de Paulo M. M. Rodrigues. Naturalmente, comentários e sugestões adicionais serão muito bem vindos.

O principal objectivo deste texto é o de apresentar duas contribuições recentes do autor na área da sazonalidade das séries temporais económicas. A primeira resulta de um trabalho em co-autoria com Antonio Montañés, da Universidad de Zaragoza, e refere-se ao estudo do comportamento da potência dos testes mais populares de raízes unitárias sazonais, os testes de Hylleberg, Engle, Granger e Yoo [HEGY] (1990), quando o ciclo sazonal (determinístico) sofre uma alteração abrupta. A segunda diz respeito aos problemas com a utilização dos testes de Dickey e Fuller [DF] (1979) com dados contendo sazonalidade determinística, sobretudo quando esse padrão sazonal não é devidamente tomado em consideração. Uma vez que, em ambos os casos, o principal objectivo é o da divulgação, muitos dos pormenores contidos nesses trabalhos são omitidos ou relegados para segundo plano. O leitor mais interessado é convidado para a sua leitura.

Como se reconhece alguma marginalidade da área deste texto em relação ao centro do estudo das séries temporais económicas, a apresentação das referidas contribuições é precedida por três secções, de natureza introdutória.

Conteúdo

1	Introdução	2
1.1	Definição	2
1.2	Breve nota histórica	3
2	Principais Modelos Lineares Univariados	6
2.1	Modelo de sazonalidade determinística	6
2.2	Modelos SARIMA	9
2.3	Modelos de sazonalidade integrada	10
3	Os testes de HEGY	13
4	A Potência dos Testes de HEGY no Caso de uma Quebra no Ciclo Sazonal	15
4.1	O processo gerador de dados	16
4.2	Resultados assintóticos	17
4.3	Resultados de simulação	20
5	Testes de DF e Sazonalidade Determinística	21
5.1	A (não) similaridade dos testes de DF	22
5.2	Implicações para o trabalho empírico	25
5.3	Ilustração Empírica	27
6	Referências Bibliográficas	29

1 Introdução

1.1 Definição

Embora a noção de sazonalidade económica seja intuitivamente familiar mesmo para não economistas, convém precisar o seu significado. Nesse sentido, adopta-se a definição de Hylleberg (1992, p. 4), actualmente aceite de forma generalizada:

“Seasonality is the systematic, although not necessary regular, intra-year movement caused by the changes of the weather, the calendar and timing of decisions, directly or indirectly through the production and consumption decisions made by the agents of the economy. These decisions are influenced by endowments, the expectations and preferences of the agents, and the production techniques available in the economy.”

O principal aspecto a salientar nesta definição é que, apesar de a sazonalidade das séries económicas se apresentar, em larga medida, como um padrão que tende a repetir-se ano após ano, esse padrão não é necessariamente constante ao longo do tempo. A forma e a amplitude do ciclo sazonal podem mudar, isto é, as flutuações sazonais não são necessariamente estritamente periódicas.

A definição também separa as causas básicas, menos variáveis e de natureza mais exógena, das que são mais susceptíveis de mudar ao longo do tempo e que geralmente têm carácter mais endógeno. Entre as primeiras sobressaem “o tempo” (i.e., as condições meteorológicas) e os efeitos do calendário associados a celebrações (sobretudo religiosas, como o Natal), mas também as práticas institucionalizadas, como as que se referem às férias escolares, às datas de declaração e de pagamento de impostos, de apresentação de contas, etc. .

Por outro lado, o segundo tipo de causas está associado ao comportamento dos agentes económicos, cujos hábitos e preferências podem mudar, e muitas vezes mudam face à melhoria de condições proporcionadas pelo progresso técnico. Por exemplo, a melhoria das tecnologias de produção, de armazenamento e de transporte de produtos horto-frutícolas permite que, nos nossos dias, o seu consumo se possa efectuar, sem grande aumento de preços, em todas as estações do ano. Um outro exemplo é o da alteração do padrão de sazonalidade do consumo de energia eléctrica, com o “pico” mais alto a tender a deslocar-se do Inverno para o Verão, sobretudo como resultado da massificação da utilização de aparelhos de ar condicionado (mas também de Verões mais quentes que o usual).

1.2 Breve nota histórica

Os primeiros estudos sobre a sazonalidade económica parecem datar de meados do século XIX e ter-se-ão inspirado fortemente na ideia da decomposição das séries temporais em várias componentes não observadas, uma metodologia que tinha sido adoptada pelos astrónomos a partir do século XVII para calcular a órbita dos planetas. Também os métodos usados na meteorologia para detectar padrões periódicos parecem ter sido importantes numa fase inicial ¹.

Num artigo publicado em 1862, W. S. Jevons exprime claramente as duas possíveis atitudes perante as flutuações sazonais, manifestando já então a sua preferência pela que se viria a tornar dominante:

“Every kind of periodic fluctuation, whether daily, weekly, monthly, quarterly or yearly, must be detected and exhibited not only as a subject of study in itself, but because we must ascertain and eliminate such periodic variations before we can correctly exhibit those which are irregular or non-periodic, and probably of more interest and importance” (in Hylleberg, 1992, p. 16).

No início do século XX, com o crescimento da recolha e publicação de informação estatística sobre a economia, começaram também a ser dados os primeiros passos para a construção de metodologias de dessazonalização de séries temporais económicas. Estas foram impulsionadas sobretudo pelo trabalho de Persons, que terá sido pioneiro na explicitação formal da hipótese de componentes não observados para essas séries.

Posteriormente, já em meados do século XX, a procura de dados dessazonalizados recebeu um novo e importante impulso com o célebre trabalho de Burns e Mitchell, no N.B.E.R., sobre as flutuações cíclicas da economia. Entretanto, o forte desenvolvimento das capacidades de cálculo automático tem permitido satisfazer essa procura crescente e parece estimulá-la também. Refinamentos sucessivos têm sido introduzidos nos métodos de ajustamento sazonal, sobretudo nos anos mais recentes, aumentando a sua complexidade e alargando o fosso entre produtores e consumidores de dados. No entanto, a proporção destes últimos que encara os referidos métodos como uma “caixa preta”, que lhes permite desembaraçarem-se de uma preocupação adicional, parece vir a aumentar.

¹Exposições mais desenvolvidas sobre alguns dos tópicos aqui abordados encontram-se em Hylleberg (1992), pp. 15-25 e em Nerlove et al. (1995), cap. 1.

Entretanto, aproximadamente a partir do início da década de 70, surgiu uma corrente de investigação de contestação aos métodos de dessazonalização, inicialmente motivada pelos seus potenciais efeitos perversos na análise multivariada. De uma forma geral, as principais questões em debate eram as seguintes:

- a) será a dessazonalização individual das séries a melhor forma de proceder para efeitos de análise multivariada?
- b) (e, sobretudo) não terão as próprias flutuações sazonais das séries informação relevante que, a ser desprezada, conduzirá a distorções e a ineficiências na análise das relações entre as variáveis económicas?

Alguma da literatura deste debate é reproduzida no capítulo 1 de Hylleberg (1992), destacando-se o trabalho de Wallis (1974), pelo pioneirismo na aproximação do método de ajustamento X-11 por filtros lineares. Um bom resumo desta literatura é oferecido por Ghysels (1994) e Ghysels e Osborn (2001, cap. 5). Em Davidson e MacKinnon (2004), pp. 584-7, é apresentada uma análise simples, no contexto do modelo de regressão, em que é analisado o caso (mais favorável) de aplicação de um filtro linear uniforme a todas as séries. Os resultados obtidos podem ser sumariados da seguinte forma:

- a) no caso de exogeneidade estrita dos regressores, o estimador OLS é centrado mas ineficiente para estimar os parâmetros da relação “de interesse”, envolvendo a “componente” não sazonal das séries;
- b) a condição de pré-determinação dos regressores deixa de ser suficiente para assegurar a consistência do estimador OLS, isto é, em geral, o estimador OLS será inconsistente.

Mais recentemente, o ajustamento sazonal também tem sido questionado mesmo quando a análise é apenas univariada, uma vez que ele tende a distorcer propriedades dinâmicas importantes das séries temporais, como as da estrutura de autocorrelações [veja-se Ghysels et al. (1993)] e de invertibilidade. Com efeito, os programas usuais de dessazonalização geralmente consideram a presença de um número excessivo de raízes unitárias (sazonais e não sazonais) nas séries. Ora, a remoção inadequada dessas raízes autoregressivas introduz raízes MA não invertíveis, dificultando a modelação.

Relativamente aos testes de raízes unitárias de Dickey e Fuller [DF] (1979), Ghysels e Perron (1993) mostram que a distribuição assintótica da versão “studentizada” (τ) não é afectada por filtros lineares. Todavia, o mesmo não se passa sob a hipótese (alternativa) de estacionaridade: a “filtragem” induz um enviesamento assintótico positivo na estimação da soma dos parâmetros autoregressivos. Conjuntamente com a distorção na estrutura de autocorrelações — que, ao contrário do esperado, não permite dispensar a utilização de versões “aumentadas” do teste —, este facto tende a reduzir a potência dos testes.

Finalmente, de forma ainda mais fundamental, a hipótese central da generalidade dos procedimentos de dessazonalização é altamente questionável. De acordo com essa hipótese, qualquer série observada, y_t , pode ser decomposta em duas “componentes” ortogonais: a sazonal, y_t^s , e a que contém as “componentes de interesse”, isto é, a tendência e o ciclo, bem como a componente irregular, y_t^{ns} . Ora, mesmo aos olhos do observador casual, esta hipótese aparece como não plausível. Por exemplo, nos últimos anos, as épocas efectivas de “saldos” têm sido iniciadas muito antes do habitual (e também antes do seu início legal). Mais geralmente, é fácil observar interacções entre as “componentes” sazonal e cíclica, com as flutuações sazonais a variarem de acordo com a fase do ciclo. Franses (1996, cap. 6) sumaria alguns estudos empíricos que mostram que essas componentes não podem ser consideradas independentes. De resto, que as flutuações sazonais não são neutras relativamente à análise dos ciclos económicos pode ser ilustrado com uma notícia de jornal: *“Japão” Apaga Duas Recessões dos Anos 90. [...] Uma nova forma de contabilizar as variações sazonais na actividade económica permitiu ao Japão desfazer-se de duas das suas três recessões da década de 90: as dos segundos semestres de 1992 e 1993...* (in Público, 4/12/2001).

Por outro lado, até a ausência de interacção com a “componente” de tendência não parece razoável. O progresso técnico (de longo prazo) é parcialmente induzido pela necessidade de suavizar as flutuações sazonais decorrentes do “tempo” e, como foi referido, também lhes introduz alterações. Essa necessidade dos agentes económicos é também constrangida por outros factores que podem ter natureza de mais longo prazo, como as restrições orçamentais.

2 Principais Modelos Lineares Univariados

Da exposição anterior deverá ter ficado claro que se preconiza a modelação explícita das flutuações sazonais e não a sua remoção prévia. Ora, mesmo quando se pretende efectuar análise multivariada, torna-se indispensável iniciar o estudo pela análise univariada das séries envolvidas. Nesse domínio, o final dos anos 80 e a década de 90 foram marcados pelo debate entre duas linhas de investigação que, embora unidas na contestação à subalternização da “componente” sazonal, adoptaram perspectivas quase diametralmente opostas. Uma, liderada por J. Miron, aceita a ideia básica de que uma grande proporção das flutuações sazonais é repetitiva (e, portanto, facilmente previsível) para salientar a utilidade e a importância do estudo dessas flutuações para a identificação dos comportamentos dos agentes económicos. A outra, encabeçada por S. Hylleberg, assenta a sua estratégia precisamente na rejeição dessa ideia básica: muitas séries económicas apresentam um padrão sazonal que não é constante e que tende a mudar ao longo do tempo. O reconhecimento de que este debate transpõe, para o domínio da sazonalidade, o (incorrectamente denominado) debate “tendência determinística *versus* estocástica”, respeitante ao comportamento de longo prazo, parece ser imediato.

Assim, nesta secção apresentam-se os dois principais modelos lineares univariados, correspondentes a essas correntes: o de sazonalidade determinística e o de sazonalidade estocástica integrada ². Revêem-se também, de forma sumária, os modelos SARIMA. Como se pressupõe que o objectivo final é o da modelação multivariada, e para evitar a adopção de uma notação algo “pesada”, considera-se que os dados observados são trimestrais. Todavia, a exposição pode ser estendida com alguma facilidade aos casos de outros tipos de dados e, em particular, ao caso de séries mensais.

2.1 Modelo de sazonalidade determinística

O modelo de sazonalidade determinística é o modelo tradicionalmente empregue em econometria, e que pressupõe um padrão sazonal regular e estável no tempo, com “picos” e “vales” sempre localizados nos mesmos trimestres (ou estações).

²Também os modelos autoregressivos periódicos têm sido objecto de alguma atenção, se bem que num plano inferior, devido ao facto de necessitarem de amostras de grande dimensão. Sobre estes modelos veja-se, por exemplo, Franses (1996, caps. 7 a 9) e Ghysels et al. (2006).

O processo trimestral x_t é um processo de sazonalidade determinística quando a sua representação é dada por

$$x_t = \sum_{j=1}^4 \alpha_j D_{tj} + u_t, \quad t = 1, 2, \dots, T, \quad (1)$$

com D_{tj} ($j = 1, \dots, 4$) representando as usuais *dummies* sazonais e u_t um processo com média nula, fracamente estacionário e invertível, isto é, por exemplo, $\alpha(L)u_t = \beta(L)\epsilon_t$, com $\epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$ e $\alpha(L)$ e $\beta(L)$ polinómios no operador de desfasamento, L , com todas as raízes fora do círculo unitário. Embora u_t possa conter também alguma sazonalidade, o essencial desta é representado pelos coeficientes das *dummies* sazonais, que representam a média do processo para cada trimestre, isto é, o chamado “ciclo sazonal”. Dada a estacionaridade de u_t , o processo tenderá a apresentar um comportamento de reversão para essas médias sazonais. Note-se, no entanto, que, como a sua média varia com o trimestre, x_t não pode ser considerado um processo estacionário. Todavia, tal como para os processos estacionários em tendência, esta não estacionaridade não coloca problemas especiais e pode ser removida facilmente.

Nas aplicações empíricas, x_t representa usualmente uma série com a tendência removida — regra geral por diferenciação, i.e., $x_t = \Delta y_t$ — e que foi também previamente logaritmicada. Nesse caso, os coeficientes α_j ($j = 1, \dots, 4$) representam as taxas médias (aproximadas) de variação trimestral da série original. É assim no caso dos trabalhos de Miron e co-autores (veja-se, por exemplo, Miron, 1994), onde o modelo (1) é empregue para extrair dos dados factos estilizados sobre o ciclo sazonal e para comparar e estabelecer semelhanças entre as flutuações sazonais e as cíclicas. Para esses efeitos, são utilizados:

- 1) o coeficiente de determinação (R^2), que é interpretado como representando a proporção da variação sazonal que pode ser atribuída à sazonalidade determinística;
- 2) as estimativas OLS dos coeficientes, para descrever e comparar os padrões sazonais das várias variáveis, e
- 3) a estatística F para testar a hipótese $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$, i.e., a ausência de sazonalidade determinística.

Estes procedimentos são criticados em Franses et al. (1995) e em Lopes (1999), pois desprezam totalmente a possibilidade de as séries conterem raízes unitárias sazonais (veja-se a subsecção 2.3), caso em que se tornam inválidos, produzindo frequentemente inferências errôneas.

O modelo de (1) não permite isolar a média global ou não condicional de x_t , $\mu = E(x_t)$, dos desvios trimestrais em relação a essa média. Para esse efeito, a representação adequada é dada por

$$x_t = \mu + \sum_{j=1}^4 \gamma_j D_{tj} + u_t, \quad (2)$$

onde $\mu = (1/4) \sum_{j=1}^4 \alpha_j$ e $\gamma_j = \alpha_j - \mu$ ($j = 1, \dots, 4$) representam os referidos desvios médios trimestrais relativamente à média global, ou seja, $\sum_{j=1}^4 \gamma_j = 0$. Sem a imposição desta restrição, contudo, o modelo constitui um exemplo clássico da “armadilha das variáveis artificiais”. Impondo-a e tomando, por exemplo, o quarto trimestre como padrão ou referência, tem-se

$$x_t = \mu + \sum_{j=1}^3 \gamma_j D_{tj}^* + u_t,$$

onde $D_{tj}^* = D_{tj} - D_{t4}$, $j = 1, 2, 3$.

Ainda uma forma alternativa de representação que pode, por vezes, revelar-se útil, é a forma trigonométrica, que permite relacionar os coeficientes das *dummies* com as frequências (espectrais) sazonais (veja-se, por exemplo, Ghysels e Osborn, 2001, pp. 21-24). Para o caso de dados trimestrais, esta é dada por

$$x_t = \mu + \theta_1 \cos(\pi t/2) + \beta_1 \sin(\pi t/2) + \theta_2 \cos(\pi t) + u_t,$$

cujos coeficientes estão relacionados com os da equação (2) através das igualdades $\gamma_1 = \mu + \beta_1 - \theta_2$, $\gamma_2 = \mu - \theta_1 + \theta_2$, $\gamma_3 = \mu - \beta_1 - \theta_2$ e $\gamma_4 = \mu + \theta_1 + \theta_2$. Naturalmente, esta representação torna claro que, para dados trimestrais, as frequências sazonais são $\pi/2$ (a frequência fundamental) e π (a frequência de Nyquist). Os coeficientes da frequência $\pi/2$, θ_1 e β_1 , referem-se ao ciclo anual, pois estão associados a funções que completam um ciclo em cada quatro trimestres, assumindo os valores 1, 0, -1 e 0. Note-se que tanto θ_1 como β_1 estão associados a dois semi-ciclos semestrais: θ_1 ao dos segundo e quarto trimestres e β_1 ao dos primeiro e terceiro. Por outro

lado, o coeficiente da frequência π , θ_2 , corresponde ao ciclo semi-anual (semestral), pois está associado a uma função $[\cos(\pi t)]$ cujos valores se repetem em cada dois períodos, alternando entre -1 e 1 .

2.2 Modelos SARIMA

É sabido que o primeiro passo da análise tradicional de Box & Jenkins consiste na diferenciação ordinária e/ou sazonal da série, tantas vezes quantas as necessárias para atingir a estacionaridade. Representem-se com d e D , respectivamente, esses números de vezes, isto é, de aplicação dos filtros $(1 - L)$ e $(1 - L^4)$. A série assim obtida, $\Delta_4^D \Delta^d y_t = (1 - L^4)^D (1 - L)^d y_t$, é modelada como um processo ARMA, com polinómios AR e MA distintos para as componentes sazonal e não sazonal. Tem-se, assim, o modelo geral multiplicativo ARIMA sazonal, SARIMA $(p, d, q) \times (P, D, Q)$,

$$\phi_p(L) \Phi_P(L^4) (1 - L^4)^D (1 - L)^d y_t = \theta_q(L) \Theta_Q(L^4) \epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$$

onde $\phi_p(L)$, $\Phi_P(L)$, $\theta_q(L)$ e $\Theta_Q(L)$ são polinómios invertíveis de ordens p , P , q e Q , respectivamente.

A versão mais conhecida destes modelos é o muito parcimonioso “modelo das linhas aéreas”,

$$\Delta_4 \Delta y_t = (1 - \theta_1 L)(1 - \theta_4 L) \epsilon_t,$$

onde $\phi_p(L) = \Phi_P(L) = 1$, $d = D = 1$ e $|\theta_1| < 1$ e $|\theta_4| < 1$, que tem sido frequentemente usado com muito êxito em previsão. Todavia, a utilização geral deste modelo em economia, sobretudo para fins descritivos, permanece muito limitada. Algumas justificações podem ser dadas para este facto:

- a) a falta de suporte empírico, fornecido por métodos estatísticos formais, para a utilização conjunta dos filtros $(1 - L)$ e $(1 - L^4)$;
- b) as dificuldades que resultam da presença de componentes MA, e
- c) a dificuldade em atribuir aos coeficientes uma interpretação económica clara (que o modelo de sazonalidade determinística permite).

Relativamente a este último ponto, pode ler-se em Davidson e MacKinnon (2004, p. 581): “*By themselves, seasonal ARMA processes cannot capture one important feature of seasonality, namely the fact that different seasons of the year have different characteristics: summer is not just winter with a different label.*”

2.3 Modelos de sazonalidade integrada

Embora tendo por origem mais próxima os modelos da subsecção anterior, os modelos de sazonalidade integrada podem, também, ser vistos como tendo por base o modelo tradicional de sazonalidade estocástica estacionária e, em particular, o AR(1) sazonal [SAR(1)] estacionário,

$$x_t = \phi_4 x_{t-4} + \epsilon_t, \quad |\phi_4| < 1,$$

ao qual podem ser acrescentadas *dummies* sazonais. De facto, o processo mais simples deste tipo é o passeio aleatório sazonal, que se obtém com $\phi_4 = 1$, isto é,

$$x_t = x_{t-4} + \epsilon_t \quad \text{ou} \quad \Delta_4 x_t = \epsilon_t. \quad (3)$$

Este processo é também o exemplo mais simples do processo sazonal de raízes unitárias, pois o filtro de diferenciação sazonal pode ser decomposto da seguinte forma:

$$\begin{aligned} \Delta_4 &= (1 - L^4) = (1 - L)S(L), \\ &= (1 - L)(1 + L + L^2 + L^3), \\ &= (1 - L)(1 + L)(1 + iL)(1 - iL), \end{aligned}$$

onde $S(L)$ representa o filtro de soma móvel $(1 + L + L^2 + L^3)$ e $i^2 = -1$. Isto é, o passeio aleatório trimestral contém quatro raízes unitárias: a usual raiz 1, correspondente à frequência de longo prazo, a raiz -1 , correspondente à frequência semi-anual (π), e o par $\pm i$, relativo à frequência anual ($\pi/2$).³

Resolvendo recursivamente a equação (3), obtém-se

$$\begin{aligned} x_t &= \epsilon_t + \epsilon_{t-4} + \epsilon_{t-8} + \dots \\ &= \epsilon_t + \sum_{i=1}^{\text{int}[(t-1)/4]} \epsilon_{t-4i} + \sum_{j=1}^4 x_{j-4} D_{tj}, \end{aligned} \quad (4)$$

onde x_{j-4} ($j = 1, \dots, 4$) representam os valores iniciais. Esta representação torna claro que estes processos podem conter uma componente sazonal determinística (que pode ser importante), incorporada através dos valores iniciais. Ela permite ainda constatar que:

³Na realidade, como se pode constatar usando coordenadas polares, a raiz $-i$ corresponde à frequência $(3/2)\pi$. Todavia, os ciclos com esta frequência são indistinguíveis dos que correspondem à frequência $\pi/2$.

- i) o passeio aleatório trimestral é composto por quatro passeios aleatórios independentes, um para cada trimestre;
- ii) as suas propriedades são muito semelhantes às do passeio aleatório usual e, em particular:
 - a) todos os choques têm um efeito permanente sobre os valores assumidos pelo processo para cada um dos trimestres, de forma que alguns deles, de magnitude elevada, podem alterar o padrão sazonal de forma duradoura;
 - b) a sua variância cresce linear e anualmente desde o início do processo.

Com exceção de i), estas propriedades também caracterizam um processo sazonal integrado geral como

$$\phi(L)\Delta_4 x_t = \theta(L)\epsilon_t, \quad (5)$$

com $\phi(L)$ e $\theta(L)$ polinómios invertíveis. Esta exceção não invalida, no entanto, a extensão a este caso geral de uma importante implicação de i): não pode existir cointegração entre as séries dos vários trimestres de um processo sazonal integrado (veja-se, por exemplo, Ghysels e Osborn, 2001, pp. 46-7), o que retira plausibilidade à utilização destes modelos para séries económicas. Destas características, decorre ainda uma outra implicação pouco razoável: a volatilidade sazonal permitida parece exceder substancialmente a observada empiricamente, isto é, “o Verão pode tornar-se Inverno”. Todavia, a equação (4) também permite constatar que, se a variação sazonal dos valores iniciais for grande relativamente à variância dos choques (σ_ϵ^2), o padrão sazonal observado (em amostras finitas) pode ser bastante estável durante muito tempo. Assim, o modelo de sazonalidade integrada deve ser visto apenas como uma aproximação parcimoniosa e alternativa ao de sazonalidade determinística, e mais adequada que este para séries com um padrão sazonal bastante variável ao longo do tempo. Como é notado em Engle et al. (1993), a escolha entre os dois modelos depende do grau de variação desse padrão.

Por outro lado, embora o modelo de sazonalidade integrada necessite da presença de uma componente determinística para se afirmar como aproximação plausível, não é razoável que esta possa ser introduzida de forma explícita nas equações (3) ou (5), através do conjunto usual de *dummies* sazonais. Com efeito, procedendo como em (4), observar-se-ia o surgimento de tendências distintas para os vários trimestres,

isto é, o processo apresentaria também tendências trimestrais divergentes. Embora em pequenas amostras algumas (poucas) séries pareçam ser bem aproximadas por modelos deste tipo, em geral, um comportamento deste género não é sustentável em termos económicos. Desta forma, quando se pretende representar a tendência da série, o procedimento usual consiste em introduzir um termo de deriva (*drift*) comum aos vários trimestres, como no modelo

$$\phi(L)\Delta_4 y_t = \gamma + \theta(L)\epsilon_t.$$

Finalmente, a decomposição do operador de diferenciação sazonal ou anual apresentada atrás também é útil para clarificar dois aspectos:

- a) contrariamente à associação tradicional de sazonalidade estocástica, tanto estacionária como integrada, exclusivamente ao desfasamento de ordem 4 (anual), desfasamentos de ordens inferiores a esta podem ser sazonais;
- b) (também contrariamente à ideia tradicional) um processo integrado sazonal pode conter apenas a raiz -1 ou as raízes $\pm i$ na sua representação autoregressiva.

Relativamente a a), a decomposição é explícita para o caso integrado. Para o caso estacionário veja-se, por exemplo, Ghysels e Osborn (2001, pp. 25-6). Por outro lado, associada a b), em Lopes (2003) propõe-se a seguinte definição de processo trimestral integrado: “*the quarterly stochastic process with no deterministic component, x_t , is said to be integrated of orders d_0, d_1, d_2 , denoted $x_t \sim I(d_0, d_1, d_2)$, if $(1-L)^{d_0}(1+L^2)^{d_1}(1+L)^{d_2}x_t$ is a stationary and invertible ARMA process*”. Além de seguir a tradição da definição de Engle e Granger (1987), esta definição permite separar as diferentes raízes unitárias, associando a cada uma a respectiva periodicidade, por ordem crescente (e, só implicitamente, a frequência ⁴). Assim, por exemplo, se após a remoção da componente determinística, o processo y_t requer a aplicação do filtro $(1-L)^2(1+L)$ para se tornar estacionário, é representado como sendo $I(2, 0, 1)$.

⁴Para uma discussão mais detalhada veja-se a versão de documento de trabalho deste artigo.

3 Os testes de HEGY

Pela sua semelhança com os testes DF e pela sua simplicidade e flexibilidade, permitindo testar separadamente as raízes unitárias correspondentes aos factores do filtro de diferenciação sazonal, os testes de Hylleberg, Engle, Granger e Yoo [HEGY] (1990) são os testes mais populares de raízes unitárias sazonais.

Considere-se uma série, y_t , possivelmente contendo tendência, e represente-se com x_t a correspondente série com a componente determinística removida, com representação autoregressiva $\phi'(L)x_t = \epsilon_t$, com $\phi'(L)$ um polinómio autoregressivo de ordem p . A regressão auxiliar dos testes HEGY baseia-se na expansão do polinómio $\phi'(L)$ em torno dos pontos ± 1 e $\pm i$, e é dada por⁵

$$\Delta_4 y_t = d_t + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + \sum_{i=1}^k \gamma_i \Delta_4 y_{t-i} + \epsilon_t, \quad (6)$$

onde d_t representa a componente determinística (usualmente dada por $\sum_{j=1}^4 \alpha_j D_{tj} + \beta t$), o penúltimo termo, com $k = p - 4$, representa o resto da expansão [e é necessário para assegurar que $\epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$] e os regressores representados com $y_{1,t}$, $y_{2,t}$ e $y_{3,t}$ são transformações de y_t dadas por:

$$\begin{aligned} y_{1,t} &= (1 + L)(1 + L^2)y_t = S(L)y_t, \\ y_{2,t} &= -(1 - L)(1 + L^2)y_t = -(1 - L + L^2 - L^3)y_t, \\ y_{3,t} &= -(1 - L)(1 + L)y_t = -(1 - L^2)y_t. \end{aligned}$$

HEGY mostraram que a presença da raiz 1 implica que $\pi_1 = 0$, que a da raiz -1 implica que $\pi_2 = 0$ e que a das raízes complexas implica $\pi_3 = \pi_4 = 0$. As respectivas alternativas estacionárias são $\pi_1 < 0$, $\pi_2 < 0$ e $\pi_3 < 0$ e/ou $\pi_4 \neq 0$. Assim, de forma semelhante à regressão DF, são impostas todas as raízes unitárias sobre a variável dependente, enquanto que em cada uma das variáveis $y_{1,t}$, $y_{2,t}$ e $y_{3,t}$ é (são) preservada(s) a(s) raiz(izes) unitária(s) testada(s) e impostas as restantes.

Desta forma, as estatísticas de teste HEGY são os rácios- t de π_1 e de π_2 , t_1 e t_2 , respectivamente, e a estatística F associada a $\pi_3 = \pi_4 = 0$, F_{34} . Naturalmente,

⁵Além do artigo original de HEGY, veja-se também Burke (1996), para uma explicação mais detalhada. Uma abordagem alternativa, bastante intuitiva, é apresentada em Ghysels e Osborn (2001), pp. 60-1, ou em Ghysels et al. (2001).

sob as respectivas hipóteses nulas, nenhuma destas estatísticas tem uma distribuição assintótica *standard*. Em particular, as distribuições de t_1 e de t_2 são a DF, no caso de t_2 a versão sem tendência [veja-se HEGY e Engle et al. (1993), bem como Ghysels et al. (2001) para uma exposição mais detalhada]. Tabelas de valores críticos para algumas dimensões de amostras podem ser consultadas, por exemplo, em HEGY e em Lopes e Montañés (2005).

De forma bastante sumariada, refiram-se os seguintes aspectos adicionais (também semelhantes aos dos testes DF):

- a) a similaridade dos testes relativamente à eventual presença de deriva e de sazonalidade nos valores iniciais do processo de geração de dados (PGD) requer a presença, na regressão de teste, de um termo de tendência e do conjunto de *dummies* sazonais, respectivamente;
- b) a presença desses regressores também é necessária para que os testes HEGY apresentem potência razoável contra as alternativas de estacionaridade em tendência e de sazonalidade determinística, respectivamente [veja-se, por exemplo, Ghysels et al. (1994)];
- c) os testes podem ser afectados por problemas graves de dimensão quando o PGD contém uma componente MA sazonal negativa importante [veja-se Ghysels et al. (1994)];
- d) em pequenas amostras, o comportamento de potência dos testes é algo desapontador, mesmo contra alternativas não muito próximas das respectivas hipóteses nulas [veja-se, por exemplo, Lopes (2003)].

Sobretudo ao longo da década de 90, os testes HEGY foram frequentemente empregues para analisar as características de sazonalidade de vários conjuntos de séries macroeconómicas trimestrais. Veja-se, por exemplo, Franses (1996), Hylleberg et al. (1993) e Osborn (1990) e, para o caso português, Amado (2000) e Lopes e Veiga (1999). A conclusão a extrair é que, embora a possibilidade esteja longe de poder ser desprezada, a presença de raízes unitárias sazonais é um fenómeno muito menos generalizado que o das raízes unitárias na frequência nula. Em particular, a utilização do filtro de diferenciação sazonal recebe muito escassa sustentação empírica. Por outro lado, como seria de esperar, encontra-se mais frequentemente suporte

empírico da presença da raiz -1 que das raízes complexas. De facto, recorde-se que é a primeira que está associada a ciclos de uma frequência (π) que envolve alternância entre valores de trimestres adjacentes, sendo, portanto, mais verosímil que o padrão sazonal apresente variabilidade relativamente a esses ciclos do que, por exemplo, invertendo os papéis do “Verão e do Inverno”.

4 A Potência dos Testes de HEGY no Caso de uma Quebra no Ciclo Sazonal

O trabalho seminal de Perron (1989) sobre os efeitos das quebras de estrutura nos testes de raízes unitárias é bem conhecido. Nesse trabalho, Perron mostrou que, no caso de processos estacionários em torno de tendências segmentadas, as estatísticas DF podem ter o seu comportamento de potência seriamente afectado. Adicionalmente, para um dos casos analisados, Perron argumentou que os testes DF seriam afectados mesmo assintoticamente, isto é, tornar-se-iam inconsistentes.

Baseado nesse trabalho, Ghysels (1994, p. 298) conjecturou que “... *a similar phenomenon may appear at the seasonal frequency. More precisely, seasonal mean shifts may lead us to accept incorrectly the presence of a unit root at the so-called seasonal frequency and thus may affect the statistical properties of tests for a seasonal unit root...*”. Ou seja, tal como os testes DF, sobre a presença de uma raiz unitária na frequência de longo prazo, podem ser iludidos por uma quebra na função de tendência, também os testes HEGY poderão, frequentemente, produzir resultados espúrios em favor das raízes unitárias sazonais quando o ciclo sazonal (determinístico) sofre uma alteração.

Em particular, podemos conceber facilmente que padrões sazonais que parecem variar ao longo do tempo possam, na realidade, resultar da presença de uma (ou mais) quebras de estrutura súbitas de um ciclo sazonal determinístico. Essas quebras podem reflectir alterações nas preferências dos agentes económicos, ou nas técnicas de produção e de armazenamento, ou, talvez mais frequentemente, podem traduzir mudanças em práticas institucionalizadas, resultantes de alterações do enquadramento legal (por exemplo, sobre o período escolar ou sobre as datas fiscais). Podem também resultar de alterações nos procedimentos usados para medir as variáveis económicas. Por exemplo, no caso português, o INE alterou a metodologia

de produção dos dados das Contas Nacionais Trimestrais após 1986.

Estudos de simulação que confirmaram a conjectura de Ghysels (1994) foram efectuados por Smith e Otero (1997) e por Balcombe (1999). Todavia, estes trabalhos não abordaram o comportamento assintótico dos testes HEGY. Também o desempenho de potência dos testes em pequenas amostras, face a diferentes tipos de quebras, não ficou completamente esclarecido. O trabalho de Lopes e Montañés [LM] (2005), que a seguir se apresenta de forma abreviada, procurou precisamente preencher essas lacunas.

4.1 O processo gerador de dados

Embora o interesse principal resida na análise dos testes de raízes unitárias sazonais, para estudar também o comportamento do teste sobre a raiz 1, o PGD considerado é estacionário em tendência:

$$y_t = \sum_{j=1}^4 \alpha_j D_{tj} + \beta t + \sum_{j=1}^4 \delta_j D_{tj} I(t > \tau) + \epsilon_t, \quad t = -3, -2, \dots, T, \quad (7)$$

onde $I(t > \tau)$ é uma função indicatriz, assumindo o valor 1 quando $t > \tau$ e 0 no caso contrário. Assim, tem-se apenas uma única quebra no padrão de sazonalidade, após o período $\tau = \lambda T$, $0 < \lambda < 1$, e esta ocorre de forma abrupta, ou seja, é do tipo *outlier* aditivo (AO, *additive outlier*). Desta forma, o ciclo sazonal é representado pelos parâmetros α_j ($j = 1, \dots, 4$) até esse período, e por $\alpha_j + \delta_j$ ($j = 1, \dots, 4$) após o momento da quebra. Sem perda de generalidade, assume-se que τ corresponde a uma observação do quarto trimestre.

Contrariamente a uma hipótese por vezes postulada na literatura sobre alterações de estrutura, é assumido que $\delta_j < \infty$, $j = 1, \dots, 4$. Ou seja, não se permite que a magnitude da quebra possa crescer com T (como, por exemplo, com $\delta_j = \psi_j \tau$, $j = 1, \dots, 4$). De facto, como se pretende estudar o comportamento dos testes quando só um (alguns) dos trimestres é (são) afectado(s), não faria sentido considerar a possibilidade de apenas um (alguns) deles passar(em) a apresentar tendência(s) trimestral(is divergentes) após a data da quebra. Mais geralmente, uma hipótese desse tipo seria admissível apenas no caso em que o PGD inicial já apresentasse um

comportamento desse género, ou seja, fosse dado por um modelo como

$$y_t = \sum_{j=1}^4 \alpha_j D_{tj} + \sum_{j=1}^4 \beta_j D_{tj} t + u_t,$$

com u_t um processo estacionário e invertível, isto é, um processo que permite que a variação sazonal determinística cresça com o tempo mas que, como foi referido atrás, é pouco plausível e pouco adequado à generalidade das séries económicas. Assim, a hipótese assumida parece ser algo restritiva apenas no que respeita ao teste da raiz na frequência nula.

4.2 Resultados assintóticos

Considere-se que o PGD é dado pela equação (7), com $\delta_j < \infty$, ($j = 1, \dots, 4$), e que a versão simples, não aumentada, da regressão de teste (6) é empregue, com d_t incluindo as *dummies* sazonais e o termo de tendência. Então, tem-se:

- a) $T^{-1/2}t_1 \xrightarrow{p} -c_1$,
- b) $T^{-1/2}t_2 \xrightarrow{p} -c_2$,
- c) $T^{-1}F_{34} \xrightarrow{p} c_3$,

onde c_1 , c_2 e c_3 são constantes positivas, dadas por expressões bastante complexas (veja-se LM). Ou seja, as estatísticas t_1 e t_2 divergem para $-\infty$ à taxa $T^{1/2}$ e a estatística F_{34} diverge para $+\infty$ à taxa T . Por conseguinte, os testes HEGY mantêm-se consistentes, isto é, não são afectados assintoticamente por quebras de magnitude finita no ciclo sazonal.

Todavia, a complexidade das expressões limite e a necessidade de investigar o comportamento dos testes perante diferentes tipos de quebras requerem a particularização do estudo para vários casos. Os casos particulares considerados em LM são os seguintes:

- caso1 : $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta$,
- caso2 : $\delta_1 = \delta$, $\delta_2 = \delta_3 = \delta_4 = 0$,
- caso3 : $\delta_1 = -\delta_2 = \delta$, $\delta_3 = \delta_4 = 0$,
- caso4 : $\delta_1 = -\delta_3 = \delta$, $\delta_2 = \delta_4 = 0$,

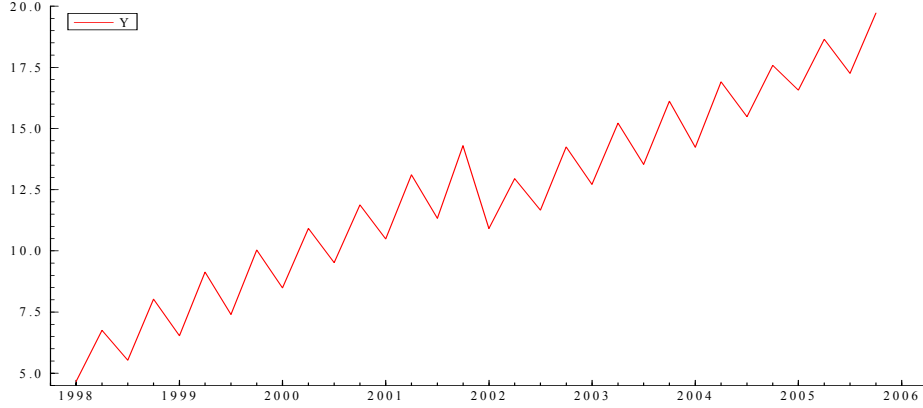


Figura 1: Um exemplo do caso 1 ($\delta < 0$).

ou seja, considerou-se um único parâmetro de quebra (δ), facto que permite simplificar a análise. No caso 1 não existe, verdadeiramente, uma alteração do padrão sazonal, mas tão somente uma mudança no nível da série (por exemplo, um *crash* que afecta igualmente todos os trimestres, como é ilustrado na figura 1). No caso 2 ocorre uma mudança do ciclo sazonal que afecta apenas um dos trimestres, provocando também uma mudança de nível. Os casos 3 e 4 são os que apresentam maior interesse para o analista da sazonalidade, pois correspondem a quebras sazonais “puras”, que não alteram o nível da série, dado que $\sum_{j=1}^4 \delta_j = 0$. Com $\delta > 0$, no caso 3 um *boom* no Inverno é compensado por um *crash* na Primavera, e no caso 4 essa compensação só ocorre no trimestre de Verão (como é ilustrado na figura 2).

A título de exemplo, considerem-se os casos 1 e 4. Para o caso 1, tem-se:

$$1.a) \quad T^{-1/2} t_1 \xrightarrow{p} -\sqrt{\frac{\sigma_\epsilon^2}{20(1-3B)B\delta^2+4\sigma_\epsilon^2}},$$

$$1.b) \quad T^{-1/2} t_2 \xrightarrow{p} -\sqrt{\frac{4(1-3B)B\delta^2+\sigma_\epsilon^2}{20(1-3B)B\delta^2+4\sigma_\epsilon^2}},$$

$$1.c) \quad T^{-1} F_{34} \xrightarrow{p} \frac{4(1-3B)B\delta^2+\sigma_\epsilon^2}{4[5(1-3B)B\delta^2+\sigma_\epsilon^2]},$$

onde $B = (1 - \lambda)\lambda$. Note-se, então, que a magnitude do parâmetro de quebra (δ) produz em t_1 um efeito diferente do que se observa em t_2 e em F_{34} : enquanto na primeira ele está presente apenas no denominador do valor limite, nas segundas δ figura em ambos os termos da fracção, como base de potências de expoentes



Figura 2: Um exemplo do caso 4 ($\delta > 0$).

idênticos. Esta diferença sugere que o comportamento de potência dos testes HEGY em pequenas amostras não deverá ser idêntico para as várias estatísticas. Enquanto que a de t_1 se deverá reduzir de forma significativa com o crescimento de δ , não se prevê que o mesmo possa ocorrer com a dos testes de raízes unitárias sazonais. Embora o teste da raiz unitária na frequência nula se mantenha consistente, o valor limite de 1.a) sugere que quebras de magnitude crescente exigirão amostras de dimensão também crescente para que o desempenho de potência seja preservado. Para os testes de raízes unitárias sazonais essa condição parece ser dispensável. Naturalmente, esta divergência não pode ser considerada surpreendente pois o ciclo sazonal não é afectado.

Por outro lado, para o caso 4 obtém-se:

$$4.a) \quad T^{-1/2} t_1 \xrightarrow{p} -\sqrt{\frac{B\delta^2 + \sigma_\epsilon^2}{6B\delta^2 + 4\sigma_\epsilon^2}},$$

$$4.b) \quad T^{-1/2} t_2 \xrightarrow{p} -\sqrt{\frac{B\delta^2 + \sigma_\epsilon^2}{6B\delta^2 + 4\sigma_\epsilon^2}},$$

$$4.c) \quad T^{-1} F_{34} \xrightarrow{p} \frac{\sigma_\epsilon^2}{6B\delta^2 + 4\sigma_\epsilon^2}.$$

Desta forma, o que se observou no caso 1 para t_1 observa-se agora (apenas) para F_{34} . Pelo contrário, os limites de 4.a) e 4.b) não surgem muito afectados pelo valor de δ . Enquanto que para t_1 este resultado dificilmente pode ser considerado

surpreendente, o mesmo não parece acontecer com t_2 , pois o ciclo sazonal foi alterado. Todavia, este *puzzle* é apenas aparente. De facto, note-se que o efeito de uma quebra que altera simetricamente as médias sazonais de dois trimestres separados por outro é semelhante ao de um grande choque que afecta o ciclo anual, como no processo $(1 + L^2)y_t = \epsilon_t$, que contém as raízes complexas. Por outro lado, essa quebra não afecta os ciclos semi-anuais, correspondentes à raiz -1 , testada com a estatística t_2 .

A utilização da representação trigonométrica da sazonalidade determinística é útil para confirmar esta intuição. Considerando, sem perda de generalidade, que $\sum_{j=1}^4 \alpha_j = 0$, os ciclos sazonais anteriores e posteriores à quebra são dados, respectivamente, por

$$sd_{1t} = \frac{\alpha_4 - \alpha_2}{2} \cos\left(\frac{\pi t}{2}\right) + \frac{\alpha_1 - \alpha_3}{2} \sin\left(\frac{\pi t}{2}\right) + \frac{\sum_{j=1}^4 (-\alpha_j)^j}{4} \cos(\pi t), \quad e$$

$$sd_{2t} = \frac{\alpha_4 - \alpha_2}{2} \cos\left(\frac{\pi t}{2}\right) + \left(\delta + \frac{\alpha_1 - \alpha_3}{2}\right) \sin\left(\frac{\pi t}{2}\right) + \frac{\sum_{j=1}^4 (-\alpha_j)^j}{4} \cos(\pi t).$$

Em resumo, neste caso espera-se que só a potência do teste das raízes complexas (associadas à frequência $\pi/2$) seja substancialmente afectada em pequenas amostras.

4.3 Resultados de simulação

Efectuou-se um estudo de simulação com o intuito de averiguar em que medida os resultados assintóticos são úteis para explicar e prever o comportamento dos testes HEGY em pequenas amostras. Continuando a considerar apenas os exemplos dos casos 1 e 4, apresentam-se, no Quadro 1, as estimativas de potência dos testes com 5% de dimensão, considerando $\epsilon_t \sim nid(0, 1)$, $\lambda = 0.5$, i.e., quebra localizada a meio da amostra, $T = 48, 96$ e 160 , e baseados em 10000 réplicas. Para o parâmetro de quebra tomou-se $\delta = \delta/\sigma_\epsilon = 0$ (ausência de quebra), 1, 3 e 5.

Em ambos os casos (bem como naqueles não reproduzidos aqui), os resultados assintóticos são bem reproduzidos em pequenas amostras:

- a) no caso 1, embora assintoticamente robusto, t_1 tem um desempenho que é claramente afectado pela dimensão da quebra, sobretudo quando esta é grande. Pelo contrário, como previsto, a potência dos testes de raízes unitárias sazonais não é afectada por quebras, nem sequer em pequenas amostras.

Quadro 1. Estimativas de potência dos testes HEGY para os casos 1 e 4

δ	$t_1(T)$			$t_2(T)$			$F_{34}(T)$		
	48	96	160	48	96	160	48	96	160
0	0.685	0.997	1.000	0.869	1.000	1.000	0.976	1.000	1.000
	caso 1: $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta$								
1	0.505	0.963	1.000	0.855	1.000	1.000	0.976	1.000	1.000
3	0.054	0.206	0.693	0.853	1.000	1.000	0.989	1.000	1.000
5	0.002	0.002	0.015	0.888	1.000	1.000	0.999	1.000	1.000
	caso 4: $\delta_1 = -\delta_3 = \delta$, $\delta_2 = \delta_4 = 0$								
1	0.605	0.995	1.000	0.827	1.000	1.000	0.841	0.999	1.000
3	0.457	0.980	1.000	0.752	1.000	1.000	0.022	0.374	0.979
5	0.521	0.987	1.000	0.815	1.000	1.000	0.000	0.000	0.018

b) No caso 4, é no teste sobre as raízes complexas (com F_{34}) que se manifesta a tensão entre as dimensões da quebra e da amostra, com a potência a reduzir-se substancialmente à medida que a primeira cresce (com T fixo) e a aumentar à medida que a segunda aumenta (com δ fixo). Pelo contrário, como esperado, o desempenho das estatísticas t_1 e t_2 mantém-se praticamente inalterado.

Finalmente, um estudo adicional (não apresentado aqui) mostra que a aproximação dada pelos valores limite aos valores médios das estatísticas em amostras finitas é de muito boa qualidade, sobretudo para t_2 e F_{34} .

5 Testes de DF e Sazonalidade Determinística

Como foi referido anteriormente, os testes de DF devem recorrer a dados não dessazonalizados. Neste contexto, como a análise da sua robustez à presença de sazonalidade estocástica não estacionária foi efectuada por Ghysels et al. (1994), Rodrigues e Osborn (1999) e Rodrigues (2000), Lopes (2006) investiga as suas propriedades quando as séries contêm sazonalidade determinística, sobretudo quando esta é negligenciada. Como é sabido, este problema insere-se no da similaridade dos testes de raízes unitárias, no que respeita, agora, aos parâmetros do ciclo sazonal.

Para este efeito, o enquadramento dos testes HEGY parece mais adequado, pois a necessidade de tomar em conta a sazonalidade determinística surge claramente da sua potencial presença nos valores iniciais do processo (recorde-se a equação (4)). Pelo contrário, quando, como no caso dos testes DF, o ponto de partida básico é o de uma equação às diferenças de primeira ordem, o problema da invariância relativamente ao ciclo sazonal parece não existir. Ele parece colocar-se apenas em relação ao valor inicial e ao parâmetro de deriva. Talvez seja este facto que justifique (parcialmente) que alguns investigadores omitam as *dummies* sazonais das suas regressões (A)DF, o mesmo não acontecendo quando efectuam as regressões dos testes HEGY.

Por outro lado, também é sabido que a qualidade da estimação da ordem da autoregressão — i.e., do parâmetro de truncatura dos desfasamentos — pode afectar, até de forma dramática, a inferência sobre a presença de raízes unitárias. Assim, procura-se também investigar o comportamento dos testes (A)DF em pequenas amostras quando, para esse efeito, é empregue o procedimento *t-sig*. Embora existam outros métodos, o método *t-sig* pode ser considerado o mais popular, pois tem dominado nas aplicações empíricas desde a publicação do trabalho de Campbell e Perron (1991).

5.1 A (não) similaridade dos testes de DF

Num artigo muito esclarecedor, Dickey et al. (1986) forneceram uma resposta clara, negativa, à questão: “a remoção das médias sazonais afecta a(s) distribuição(ões) limite da(s) estatística(s) DF?”. Todavia, o seu resultado parece ter sido mal interpretado por alguns, que parecem tê-lo visto como uma declaração de indiferença sobre a inclusão das *dummies* sazonais nas regressões de teste. Para melhor esclarecer o problema, é útil inverter (e qualificar) a questão inicial: quando a série contém sazonalidade determinística, a não remoção das médias sazonais afecta a(s) distribuição(ões) da(s) estatística(s) DF? A resposta é bem conhecida — “sim, certamente!” — porque, como é sabido, a regressão deve incluir pelo menos todas as componentes determinísticas presentes no PGD; veja-se, por exemplo, Campbell e Perron (1991), Banerjee et al. (1993), pp. 104-5, ou Kiviet e Phillips (1992). Coloca-se então, logicamente, a questão adicional: de que forma são afectadas as distribuições?

Para obter a resposta, considere-se o PGD simples, onde um único parâmetro

(δ) governa o ciclo sazonal,

$$\Delta y_t = \sum_{j=1}^4 (-1)^j \delta D_{jt} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2), \quad t = 1, \dots, T, \quad (8)$$

e, sem perda de generalidade, se assume que a amostra inclui todas as observações de um certo número de anos. Então, embora produzindo um teste similar relativamente a y_0 , a estatística $\text{DF}_{c(nd)} \equiv \tau_{c(nd)}$, obtida da estimação OLS da equação

$$\Delta y_t = \alpha + \phi y_{t-1} + v_t, \quad (9)$$

onde $\text{DF}_{c(nd)} = \hat{\phi}/\hat{\sigma}_{\hat{\phi}}$, não fornece um teste invariante ao(s) parâmetro(s) perturbador(es) do ciclo sazonal. Para esse efeito, é necessário acrescentar as *dummies* sazonais a (9) e estimar a regressão

$$\Delta y_t = \sum_{j=1}^4 \alpha_j D_{jt} + \psi y_{t-1} + \omega_t, \quad (10)$$

cuja estatística $\text{DF}_{sd} \equiv \tau_{sd} (= \hat{\psi}/\hat{\sigma}_{\hat{\psi}})$ é também invariante em relação a y_0 . Por outro lado, é sabido que a invariância relativamente à eventual presença de deriva em (8) requer a inclusão de um termo de tendência linear em (10).

Se o PGD é dado por (8), com $y_0 = 0$, mas o teste utiliza a equação (9), com a constante omitida e desprezando a sazonalidade determinística, tem-se

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

onde \Rightarrow representa convergência fraca em distribuição e $K_s = \delta/\sigma_\varepsilon$, isto é, K_s representa a sazonalidade estandardizada, e $W(r)$ representa um processo de Wiener *standard* definido em $[0, 1]$. Este resultado decorre imediatamente como corolário de uma proposição demonstrada por Franses e Haldrup (1994) para as regressões DF com séries temporais contaminadas por *outliers* aditivos (AOs), pois a equação (8) pode escrever-se sob essa forma, com os picos e vales sazonais a desempenharem o papel de *outliers*⁶. Por outro lado, se for empregue a equação (9) (incluindo a

⁶Trata-se de uma analogia não rigorosa mas que não invalida o resultado (pelo contrário, simplifica a sua demonstração). Uma abordagem alternativa, que generaliza a que aqui se apresenta, é fornecida por Demetrescu e Hassler (2004).

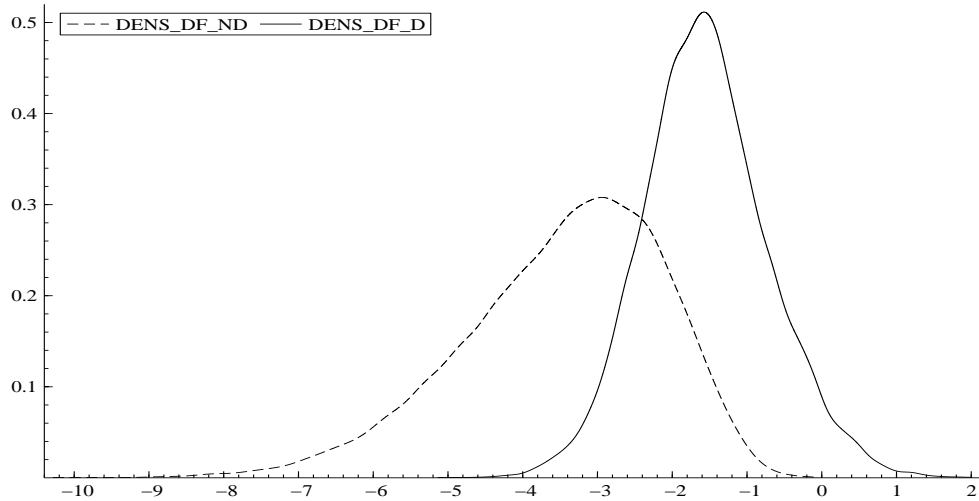


Figura 3: Distribuições de DF_{sd} (DF, à direita) e de $DF_{c(nd)}$ com $K_s = 2$ (funções densidade de probabilidade estimadas com 20000 réplicas de amostras com $T = 800$).

constante), o processo de Wiener *standard* é substituído por um processo Browniano com a média removida. Se, adicionalmente, se incluir um termo de tendência linear, passa a ter-se um movimento Browniano com a tendência removida.

Por conseguinte, deste resultado deduz-se que:

- a) quando $\delta = 0$, a distribuição limite é, naturalmente, a de DF;
- b) quando $\delta \neq 0$, a distribuição limite passa a depender de K_s e desloca-se para a esquerda à medida que este aumenta.

Um estudo de simulação mostra que este deslocamento é perceptível mesmo quando K_s assume apenas o valor 0.1 e $T = 80$. Mostra também que o resultado obtido fornece uma boa aproximação a processos gerais, com padrões sazonais mais heterogêneos que o assumido, do tipo $\Delta y_t = \sum_{j=1}^4 \delta_j D_{tj} + \epsilon_t$, $\sum_{j=1}^4 \delta_j = 0$, com $\delta \approx (1/4) \sum_{j=1}^4 |\delta_j|$. A figura 3 compara as distribuições de DF_{sd} e de $DF_{c(nd)}$ quando $K_s = 2$ e, além do deslocamento para a esquerda, a segunda também apresenta maior dispersão que a usual distribuição de DF.

5.2 Implicações para o trabalho empírico

O resultado anterior tem implicações bem nítidas sobre as propriedades dos testes DF. Uma vez que a distribuição se desloca para a esquerda, negligenciar a presença de sazonalidade determinística conduzirá a rejeições espúrias da hipótese de raiz unitária, que podem ser muito frequentes quando a sua importância relativa é grande. Intuitivamente, esta implicação era esperada, pois um ciclo sazonal não considerado conduz à percepção errônea que todas as observações contêm “choques” puramente transitórios, isto é, que a série apresenta um comportamento espúrio de reversão para a média. Quando esses “choques” são grandes relativamente ao desvio padrão dos verdadeiros (e persistentes) choques, a frequência das rejeições incorretas da hipótese nula também tenderá a ser elevada. Para o caso simples do processo da equação (8), quando $T = 80$ e é usada a equação (9), o quadro 2 revela que o problema pode ser muito grave.

Quadro 2. Estimativas da dimensão real dos testes de DF com dimensão nominal de 5% para $T = 80$

K_s	0.1	1.0	5.0
dim. est.	5.12%	20.45%	99.13%

Todavia, o processo de (8) é muito simples e é raramente considerado no trabalho empírico. De forma mais realista, considere-se que o PGD é dado por

$$\psi(L)\Delta y_t = \mu + \sum_{j=1}^4 (-1)^j \delta D_{jt} + \theta(L)\varepsilon_t,$$

onde $\psi(L)$ pode conter raízes unitárias sazonais. Também de forma mais realista, assumam-se que, em vez da equação (9), a inferência se baseia em regressões do tipo

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \sum_{i=1}^k \lambda_i \Delta y_{t-i} + \epsilon_{nd,t},$$

onde k representa o parâmetro de truncatura de defasamentos, estimado com base no método t -sig (do geral para o particular), isto é, efectuando testes de significância recursivos sobre os parâmetros λ_i , com dimensão aproximada de 5%. A respectiva estatística (A)DF, $\tau_{ct(nd)} = \hat{\phi}/\hat{\sigma}_{\hat{\phi}}$, é representada com $ADF_{ct(nd)}$.

Para estudar o comportamento do teste em pequenas amostras, considerou-se $T = 48, 80$ e 160 , e $k_{max} = 4, 8$ e 12 , respectivamente. Alguns resultados selecionados de um estudo de simulação, baseado em 10000 réplicas, são apresentados no quadro 3 (veja-se Lopes, 2006, para mais pormenores e, em particular, para as justificações dos PGDs considerados).

Quadro 3. Frequências de rejeição de $ADF_{ct(nd)}$ e $ADF_{sd,t}$ ao nível nominal de 5%

T (k_{max})	48(4)		80(8)		160(12)	
δ	$ADF_{ct(nd)}$	$ADF_{sd,t}$	$ADF_{ct(nd)}$	$ADF_{sd,t}$	$ADF_{ct(nd)}$	$ADF_{sd,t}$
PGD: $\Delta y_t = \mu + \sum_{j=1}^4 (-1)^j \delta D_{jt} + \epsilon_t - 0.8\epsilon_{t-1}$						
0	0.882	0.860	0.707	0.683	0.465	0.446
1	0.692	0.860	0.456	0.683	0.210	0.446
5	0.501	0.860	0.280	0.683	0.151	0.446
PGD: $\Delta y_t = \mu + \sum_{j=1}^4 (-1)^j \delta D_{jt} + \epsilon_t - 0.8\epsilon_{t-4}$						
0	0.323	0.300	0.261	0.240	0.274	0.257
1	0.614	0.300	0.456	0.240	0.474	0.257
5	0.676	0.300	0.455	0.240	0.469	0.257
PGD: $(1 + 0.9L)(1 + 0.4L^2)\Delta y_t = \mu + \sum_{j=1}^4 (-1)^j \delta D_{jt} + \epsilon_t$						
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

O panorama geral para o teste $ADF_{ct(nd)}$ é de distorções de dimensão generalizadas, que podem ser muito graves no caso dos processos com componentes MA negativas. Todavia, também se pode observar que, relativamente aos resultados do quadro 2, o método t -sig atenua substancialmente o problema, sobretudo quando se permite que k_{max} cresça com T , o que produz estimativas mais elevadas de k . De novo, a analogia com o modelo AO é útil para explicar este facto: os “outliers” sazonais induzem nos erros da regressão uma componente adicional do tipo MA negativo, que o método t -sig tenta capturar através de autoregressões mais longas (veja-se Franses e Haldrup, 1994, e Perron e Rodríguez, 2003). Ou seja, desfasamentos adicionais são induzidos para substituir o efeito das *dummies* sazonais omitidas. No

entanto, mesmo quando $T = 160$ (e $k_{max} = 12$) e o PGD já contém uma componente MA, observam-se ainda proporções muito elevadas de rejeições espúrias.

O quadro 3 contém ainda estimativas da dimensão real do teste $ADF_{sd,t}$ ($\tau_{sd,t} = \hat{\psi}/\hat{\sigma}_{\hat{\psi}}$), baseado na regressão similar

$$\Delta y_t = \sum_{j=1}^4 \alpha_j D_{jt} + \beta^* t + \psi y_{t-1} + \sum_{i=1}^{k^*} \gamma_i \Delta y_{t-i} + \epsilon_{sd,t},$$

que permitem avaliar o desempenho do método t -sig em circunstâncias normais. Confirmando e generalizando resultados anteriores de Taylor (2000), pode observar-se que, em pequenas amostras, esse desempenho é muito desapontador para processos com componentes MA negativos. Embora consistente, o método é afectado adversamente pela inclusão de regressores determinísticos nas equações de teste, produzindo estimadores dos parâmetros de desfasamento enviesados para zero. Daqui resultam estruturas de desfasamento sub-parametrizadas e, conseqüentemente, distorções de dimensão que podem ser substanciais. É este facto que explica que, em pequenas amostras, e como o quadro 3 ilustra, o teste similar não se revele muito superior ao não similar em termos de dimensão.

Por outro lado, também a potência dos testes $ADF_{sd,t}$ e $ADF_{ct(nd)}$ deve ser comparada, tornando-se necessária corrigi-la da dimensão em ambos os casos. Nos dois casos, a correcção deve-se à utilização do método t -sig mas, obviamente, no de $ADF_{ct(nd)}$, também à alteração da distribuição resultante da omissão das *dummies* sazonais. A este nível, a superioridade do teste $ADF_{sd,t}$ é esmagadora e deve-se, em grande medida, à “inflação” na estimação de k provocada pela referida omissão. No entanto, a relevância empírica deste resultado é limitada pelo facto de, na prática, não serem empregues valores críticos corrigidos.

5.3 Ilustração Empírica

Para ilustrar empiricamente esta análise, considere-se um exemplo que recorre a um conjunto de séries temporais portuguesas (veja-se o quadro 4). A metodologia empregue é bastante simples: a) quando as estimativas do parâmetro de truncatura de desfasamentos (\hat{k} e \hat{k}^*) são idênticas, dá-se preferência à utilização do valor p associado à estatística $ADF_{sd,t}$; b) no caso contrário, e em particular quando $\hat{k} > \hat{k}^*$, recorre-se a uma investigação mais profunda, utilizando os instrumentos

da modelação ARIMA, para tentar confirmar a suspeita que a primeira se encontra “inflacionada”.

Quadro 4. Resultados empíricos para algumas séries temporais económicas portuguesas

	amostra	k_{max}	$ADF_{sd[t]}(\hat{k})$ [val.- p]	$ADF_{c[t](nd)}(\hat{k}^*)$ [val.- p]
PIB	77:1–98:4	8	−2.02 (1) [0.59]	−3.29 (6) [0.06]
Cons. Priv.	77:1–98:4	8	−2.49 (4) [0.35]	−2.67 (4) [0.25]
Cons. Púb.	77:1–98:4	8	−2.20 (5) [0.50]	−2.22 (5) [0.51]
FBCF	77:1–98:4	8	−2.16 (8) [0.52]	−2.23 (8) [0.48]
Exportações	60:1–98:4	12	−2.92 (7) [0.18]	−2.92 (7) [0.15]
Importações	60:1–98:4	12	−2.89 (12) [0.19]	−2.94 (12) [0.13]
Inflação	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]

Notas: 1) todas as séries, excepto a da taxa de inflação, foram previamente logaritmizadas; 2) os valores- p das estatísticas $ADF_{sd[t]}$ foram estimados por simulação, com base em 50000 réplicas; 3) os valores- p das estatísticas $ADF_{c[t](nd)}$ foram estimados com base numa rotina incorporada no TSP 4.5.

Para a maioria das séries, o suporte à hipótese de raiz unitária é aproximadamente o mesmo, quer se considere ou não a presença de sazonalidade determinística. Todavia, para três delas observam-se discrepâncias importantes. As séries para os consumos privado e público, FBCF, exportações, importações e inflação, constituem o primeiro grupo. Nota-se, no entanto, uma ligeira tendência para o teste adequado proporcionar mais apoio à hipótese nula. Também é de notar a diferença nos comprimentos de desfasamento estimados para a série da inflação (presumindo-se que k esteja sobre-estimado no caso de $ADF_{c(nd)}$).

As restantes três séries parecem ilustrar a análise anterior, i.e., parecem fornecer exemplos de rejeições espúrias pelo teste $ADF_{ct(nd)}$. Como $\hat{k} = \hat{k}^*$, o caso da série do IPI de electricidade é o mais simples, pois seguir a regra rígida da dimensão de 5% conduz à inversão da decisão. Os restantes dois casos são ainda mais interessantes: desprezar a sazonalidade determinística nas séries do PIB e do IPI–total reduz drasticamente o suporte à hipótese de raiz unitária, e uma análise mais detalhada mostra que, pelo menos aparentemente, as estatísticas $ADF_{ct(nd)}$ estão a favorecer de forma

incorrecta a hipótese de estacionaridade em tendência. Com efeito, considerando devidamente os termos determinísticos, uma análise baseada na metodologia de Box & Jenkins suporta as autoregressões mais curtas, associadas às estatísticas $ADF_{sd,t}$. O escasso apoio dado pela regressão não similar à presença de uma raiz unitária parece ser espúrio.

Em resumo, pode concluir-se que a percepção relativamente comum, com origem na ideia da decomposição das séries temporais em componentes independentes, de que a sazonalidade determinística nada tem a ver com os testes sobre as suas propriedades de longo prazo, é incorrecta.

6 Referências Bibliográficas

- Amado, C. A. de O. (2000). *Modelização da sazonalidade: alguma evidência empírica para a economia Portuguesa*, Dissertação de mestrado em Economia, Faculdade de Economia da Universidade do Porto.
- Balcombe, K. (1999). Seasonal unit root tests with structural breaks in deterministic seasonality, *Oxford Bulletin of Economics and Statistics*, 61, 569-82.
- Banerjee, A., Dolado, J., Galbraith, J. W. e Hendry, D. F. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press.
- Burke, S. P. (1996). Some reparameterizations of lag polynomials for dynamic analysis, *Oxford Bulletin of Economics and Statistics*, 58, 373-89.
- Campbell, J. Y. e Perron, P. (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots, *NBER Macroeconomics Annual*, 141-219.
- Davidson, J. e MacKinnon, R. (2004). *Econometric Theory and Methods*, Oxford University Press.
- Demetrescu, M. e Hassler, U. (2004). Effect of neglected deterministic seasonality on unit root tests, documento de trabalho, Goethe-University Frankfurt.

- Dickey, D. A. e Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-31.
- Dickey, D. A., Bell, W. R. e Miller, R. B. (1986). Unit roots in time series models: tests and implications, *American Statistician*, 40, 12-26.
- Engle, R. F., Granger, C. W. J., Hylleberg, S. e Lee, H. S. (1993). Seasonal cointegration: the Japanese consumption function, *Journal of Econometrics*, 55, 275-298.
- Franses, P.H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*, Oxford University Press.
- Franses, P. H. e Haldrup, N. (1994). The effects of additive outliers on tests for unit roots and cointegration, *Journal of Business and Economic Statistics*, 12, 471-8.
- Franses, P. H., Hylleberg, S. e Lee, H. S. (1995). Spurious deterministic seasonality, *Economics Letters*, 48, 249-56.
- Ghysels, E. (1994). On the economics and econometrics of seasonality, in Sims, C. A., ed., *Advances in Econometrics, Sixth World Congress*, vol. I, 257-322, Cambridge University Press.
- Ghysels, E., Lee, H. S., e Noh, J. (1994). Testing for unit roots in seasonal time series, *Journal of Econometrics*, 62, 415-442.
- Ghysels, E., Lee, H. S. e Siklos, P. (1993). On the (mis)specification of seasonality and its consequences: an empirical investigation with U.S. data, *Empirical Economics*, 18, 747-760.
- Ghysels, E. e Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series*, Cambridge University Press.
- Ghysels, E., Osborn, D. e Rodrigues, P. M. M. (2006). Forecasting seasonal time series, *mimeo*, a publicar in Elliot, G., Granger, C. W. J. e Timmermann, A., eds., *Handbook of Economic Forecasting*, Elsevier.

- Ghysels, E. e Perron, P. (1993). The effect of seasonal adjustment filters on tests for a unit root, *Journal of Econometrics*, 55, 57-98.
- Ghysels, E., Osborn, D. R. e Rodrigues, P. M. M. (2001). Seasonal nonstationarity and near-nonstationarity, in Baltagi, B. H., ed., *A Companion to Theoretical Econometrics*, Blackwell Publishing.
- Hylleberg, S. (1992). *Modelling Seasonality*, Oxford University Press.
- Hylleberg, S., Engle, R. F., Granger, C. W. J. e Yoo, B. S. (1990). Seasonal integration and cointegration, *Journal of Econometrics*, 49, 215-238.
- Hylleberg, S., Jorgensen, C. e Sorensen, N. K. (1993). Seasonality in macroeconomic time series, *Empirical Economics*, 18, 321-35.
- Kiviet, J. F. e Phillips, G. D. A. (1992). Exact similar tests for unit roots and cointegration, *Oxford Bulletin of Economics and Statistics*, 54, 349-67.
- Lopes, A. C. B. da S. e Veiga, M. H. (1999). Sazonalidade estocástica não estacionária: alguma evidência empírica para a economia portuguesa, *Economia*, XXIII, 3-25.
- Lopes, A. C. B. da S. (1999). Spurious deterministic seasonality and autocorrelation corrections with quarterly data: further Monte Carlo results, *Empirical Economics*, 24, 341-59.
- Lopes, A. C. B. da S. (2003). The order of integration for quarterly macroeconomic time series: a simple testing strategy, *Empirical Economics*, 28, 783-94.
- Lopes, A. C. B. da S. e Montañés, A. (2005). The behavior of HEGY tests for quarterly time series with seasonal mean shifts, *Econometric Reviews*, 24, 83-108.
- Lopes, A. C. B. da S. (2006). Deterministic seasonality in Dickey-Fuller tests: should we care?, *Empirical Economics*, 31, 165-82.
- Miron, J. A. (1994). The economics of seasonal cycles, in Sims, C. A. (ed.), *Advances in Econometrics*, Sixth World Congress, vol. I, 213-255, Cambridge University Press.

- Nerlove, M., Grether, D. M. e Carvalho, J. L. (1995). *Analysis of Economic Time Series, a Synthesis* (revised ed.), Academic Press.
- Osborn, D. R. (1990). A survey of seasonality in UK macroeconomic variables, *International Journal of Forecasting*, 6, 327-36.
- Perron, P. (1989). The great crash, the oil shock and the unit root hypothesis, *Econometrica*, 57, 1361-402.
- Perron, P. e Rodríguez, G. (2003). Searching for additive outliers in nonstationary time series, *Journal of Time Series Analysis*, 24, 193-220.
- Rodrigues, P. M. M. (2000). A note on the application of the DF test to seasonal data, *Statistics and Probability Letters*, 47, 171-5.
- Rodrigues, P. M. M. e Osborn, D. R. (1999). Performance of seasonal unit root tests for monthly data, *Journal of Applied Statistics*, 26, 985-1004.
- Smith, J. e Otero, J. (1997). Structural breaks and seasonal integration, *Economics Letters*, 56, 13-19.
- 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, 293-304.
- Wallis, K. F. (1974). Seasonal adjustment and relations between variables, *Journal of the American Statistical Association*, 69, 18-31, reimpresso in Hylleberg, S. (1992), 27-59.