

"A FEA e a USP respeitam os direitos autorais deste trabalho. Nós acreditamos que a melhor proteção contra o uso ilegítimo deste texto é a publicação online. Além de preservar o conteúdo motiva-nos oferecer à sociedade o conhecimento produzido no âmbito da universidade pública e dar publicidade ao esforço do pesquisador. Entretanto, caso não seja do interesse do autor manter o documento online, pedimos compreensão em relação à iniciativa e o contato pelo e-mail [bibfea@usp.br](mailto:bibfea@usp.br) para que possamos tomar as providências cabíveis (remoção da tese ou dissertação da BDTD)."

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE  
DEPARTAMENTO DE ECONOMIA

MUDANÇAS NO PADRÃO SAZONAL:  
UMA COMPARAÇÃO ENTRE O DESEMPENHO DOS MODELOS  
AUTO-REGRESSIVO E ESTRUTURAL

CARLOS MASSAYUKI ASSATO

ORIENTADOR: PROF. DR. PAULO PICCHETTI

SÃO PAULO

2003

Reitor da Universidade de São Paulo  
Prof. Dr. Adolpho José Melfi

Diretor da Faculdade de Economia, Administração e  
Contabilidade  
Prof. Dra. Maria Teresa Leme Fleury

Chefe do Departamento de Economia  
Prof. Dra. Elisabeth Maria Mercier Querido Farina

DEDALUS - Acervo - FEA



20600025662

MUDANÇAS NO PADRÃO SAZONAL:  
UMA COMPARAÇÃO ENTRE O DESEMPENHO DOS MODELOS  
AUTO-REGRESSIVO E ESTRUTURAL

ALUNO: CARLOS MASSAYUKI ASSATO

ORIENTADOR: PROF. DR. PAULO PICCHETTI

Dissertação apresentada ao  
Departamento de Economia da Faculdade  
de Economia, Administração e  
Contabilidade da Universidade de São  
Paulo para a obtenção do título de  
Mestre em Economia

SÃO PAULO

2003

T330.015195 A844m

T85526



20600025662



Powered by MidProStar - www.logprocess.com.br

## FICHA CATALOGRÁFICA

Elaborada pela Seção de Publicações e Divulgação do SBD/FEA/USP

Assato, Carlos Massayuki

Mudanças no padrão sazonal : uma comparação entre o desempenho dos modelos auto-regressivo e estrutural / Carlos Massayuki Assato. -- São Paulo : FEA/USP, 2003.

56 f.

Dissertação (Mestrado) – Universidade de São Paulo, 2003  
Bibliografia.

1. Econometria 2. Sazonalidade 3. Análise de regressão e de correlação I. Faculdade de Economia, Administração e Contabilidade da USP II. Título.

CDD – 330.015195

## AGRADECIMENTOS

- Ao Prof. Paulo Picchetti pelas aulas ministradas durante o curso desde a fase preliminar até a conclusão desta dissertação e na defesa da genialidade de Duke Ellington e Charlie Parker.
- À FINE e CAPES pelo apoio financeiro.
- Ao Departamento de Economia e aos professores que ministraram as matérias durante o Mestrado.
- Aos Profs. Vera Fava (curso de Econometria II que foi muito importante nesta dissertação) e Denisard Alves na qualificação deste trabalho.
- Ao Prof. Flávio Saes pelos melhores cursos que tive durante a graduação.
- Aos colegas do curso de Mestrado, especialmente Guilherme, Alexandre, Léo, Flávio e Adriano.
- Aos amigos Leandro Loyola, Fábio Negrão, Fábio Rosas, Reinaldo Prado, Paulo Mazieri, Flávio Hourneaux Jr., Adriana e Rogério pelo incentivo e apoio durante todo o curso.
- A José Carlos Souza Santos, professor do Departamento de Economia e amigo da Linear DTVM pelo apoio e ajuda desde a preparação para ANPEC.
- Ao amigo Marcos Mollica pelo incentivo desde a primeira conversa sobre o Mestrado no IPE, ênfase nos estudos econômicos e pela amizade desde a Linear DTVM.
- Por fim, em memória de meus pais, aos meus irmãos Edson e Thiemi e em especial à Carmem pelo amor, dedicação e paciência.

## RESUMO

Este trabalho propõe realizar uma comparação entre os Modelos Auto-Regressivo e Estrutural para explicar séries econômicas que apresentam um padrão sazonal que se altera lentamente com o decorrer do tempo. De acordo com a especificação do Processo Gerador de Dados utilizado neste trabalho, os resultados obtidos confirmam que o Modelo Estrutural se adapta melhor aos dados quanto menor a magnitude dos choques sazonais vis-à-vis os choques de interesse primário. Além disso, apresentamos um estudo sobre a importância da inclusão do efeito de feriados móveis em séries econômicas brasileiras.

## ABSTRACT

This dissertation intends to compare Autoregressive and Structural Models in explaining economic time series that have slowly changing seasonal patterns. Based on a specified Data Generating Process, results confirm that Structural Models are better than its rival as smaller the seasonal disturbances comparing with prime interest disturbances. Moreover, we present a study about the importance of holiday effects in Brazilian economic time series.

## SUMÁRIO

1. APRESENTAÇÃO .....	1
2. DESCRIÇÃO DOS MODELOS .....	5
2.1 - BSM .....	5
2.2 - Modelo AR .....	7
3. EXPERIMENTO DE MONTE CARLO .....	9
3.1 - DGP .....	9
3.2 - Estimaco do BSM .....	12
3.3 - Modelo AR - $\beta = 1$ .....	12
3.4 - Modelo AR - $\beta < 1$ .....	15
3.5 - Estatísticas de Teste .....	17
4. RESULTADOS .....	21
4.1 - Especificaco com raiz unitária sazonal .	21
4.2 - Especificaco com $\beta < 1$ .....	24
4.2.1 - Amostras com dimenses diferentes .	24
4.2.2 - Alteraco do parâmetro sazonal ....	28
5. PARTE EMPÍRICA .....	29
6. EFEITO FERIADO .....	36
6.1 - Modelo geral .....	36
6.2 - Consultas ao crédito .....	38
6.3 - Efeito diário constante .....	40
CONSIDERAÇÕES FINAIS .....	43
BIBLIOGRAFIA .....	45
APÊNDICE - ALGORITMOS .....	47



## 1. APRESENTAÇÃO

O problema da sazonalidade em séries de tempo tem sido objeto de pesquisa em vários trabalhos econométricos nos últimos anos.

Esta dissertação tem por objetivo confrontar um dos pontos de discussão desta área: a comparação entre os Modelos Estrutural e Auto-Regressivo para modelar séries de tempo com componente sazonal.

Antes, porém, é necessário apresentar de forma breve as duas principais áreas de estudos sobre sazonalidade, que estão ligadas de forma indireta à este trabalho.

A primeira área concerne aos chamados programas de ajustamento sazonal, ou seja, à utilização de procedimentos para "limpar" séries econômicas dos efeitos sazonais. Dentre os vários programas, podemos citar o TRAMO/SEATS<sup>1</sup> e o X-12<sup>2</sup> do Census Bureau norte-americano. Este último é o programa mais utilizado por muitos economistas.

A segunda área refere-se à construção de testes para a presença de raízes unitárias sazonais. O principal estudo nesta área é o trabalho de Hylleberg et al. (1990) que é denominado na literatura econométrica de HEGY. Este teste é uma generalização do teste de raiz unitária de Dickey-Fuller e, como este último, é baseado nas defasagens da série em estudo. O grande mérito deste trabalho é que ele permite o teste em todas as frequências sazonais e, inclusive, na frequência zero.

No entanto, estas duas áreas apresentam vários pontos de controvérsia entre os estudiosos e não há um consenso entre eles.

---

<sup>1</sup> TRAMO (Time series Regression with ARIMA noise, Missing observations and Outliers) e SEATS (Signal Extraction in ARIMA Time Series), para detalhes ver Ghysels e Osborn (2001), pág. 108-114.

<sup>2</sup> Para maiores detalhes sobre este programa, Findley et al. (1998).

Uma crítica com relação à utilização dos programas de ajustamento sazonal ocorre na utilização de uma estrutura ARIMA não invertível. "Se um programa de ajustamento sazonal padrão (...) for utilizado, além dos problemas de não-invertibilidade, é provável que o ajuste seja sub-ótimo e sérias distorções ocorram"<sup>3</sup>.

Para os testes para a presença de raízes unitárias sazonais, pelo fato destes testes utilizarem a mesma estrutura do teste de raiz unitária de Dickey-Fuller aumentado, eles apresentam problemas com suas propriedades estatísticas. Por exemplo, uma destas críticas, mostra que "uma aproximação auto-regressiva pode levar a erros de especificação nos testes de raízes unitárias quando o parâmetro MA for representativo"<sup>4</sup>.

É impressionante notar que ao realizar um estudo econométrico observamos uma diferença de performance entre vários métodos com resultados, em algumas ocasiões, conflitantes.

Tome, por exemplo, a discussão apresentada em Taylor e Leybourne (1999). Neste estudo, os autores comparam a estatística de teste proposta por eles, denominada de  $c_{i,T}$  ( $i$  = frequência a ser analisada e  $T$  = tamanho da amostra) com o teste HEGY. Eles concluem que: "em geral, os testes HEGY são mais poderosos que  $c_{i,T}$  (...), embora este seja mais poderoso para pequenos desvios da raiz unitária de interesse quando  $T$  for pequeno". Além disso, os autores observam que "o tamanho das estatísticas de teste HEGY tendem a aumentar com o número de defasagens, enquanto que para  $c_{i,T}$  ocorre o inverso"<sup>5</sup>.

---

<sup>3</sup> Harvey e Scott (1994), pág. 1332. Todas as citações em inglês deste trabalho estão traduzidas.

<sup>4</sup> Schwert (1989), pág. 154.

<sup>5</sup> Taylor e Leybourne (1999), pág. 277.

Diante desta breve discussão, nosso objetivo é analisar um outro ponto de discussão na literatura sobre sazonalidade proposto por Harvey e Scott (1994).

Neste estudo, eles analisam as séries de Consumo e Renda para a Grã-Bretanha entre 1958-92. Tais séries têm um padrão sazonal que se altera lentamente com o decorrer do tempo.

Segundo estes autores, para séries de tempo com esta característica sazonal o Modelo Estrutural, tal como proposto por Harvey (1989), apresenta melhor performance que o Modelo Auto-Regressivo, descrito em detalhes, por exemplo, em Hamilton (1994)<sup>6</sup>.

A argumentação é que o Modelo Estrutural tem sucesso em separar os choques sazonais daqueles de "interesse primário" (ambos com persistência no longo prazo), sendo que os primeiros são de menor magnitude que os últimos.

A proposta principal desta dissertação é realizar um estudo econométrico em séries univariadas que apresentam estas características através de uma simulação de dados (através de experimento de Monte Carlo) e a utilização de estatísticas de teste para verificar a performance dos Modelos Estrutural e Auto-Regressivo.

A estrutura desta dissertação é compreendida em seis capítulos.

No capítulo 2, descrevemos as características básicas dos Modelos Estrutural e Auto-Regressivo, que são amplamente utilizados na literatura de séries de tempo.

---

<sup>6</sup> Veja Hamilton (1994), cap. 3, págs. 43-71.

No capítulo 3, apresentamos com detalhes a metodologia para a realização do experimento de Monte Carlo. Para facilitar a exposição, dividimos este capítulo em três partes. A primeira, descreve a estrutura do Processo Gerador de Dados. A segunda parte refere-se à construção dos algoritmos dos Modelos Estrutural e Auto-Regressivo. Finalmente, verificamos a adequação de cada um destes modelos com a elaboração de estatísticas de teste.

Os resultados e a discussão do experimento de Monte Carlo são apresentados no capítulo 4.

No capítulo 5, realizamos um teste empírico com uma série econômica brasileira que apresenta um padrão sazonal com as características aqui descritas.

Finalmente, no último capítulo como um estudo complementar, introduzimos a discussão do efeito feriado, proposto por Bell e Hillmer (1983) que parece ser muito importante para a série brasileira que utilizamos no capítulo 5.

Após estes capítulos, apresentamos as considerações finais, a bibliografia e no apêndice, os algoritmos descritos no capítulo 3.

## 2. DESCRIÇÃO DOS MODELOS

### 2.1 - Modelo Estrutural Básico (BSM)<sup>7</sup>

O modelo de componentes não observados apresenta os termos de uma série a ser analisada -  $y_t$  - de forma aditiva. Permite-se que tais termos sejam estocásticos, ou seja, podem variar ao longo do tempo.

Como um caso particular deste modelo, utilizaremos o Modelo Estrutural Básico - BSM, que pode ser descrito da seguinte forma:

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

onde:

$\mu_t$  - tendência

$\gamma_t$  - componente sazonal

$\varepsilon_t$  - choque  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

A tendência,  $\mu_t$ , é desmembrada da seguinte forma:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \zeta_t$$

onde:

$\beta_t$  - declividade da tendência

$\eta_t$  - choque na tendência  $\eta_t \sim N(0, \sigma_\eta^2)$

$\zeta_t$  - choque na declividade  $\zeta_t \sim N(0, \sigma_\zeta^2)$

A representação da tendência estocástica utilizada acima é o caso mais geral. Note que a tendência se torna determinista quando  $\sigma_\eta^2$  e  $\sigma_\zeta^2$  forem nulas.

---

<sup>7</sup> Neste trabalho, utilizaremos as siglas que se tornaram padrão na literatura econométrica.

Com relação ao componente sazonal, o BSM permite duas formulações básicas (por dummies ou na forma trigonométrica). Neste trabalho, utilizaremos a sazonalidade trigonométrica, que foi a forma utilizada em Harvey e Scott (1994).

A sazonalidade trigonométrica é apresentada da seguinte forma:

$$\gamma_t = \sum_{j=1}^{\frac{s}{2}} \gamma_{j,t}$$

$$\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \gamma_{j,t-1}^* \sin \lambda_j + \omega_{j,t}$$

$$\gamma_{j,t}^* = -\gamma_{j,t-1} \sin \lambda_j + \gamma_{j,t-1}^* \cos \lambda_j + \omega_{j,t}^*$$

para  $j = 1, \dots, \frac{s}{2}$  e  $\lambda_j = 2\pi \frac{j}{s}$

onde:

$s$  = número de períodos na série  $y_t$

$\omega_t \sim N(0, \sigma^2_\omega)$  e  $\omega_{j,t}^* \sim N(0, \sigma^2_\omega)$

Os componentes  $\varepsilon_t$ ,  $\eta_t$ ,  $\zeta_t$ ,  $\omega_{j,t}$  e  $\omega_{j,t}^*$  são não correlacionados entre si. Esta é uma das principais hipóteses do BSM. Note que as variâncias dos componentes irregulares na especificação sazonal,  $\omega_{j,t}$  e  $\omega_{j,t}^*$ , são iguais. Harvey (1989) argumenta que esta restrição impõe um custo baixo para a adequação do modelo aos dados e não há desperdício de tempo em termos computacionais.

Finalmente, como apontado acima com a tendência, o componente sazonal é determinista se sua variância for nula.

## 2.2 - Modelo Auto-Regressivo (AR)

Para a descrição deste modelo, utilizamos a notação apresentada em Hamilton (1994)<sup>8</sup>. Seguiremos também sua forma de apresentação.

Um processo  $\varepsilon_t$  é um ruído branco se:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \text{ e } E(\varepsilon_t \varepsilon_s) = 0 \text{ para todo } t \neq s$$

Uma série  $y_t$  é fracamente estacionária se:

$$E(y_t) = \mu \text{ para todo } t$$

$$E(y_t - \mu)(y_{t-j} - \mu) = \gamma_j \text{ para todo } t \text{ e qualquer } j$$

A partir destas duas representações, considere dois processos para a série  $y_t$ .

Primeiramente, um processo Auto-Regressivo - AR - de p-ésima ordem:

$$y_t = \mu + (\phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p) y_t + \varepsilon_t$$

onde:

L - operador defasagem

Caso as raízes do polinômio  $(\phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p)$  estejam fora do círculo unitário,  $y_t$  possui uma representação AR estacionária de p-ésima ordem.

Em segundo lugar, considere agora que a série  $y_t$  possa ser representada por um processo de média móvel (MA) de q-ésima ordem:

$$y_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t$$

---

<sup>8</sup> Veja nota 6.

Caso as raízes do polinômio  $(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)$  estejam fora do círculo unitário, podemos escrever este processo MA como um processo AR infinito invertendo o operador MA da seguinte forma:

$$(1 + \eta_1 L + \eta_2 L^2 + \dots) (y_t - \mu) = \varepsilon_t$$

onde:  $(1 + \eta_1 L + \eta_2 L^2 + \dots) = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)^{-1}$

Esta é a chamada representação MA invertível.

Consideramos neste trabalho que as séries geradas são invertíveis e tenham uma representação AR.

Finalmente, a partir de uma série  $y_t$  para se encontrar a especificação adequada, realizam-se as etapas de identificação (funções de auto-correlação e auto-correlação parcial), estimação e diagnóstico (adequação do modelo escolhido aos dados).



### 3. EXPERIMENTO DE MONTE CARLO

Para realizarmos este experimento devido ao grande número de replicações e à necessidade de trabalharmos com matrizes de grande dimensão, utilizamos a linguagem de programação Ox e o "pacote estatístico" SSFPack 2.2, que transforma as matrizes para a forma de espaço-estado, detalhado em Koopman et al. (1999).

Os algoritmos completos estão descritos no apêndice deste trabalho. A seguir, apresentamos as três etapas de programação.

#### 3.1 - Processo Gerador de Dados (DGP)

No estudo de Harvey e Scott (1994) são utilizados dados trimestrais. Neste trabalho, utilizamos séries mensais devido à maior disponibilidade deste tipo de periodização para as séries econômicas brasileiras.

Os componentes do DGP, no domínio do tempo, são descritos em duas equações. A primeira, a equação principal, apresenta-se da seguinte forma:

$$y_t = y_{t-1} + s_t + \varepsilon_t \quad (1)$$

onde:

$y_t$  - valor da série em  $t$

$s_t$  - componente sazonal em  $t$

$\varepsilon_t$  - choque aleatório em  $t$  -  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  -  $E(\varepsilon_t \varepsilon_{t-j}) = 0$  para qualquer  $j$

Note que, propositalmente utilizamos uma estrutura com raiz unitária. Isto deve-se a duas razões.

A primeira diz respeito à discussão recorrente em econometria de séries de tempo onde os processos com raiz unitárias parecem ser a regra,

principalmente em séries macroeconômicas. No entanto, ocorrem críticas com relação aos testes de raízes unitárias quando o DGP apresenta um processo de "quase" raiz unitária.

Neste caso, Campbell e Perron (1991)<sup>9</sup> mostram que "a restrição de raiz unitária pode ajudar a aumentar a eficiência das estimativas (isto é, reduzir o erro quadrático médio) mesmo que as variáveis do modelo não tenham uma raiz unitária mas sejam quase integradas".

A segunda razão é que após a diferenciação de (1), podemos caracterizar a sazonalidade de forma mais clara e verificar sua mudança com o decorrer do tempo.

A segunda equação, referente ao componente sazonal -  $s_t$  - é construída como segue:

$$s_t = \beta s_{t-12} + \eta_t \quad (2)$$

onde:

$\beta$  - parâmetro relacionado ao termo sazonal

$\eta_t$  - choque aleatório em  $t$  -  $\eta_t \sim N(0, \sigma_\eta^2)$  e  $E(\eta_t \eta_{t-j}) = 0$  para qualquer  $j$

Podemos reescrever a equação (2) de outra forma:

$$s_{i,j} = \beta s_{i,j-1} + \eta_{i,j} \quad (2)^*$$

onde:

$s_{i,j}$  - componente sazonal no mês  $i$  e ano  $j$

Nas simulações do capítulo 4, utilizamos diferentes valores para  $\beta$ , primeiramente para  $\beta = 1$ , que é a fonte de discussão de Harvey e Scott

---

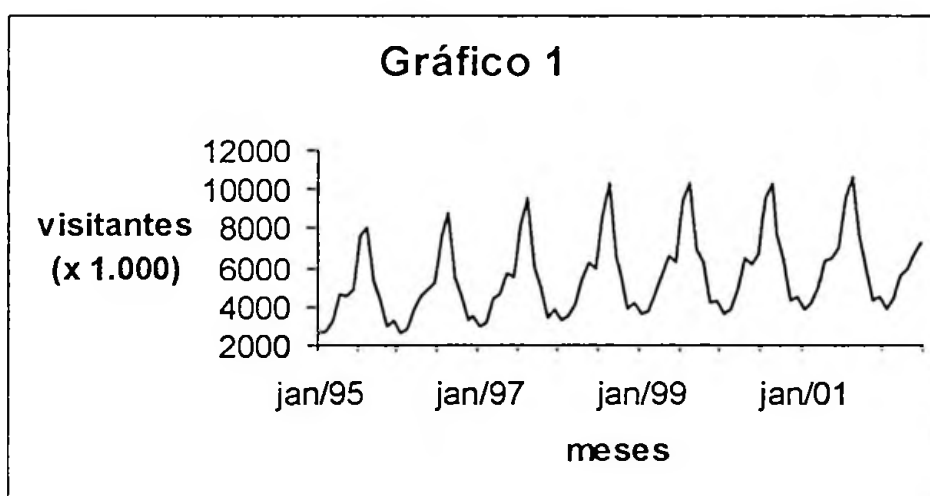
<sup>9</sup> Campbell e Perron (1991), pág. 22 e 23. Os autores fizeram uma revisão dos testes de raízes unitárias e cointegração no começo da década de noventa. Há vários resultados interessantes, mas que não cabem no escopo deste trabalho.

(1994). Além disso, realizamos um estudo para valores de  $\beta$  próximos à raiz unitária, que também caracterizam uma longa permanência dos choques.

Por hipótese, os choques de (1) e (2) são não correlacionados em qualquer período.

Para a construção das séries, utilizamos valores iniciais para  $y_0 = 100$  e os doze primeiros valores do termo sazonal agrupados em um vetor  $s_0$  - de  $s_{1,0}$  a  $s_{12,0}$  - notação de (2)\*.

Para os valores iniciais do vetor  $s_0$  foram escolhidos os parâmetros de uma série com um componente sazonal bem definido - a quantidade mensal total de visitantes na Espanha. Utilizamos os dados disponíveis no sítio do Instituto Nacional de Estatística da Espanha (INE), que compreende os anos entre 1995 a 2002, como apresentado abaixo.



Para  $\varepsilon_t$  e  $\eta_t$ , foram gerados números aleatórios com uma Distribuição Normal Padronizada. Com estas matrizes, realizamos os passos seguintes para a construção de  $y_t$ .

Além disso, descartamos as primeiras vinte e quatro observações de cada série para evitar o problema de valores iniciais como apontado em Schwert (1989). As doze últimas observações foram separadas para os

testes da capacidade de previsão dos dois modelos como é descrito seção 3.5, abaixo.

### 3.2 - Estimação do BSM

Para a estimação do BSM foi construído um algoritmo onde a série  $y_t$  é representada na forma de espaço-estado para todos os componentes (tendência, declividade e sazonalidade) como apresentado na seção 2.1.

Para cada série foram estimados 4 hiperparâmetros ( $\sigma^2_\varepsilon$ ,  $\sigma^2_\eta$ ,  $\sigma^2_\zeta$  e  $\sigma^2_{\omega,j}$  para  $j = 1, 2, \dots, 11$ ). A estimação é feita pelo método da máxima verossimilhança como discutido em detalhes em Harvey (1989). Como apresentado ao final do seção da 2.1 consideramos a mesma variância para todo  $j$  no termo sazonal.

### 3.3 - Modelo AR com Raiz unitária sazonal - $\beta = 1$

Considere  $\Delta$  e  $\Delta_{12}$  como a primeira e décima segunda diferenças respectivamente, a esperança -  $E(\cdot)$  -, as  $j$ -ésimas autocovariâncias -  $\gamma(j)$  - e as  $j$ -ésimas autocorrelações -  $\rho(j)$ .

Retomando as equações (1) e (2), com  $\beta = 1$ , temos:

$$y_t = y_{t-1} + s_t + \varepsilon_t \quad (1)$$

$$\Delta y_t = s_t + \varepsilon_t$$

$$s_t = s_{t-12} + \eta_t \quad (2)$$

Utilizando Harvey (1989)<sup>10</sup>, "com um modelo linear, os componentes caracterizados por diferentes choques, podem ser combinados em um modelo com um choque único, chamado de forma reduzida":

$$y_t^* = \Delta \Delta_{12} y_t = \eta_t + \Delta_{12} \varepsilon_t$$

Seguindo:

$$E(y_t^*) = 0$$

$$E(y_t^*)^2 = \sigma_\eta^2 + 2 \sigma_\varepsilon^2$$

$$\gamma(1) = 0$$

(...)

$$\gamma(12) = -\sigma_\varepsilon^2$$

$$\gamma(j) = 0 \text{ para todo } j > 12$$

$$\rho(12) = -\frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + 2 \sigma_\varepsilon^2} \quad (3)$$

Agora considere um processo  $x_t$  - MA(1)<sub>12</sub>:

$$x_t = \omega_t + \theta \omega_{t-12} \quad \omega_t \sim N(0, \sigma_\omega^2)$$

$$E(x_t) = 0$$

$$E(x_t)^2 = \sigma_\omega^2 + \theta^2 \sigma_\omega^2 = \sigma_\omega^2 (1 + \theta^2)$$

$$\gamma(1) = 0$$

(...)

$$\gamma(12) = \theta \sigma_\omega^2$$

$$\gamma(j) = 0 \text{ para todo } j > 12$$

$$\rho(12) = \frac{\theta}{1 + \theta^2} \quad (4)$$

---

<sup>10</sup> Veja a discussão das págs. 51 a 68. A citação está na página 67.

Em suma, a especificação das equações (1) e (2) com  $\beta=1$  é a mesma de um processo MA(1)<sub>12</sub>. Igualando (3) e (4) e chamando  $q = \frac{\sigma_{\eta}^2}{\sigma_{\epsilon}^2}$  de "razão do sinal dos choques"<sup>11</sup>:

$$\frac{\theta}{1+\theta^2} = -\frac{\sigma_{\epsilon}^2}{\sigma_{\eta}^2 + 2\sigma_{\epsilon}^2}$$

$$\frac{\theta}{1+\theta^2} = -\frac{1}{q+2} \quad (5)$$

Resolvendo (5) para  $\theta$ , o parâmetro do processo MA(1)<sub>12</sub>, temos:

$$\theta = \frac{-(2+q) \pm \sqrt{q^2 + 4q}}{2} \quad (6)$$

No entanto, em (6) descarta-se o sinal negativo, pois o mesmo implica em um processo não-invertível.

Com o sinal positivo,  $0 \leq q \leq \infty$  corresponde a  $-1 \leq \theta \leq 0$ . Note que para alguns valores de  $q$ , temos:

$q$	$\theta$
0.20	-0.64
0.10	-0.73
0.01	-0.90

Nos resultados do próximo capítulo consideramos os valores da tabela acima para a comparação entre os Modelos Auto-Regressivo e Estrutural.

Veja que a "razão do sinal dos choques" definirá a magnitude do coeficiente do processo MA.

Aqui está o núcleo da crítica de Harvey e Scott (1994) sobre a inadequação do Modelo AR quando os choques sazonais têm magnitude

<sup>11</sup> Harvey (1989), pág. 68 apresenta toda a argumentação que segue.

inferior aos choques da equação principal (ou de interesse primário como eles definem).

Além disso, quando  $\sigma_{\eta}^2 = 0$  (componente determinista na equação sazonal) temos um processo MA estritamente não-invertível.

### 3.4 - Modelo Auto-Regressivo para $\beta < 1$

Retomando as mesmas notações que utilizamos na seção anterior para o caso da raiz unitária sazonal e utilizando o método iterativo, a partir de (2)\*, podemos encontrar uma equação geral para o componente sazonal.  $s_{i,j}$ . Lembre-se que:

$i = \text{mês}, j = \text{ano}$

$$y_t = y_{t-1} + s_t + \varepsilon_t \quad (1)$$

$$s_{i,j} = \beta s_{i,j-1} + \eta_{i,j} \quad (2)^*$$

$$s_{1,1} = \beta s_{1,0} + \eta_{1,1}$$

$$s_{2,1} = \beta s_{2,0} + \eta_{2,1}$$

(...)

$$s_{12,1} = \beta s_{12,0} + \eta_{12,1}$$

$$s_{1,2} = \beta s_{1,1} + \eta_{1,2} = \beta (\beta s_{1,0} + \eta_{1,1}) + \eta_{1,2} = \beta^2 s_{1,0} + \beta \eta_{1,1} + \eta_{1,2}$$

(...)

$$s_{1,3} = \beta s_{1,2} + \eta_{1,3} = \beta (\beta^2 s_{1,0} + \beta \eta_{1,1} + \eta_{1,2}) + \eta_{1,3} =$$

$$= \beta^3 s_{1,0} + \beta^2 \eta_{1,1} + \beta \eta_{1,2} + \eta_{1,3}$$

(...)

$$s_{i,j} = \beta^j s_{i,0} + \sum_{k=0}^{j-1} \beta^k \eta_{i,j-k} \quad (7)$$

De (7) e como  $\beta < 1$ , por hipótese, temos:

$$\lim_{j \rightarrow \infty} s_{i,j} = \sum_{k=0}^{\infty} \beta^k \eta_{i,j-k} \quad (8)$$

A partir de (1) e (8) temos:

$$y_t = y_{t-1} + s_t + \varepsilon_t$$

$$y_t = y_{t-1} + \sum_{k=0}^{\infty} \beta^k \eta_{1,j-k} + \varepsilon_t$$

$$\Delta y_t = \sum_{k=0}^{\infty} \beta^k \eta_{1,j-k} + \varepsilon_t$$

$$E(\Delta y_t) = 0$$

$$\gamma(0) = E(\Delta y_t)^2 = \sigma_{\eta}^2 \sum_{k=0}^{\infty} \beta^{2k} + \sigma_{\varepsilon}^2$$

$$\gamma(j) = E[(\Delta y_t) (\Delta y_{t-j})] = 0 \text{ para todo } j \neq 12, 24, 36, \dots$$

$$\gamma(12) = E[(\Delta y_t) (\Delta y_{t-12})] = \sigma_{\eta}^2 (\beta + \beta^3 + \beta^5 + \dots)$$

$$\gamma(24) = E[(\Delta y_t) (\Delta y_{t-24})] = \sigma_{\eta}^2 (\beta^2 + \beta^4 + \beta^6 + \dots)$$

(...)

$$\rho(j) = 0 \text{ para todo } j \neq 12, 24, 36, \dots$$

$$\rho(12) = \frac{\sigma_{\eta}^2 (\beta + \beta^3 + \beta^5 + \dots)}{\sigma_{\eta}^2 \sum_{k=0}^{\infty} \beta^{2k} + \sigma_{\varepsilon}^2}$$

$$\rho(24) = \frac{\sigma_{\eta}^2 (\beta^2 + \beta^4 + \beta^6 + \dots)}{\sigma_{\eta}^2 \sum_{k=0}^{\infty} \beta^{2k} + \sigma_{\varepsilon}^2}$$

(...)

Portanto, a partir das autocorrelações, nota-se que há "picos" decrescentes há cada doze meses. Após diferenciarmos  $y_t$ , escolhemos um modelo AR que contemple apenas os períodos sazonais.

Com relação à estimação dos modelos AR (tanto para  $\beta = 1$  como  $\beta < 1$ ), dado que temos a restrição sobre os parâmetros intermediários (nulos)



entre os períodos sazonais, utilizamos o método de mínimos quadrados ordinários. Isto deve-se ao fato de que caso optássemos pelo método da máxima verossimilhança o esforço computacional seria imenso, o que tornaria o estudo inviável.

### 3.5 - Estatísticas de teste

Para os dados dentro da amostra, a partir dos parâmetros calculados nos dois modelos obtém-se a série estimada -  $y_{est}$  - seguindo o "método de um passo à frente":

$$y_{est}(l) = E(y_{t+l} | y_{t+l-1})$$

Para cada um dos modelos, há pontos iniciais distintos (diferentes valores para  $l$ ). No BSM, o  $y_{est}$  é calculado a partir da observação número 15 (hiperparâmetros do erro, tendência, declividade e onze para a sazonalidade). No modelo  $AR(3)_{12}$  por exemplo, após a primeira e décima-segunda diferenciação, o ponto inicial situa-se na 50ª observação.

Isto posto, construímos a variância do erro de predição (vep) da seguinte forma:

$$r_t = y_t - y_{est}$$

$$vep = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2$$

onde:

$r_t$  - erro entre as observações "reais" e os valores estimados em  $t$

$\bar{r}$  - média dos erros

$T$  - tamanho da série

Para os dados fora da amostra, ou seja, no tocante à previsão dos modelos, dividimos as estatísticas em dois conjuntos. Primeiramente, apresentamos as estatísticas "tradicionais":

- Erro quadrático médio (EQM)

$$\text{EQM} = \left[ \frac{1}{J} \sum_{j=1}^J e^2(j) \right]^{\frac{1}{2}}$$

- Erro absoluto médio (EAM)

$$\text{EAM} = \frac{1}{J} \sum_{j=1}^J |e(j)|$$

- Erro percentual absoluto médio (EPAM)

$$\text{EPAM} = \frac{1}{J} \sum_{j=1}^J \left| \frac{e(j)}{y_{t,j}} \right|$$

onde:

$$y_{\text{prev}}(j) = E(y_{t+j} | y_{t+j-1})$$

$$e(j) = \hat{y}_t - y_{\text{prev}}$$

j = período a ser estimado fora da amostra - j = 1, 2, ..., 12

O segundo conjunto de estatísticas segue de Theil (1965). Este autor apresenta diversas estatísticas para previsão e a principal delas ficou conhecida na literatura econométrica como Coeficiente de Desigualdade de Theil (1965) - U:

$$U = \left( \frac{\sum_{j=1}^J e^2(j)}{\sum_{j=1}^J y_{t+j}^2} \right)^{\frac{1}{2}}$$

onde:

$y_{t+j}$  = valor observado em j

Note que  $U$  é a raiz quadrada da razão entre o quadrado do EQM e a média do quadrado dos valores "reais".

Mas o ponto mais interessante no livro de Theil, é o desmembramento do numerador de  $U^2$  em três termos:

$$\frac{1}{J} \sum_{j=1}^J (y_{t+j} - y_{\text{prev em } j})^2 = (\underline{y}_{\text{prev}} - \underline{y}_t)^2 + (s_{\text{prev}} - s_y)^2 + 2(1 - r) s_{\text{prev}} s_y \quad (9)$$

onde:

$\underline{y}_{\text{prev}}$  - média de  $y_{\text{prev}}$

$\underline{y}_t$  - média de  $y_{t+j}$

$s_{\text{prev}}$  - desvio padrão de  $y_{\text{prev}}$

$s_y$  - desvio padrão de  $y_{t+j}$

$r$  - coeficiente de correlação entre os  $y_{\text{prev}}$  e  $y_{t+j}$

Ao dividir cada um dos três componentes (9) pela soma, temos o "Primeiro Conjunto de Proporções de Desigualdade":

- Proporção do viés -  $U^M = (\underline{y}_{\text{prev}} - \underline{y}_t)^2 / \text{EQM}^2$

- Proporção da variância -  $U^S = (s_{\text{prev}} - s_y)^2 / \text{EQM}^2$

- Proporção da covariância -  $U^C = [2(1 - r) s_{\text{prev}} s_y] / \text{EQM}^2$

A partir destes componentes, Theil argumenta que "(...) tal decomposição é útil pois coloca uma luz sobre a natureza dos erros de previsão" <sup>12</sup>.

Além da expressão (9), Theil apresenta o "Segundo Conjunto de Proporções de Desigualdade", que também é obtido a partir de  $U^2$ :

$$\frac{1}{J} \sum_{j=1}^J (y_{t+j} - y_{\text{prev em } j})^2 = (\underline{y}_{\text{prev}} - \underline{y}_t)^2 + (s_{\text{prev}} - r s_y)^2 + (1 - r^2) s_y^2 \quad (10)$$

Procedendo da mesma forma que (9), temos:

---

<sup>12</sup> Theil (1965), pág. 32.

- Proporção da Regressão -  $U^R = (s_{\text{prev}} - r s_y)^2 / EQM^2$

- Proporção dos Ruídos -  $U^D = (1 - r^2) s_y^2 / EQM^2$

O primeiro termo de (10) é o mesmo que em (9) -  $U^M$ . Obviamente, como no caso anterior a soma dos termos  $U^M$ ,  $U^R$  e  $U^D$  é igual a unidade.

Antes de finalizarmos esta seção, é necessário apresentar uma crítica aos dois Conjuntos de Desigualdade de Theil feita por Granger e Newbold (1977)<sup>13</sup>.

Considerando um modelo AR(1) com um amostra tendendo ao infinito, os autores mostram que a única restrição do "Primeiro Conjunto de Desigualdade" é que  $U^s \geq 0$  e  $U^c \leq 1$ . Ou seja, não é possível tirar conclusões sobre a magnitude de tais estatísticas.

Diante desta crítica, no capítulo seguinte, serão apresentadas todas as estatísticas de teste aqui discutidas, porém a atenção estará voltada para a vep e as estatísticas tradicionais.

---

<sup>13</sup> Granger e Newbold (1977), pág. 288.

## 4. RESULTADOS

Este capítulo é dividido em duas partes para DGPs distintos. A primeira parte refere-se à especificação com raiz unitária sazonal e a segunda diz respeito à análise da sensibilidade do parâmetro sazonal ( $\beta < 1$ ) e diferentes tamanhos de amostra.

### 4.1 - Especificação com raiz unitária sazonal

Na seção 3.3, mostramos como Harvey e Scott (1994) a partir das equações (1) e (2):

$$y_t = y_{t-1} + s_t + \varepsilon_t \quad (1)$$

$$s_t = \beta s_{t-12} + \eta_t \quad (2)$$

argumentam que no caso da raiz unitária sazonal,  $y_t$  transforma-se em um processo  $(0,1,0) \times (0,1,1)_{12}$  e que o parâmetro  $\theta$  de  $MA(1)_{12}$  tem um relação direta com  $q$ , a "razão do sinal dos choques",  $\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$ :

$$\theta = \frac{-(2+q) \pm \sqrt{q^2 + 4q}}{2} \quad (6)$$

Abaixo apresentamos os experimentos de Monte Carlo com diferentes valores para  $q$ . Utilizamos amostras com 600 observações, que correspondem a cinquenta anos. As primeiras vinte e quatro observações foram descartadas como mencionado no final da seção 3.1. Encontramos a convergência das estatísticas com 25.000 replicações.

No caso do Modelo AR, após realizar a transformação dos vetores  $y_t$  na primeira e décima-segunda diferença -  $y_t^*$  - utilizamos a especificação  $AR(3)_{12}$ .

A tabela 1 mostra o percentual em que as estatísticas do Modelo AR apresenta valores absolutos maiores que o BSM, ou seja, a proporção dentre todas as replicações em que o Modelo AR apresentou uma performance inferior sobre o BSM<sup>14</sup>.

Tabela 1			
Estatísticas de teste - AR(3) <sub>12</sub>			
q ->	0.20	0.10	0.01
vep	87.65%	99.07%	99.99%
EQM	53.38%	57.95%	62.09%
EAM	52.82%	57.33%	61.70%
EPAM	52.88%	57.36%	61.74%
U	53.38%	57.95%	62.09%
U <sup>M</sup>	49.80%	49.42%	49.88%
U <sup>S</sup>	49.05%	49.48%	50.04%
U <sup>C</sup>	50.93%	50.50%	50.34%
U <sup>R</sup>	49.08%	49.59%	49.94%
U <sup>D</sup>	50.92%	50.54%	50.34%

Na primeira linha, notamos um resultado interessante com relação à convergência da variância do erro de predição. A performance do BSM dentro da amostra, mostra que este supera seu rival rapidamente à medida em que  $q$  decresce.

Para a capacidade de previsão, nas estatísticas "tradicionais" e no Coeficiente de Desigualdade de Theil, U, ocorre o mesmo fato porém de forma mais lenta.

Em seguida, utilizamos uma outra especificação para o Modelo AR no intuito de verificar a importância do número de defasagens nos resultados das estatísticas. Na tabela 2 utilizamos a especificação AR(4)<sub>12</sub>.

<sup>14</sup> No algoritmo para  $q=0.01$ , consideramos a declividade da tendência como determinista. Colocamos esta restrição devido à economia de tempo computacional.

Tabela 2			
Estatísticas de teste - AR(4) <sub>12</sub>			
q ->	0.20	0.10	0.01
vep	57.97%	93.62%	99.60%
EQM	51.55%	56.30%	60.69%
EAM	51.17%	55.70%	60.20%
EPAM	51.17%	55.69%	60.20%
U	51.55%	56.30%	60.69%
U <sup>M</sup>	49.41%	49.46%	50.23%
U <sup>S</sup>	49.34%	49.83%	50.08%
U <sup>C</sup>	50.93%	50.20%	49.72%
U <sup>R</sup>	49.22%	49.82%	50.17%
U <sup>D</sup>	50.92%	50.22%	49.76%

Novamente, na variância do erro de previsão o BSM apresenta valores menores que Modelo AR. Nas estatísticas "tradicionais" e no Coeficiente de Desigualdade de Theil, U, a diferença é menor, porém ainda em favor do BSM.

As duas tabelas acima confirmam duas assertivas apresentadas por Harvey e Scott.

Primeiro, é que quanto menor o valor de q, ou seja, menor a razão entre a magnitude do choque na equação sazonal e do choque na equação principal a performance do BSM cresce vis-à-vis seu rival. Ou colocado de outra forma, à medida em que o parâmetro  $\theta$  do processo MA se aproxima da região de não-invertibilidade, o BSM apresenta melhores resultados que o Modelo AR.

O "principal fato estilizado da sazonalidade é que ela se desenvolve lentamente de uma forma não estacionária. Um modelo que exhibe estas

características parece se adequar à maioria das séries econômicas quando incluído como um componente em um Modelo Estrutural"<sup>15</sup>.

Em segundo lugar, "a habilidade (das representações auto-regressivas) para lidar com padrões sazonais que se alteram lentamente é limitada. Defasagens maiores parecem ser necessárias e a dinâmica resultante parece confundir os efeitos sazonais com as respostas dinâmicas de interesse primário "<sup>16</sup>

#### 4.2 - Especificação com $\beta < 1$

Ao confirmarmos a proposição dos autores no caso da raiz unitária sazonal, estendemos nosso estudo para verificar como os dois modelos se comportam com especificações próximas à raiz unitária sazonal. Realizamos o estudo em duas vias distintas. A primeira com  $\beta = 0.99$  e diferentes dimensões de amostra e a segunda com diferentes valores para  $\beta$  e amostra constante.

Relembrando a discussão ao final da seção 3.4 em que apresentamos um correlograma decrescente nos períodos sazonais após a primeira diferença em  $y_t$ , utilizamos uma especificação com três defasagens - Modelo AR -  $(0,1,0) \times (3,0,0)_{12}$ .

##### 4.2.1 - Amostras com dimensões diferentes

Realizamos três simulações. A primeira delas foi feita com 108 observações (correspondente a nove anos). As demais com 180 e 600

---

<sup>15</sup> Harvey e Scott (1994), pág. 1342.

<sup>16</sup> op. cit, pág. 1324-1325.



observações como no item anterior. Além disso, utilizamos o parâmetro sazonal,  $\beta = 0.99$ , propositalmente para haver uma "quase" raiz unitária sazonal e que os choques tivessem um efeito prolongado nas séries. Os resultados estão na Tabela 3.

Estatísticas de Teste $s_t = 0.99 s_{t-1} + \eta_t$			
amostra ->	108	180	600
Vep	1.90%	0.00%	0.00%
EQM	31.59%	33.37%	40.38%
EAM	31.86%	33.18%	40.09%
EPAM	32.07%	33.35%	40.17%
U	31.59%	33.37%	40.38%
U <sup>M</sup>	50.29%	50.25%	51.34%
U <sup>S</sup>	23.61%	26.19%	35.02%
U <sup>C</sup>	68.34%	66.38%	59.86%
U <sup>R</sup>	22.64%	25.29%	33.61%
U <sup>D</sup>	68.98%	66.96%	60.92%

A primeira observação a ser feita é que o Modelo AR apresenta melhor resultado em todas as simulações com relação à estatística dentro da amostra (variância do erro de predição - vep).

Em segundo lugar, verificamos que as estatísticas "tradicionais" apontam que o Modelo AR tem melhor resultado que o BSM independentemente do tamanho da amostra, embora a diferença entre os modelos tende a cair com o aumento da amostra. Note que para a simulação com 600 observações as estatísticas estão por volta de 60% a favor do Modelo AR. O mesmo pode ser dito com relação ao coeficiente de desigualdade de Theil - U.

As proporções de desigualdade apresentaram resultados díspares. Ao final da seção 3.5, colocamos em perspectiva a crítica à interpretação destas estatísticas feita por Granger e Newbold.

Mas como um exercício econométrico, retomando as definições das proporções de desigualdade, verificamos que, obviamente, existe um *trade-off* entre os componentes:

- Proporção do viés -  $U^M = (\underline{y}_{prev} - \underline{y}_t)^2 / EQM^2$
- Proporção da variância -  $U^S = (s_{prev} - s_y)^2 / EQM^2$
- Proporção da covariância -  $U^C = [2 (1 - r) s_{prev} s_y] / EQM^2$
- Proporção da Regressão -  $U^R = (s_{prev} - r s_y)^2 / EQM^2$
- Proporção dos Ruídos -  $U^D = (1 - r^2) s_y^2 / EQM^2$

onde:  $U^S + U^C = U^R + U^D = 1 - U^M$

Na tabela 3, a proporção do viés -  $U^M$  apresenta um valor constante nos três tamanhos de amostra ao redor de 50%. Veja que  $U^M$  representa a diferença entre as médias dos valores observados e previstos.

Não ocorre diferença significativa entre os dois modelos com relação a esta estatística.  $U^M$  corresponde a cerca de 50% do Erro Quadrático Médio. Este é um tipo de erro sério, pois "(...) quem faz as previsões deve ter a habilidade de reduzir estes tipos de erros com o passar do tempo"<sup>17</sup>.

Com relação às outras proporções de desigualdade, note que todas as estatísticas ( $U^S$ ,  $U^C$ ,  $U^R$  e  $U^D$ ) utilizam o desvio padrão, o coeficiente de correlação ou ambos.

Na Tabela 4 apresentamos os resultados das médias destas duas estatísticas com relação aos dados da Tabela 3.

---

<sup>17</sup> Theil (1965), pág. 32.

Tabela 4 - Média das 25,000 repetições

n° observações	108		180		600	
estatísticas	AR	Estrut	AR	Estrut	AR	Estrut
desv pad	32.7227	34.1446	30.8227	32.0889	21.6970	22.4827
coef correl	0.99909	0.99916	0.99899	0.99902	0.99796	0.99807

Observamos duas características distintas. A primeira refere-se à queda dos desvios-padrão e dos coeficientes de correlação com o aumento do tamanho da amostra. A segunda característica é que a diferença entre as estatísticas de cada modelo com aumento da amostra também diminui. Isto se reflete nas estatísticas de teste, onde os dois modelos vão apresentando resultados mais próximos à medida que a amostra cresce.

#### 4.2.2 - Alteração do parâmetro sazonal

Retomando a equação (2)\*:

$$s_{i,j} = \beta s_{i,j-1} + \eta_{i,j} \quad (2)^*$$

Realizamos um estudo com diferentes valores para  $\beta$ , todos próximos à raiz unitária utilizando amostras com 108 observações, que corresponderiam a nove anos, período importante para os dados da economia brasileira pois refere-se ao período após a implantação do Plano Real.

Estatísticas de Teste - $s_t = \beta s_{t-1} + \eta_t$			
$\beta \rightarrow$	0.90	0.95	0.99
vep	0.00%	0.00%	1.90%
EQM	21.56%	21.87%	31.59%
EAM	22.54%	22.64%	31.86%
EPAM	22.41%	22.51%	32.07%
U	21.56%	21.87%	31.59%
U <sup>M</sup>	55.00%	51.38%	50.29%
U <sup>S</sup>	22.68%	20.55%	23.61%
U <sup>C</sup>	67.71%	71.93%	68.34%
U <sup>R</sup>	20.02%	19.02%	22.64%
U <sup>D</sup>	71.48%	73.93%	68.98%

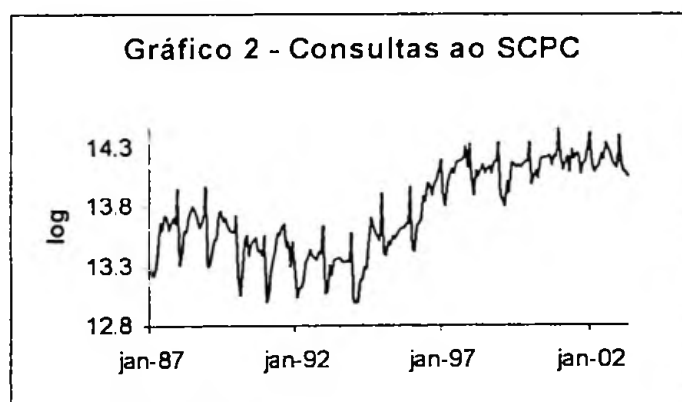
O que podemos observar nesta tabela é que mesmo quando o processo sazonal, como especificado na equação (2)\*, se aproxima da raiz unitária o modelo AR ainda apresenta uma performance bastante superior ao BSM nas estatísticas "tradicionais". A única estatística que suscita algum "equilíbrio" entre os modelos é U<sup>M</sup>, mesmo fato ocorreu na seção anterior.

Portanto, a conclusão destas duas seções complementares ao objeto principal da dissertação (a especificação de raiz unitária sazonal) é que o modelo AR (0,1,0) x (3,0,0)<sub>12</sub> apresenta resultados melhores que o BSM porém a diferença tende a cair com o aumento das amostras e com a aproximação do parâmetro  $\beta$  em direção à raiz unitária sazonal.

## 5. PARTE EMPÍRICA

Nesta seção do trabalho, utilizamos a série de consultas ao Serviço Central de Proteção ao Crédito (SCPC) fornecido pela Associação Comercial de São Paulo (ACSP). No entanto, verificamos que os dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA) contemplavam uma amostra maior que da ACSP (veja o Gráfico 2).

O período compreende o mês de janeiro de 1987 a abril de 2003. Os últimos doze meses foram retirados para verificar a capacidade de previsão dos dois modelos. Desta forma, temos 184 observações para a estimação. Os gráficos abaixo apresentam o logaritmo natural da série e o componente sazonal do BSM, cujo valor da variância é apresentado mais abaixo. Veja a mudança lenta na sazonalidade com o decorrer da série.



Discutiremos primeiramente o modelo AR. Ao analisar visualmente o gráfico, podemos suspeitar que ocorreu uma "quebra" após a decretação do Plano Real em julho de 1994.

Em virtude desta suspeita, realizamos o teste de "mudança estrutural" com a presença de uma tendência determinista ou estocástica na série. Este procedimento foi elaborado por Perron (1989) e sumarizado em Enders (1995)<sup>18</sup>. Seguiremos os passos deste último.

As hipóteses nula e alternativa do modelo mais geral do teste de Perron são:

$$H_0: y_t = \alpha + \beta_1 y_{t-1} + \mu_1 D_P + \mu_2 D_L + \varepsilon_t$$

$$A_0: y_t = \alpha + \beta_2 t + \mu_2 D_L + \mu_3 D_T + \varepsilon_t$$

onde:

$y_t$  - série do logaritmo natural do SCPC

$\alpha$  - constante

$t$  - tendência determinista

$D_P$  - dummy de "pulso" para agosto de 1994 = 1 e 0 para os demais meses

$D_L$  - dummy de "nível" para todos os meses após agosto de 1994 (inclusive) = 1 e 0 para os anteriores

$D_T$  - dummy de "tendência" para agosto de 1994 = 1, setembro de 1994 = 2,...

$\varepsilon_t$  - choque

Note a diferença das três dummies destas especificações.  $D_P$  representa um choque único no processo de raiz unitária e que altera o nível da série após julho de 1994.  $D_L$  diz respeito a mudança na constante nos

---

<sup>18</sup> Veja pp. 243-251.

dois processos e  $D_T$  representa uma mudança na declividade da tendência determinista.

Enders mostra que a estimação dos parâmetros pode ser realizada em uma só etapa com a seguinte equação:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 t + \mu_1 D_P + \mu_2 D_L + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t \quad (11)$$

onde:

$\sum_{i=1}^k \rho_i \Delta y_{t-i}$  - forma paramétrica "aumentada" para captar as autocorrelações até a ordem k

Realizamos a regressão da equação (11) utilizando o método de mínimo quadrados ordinários e os resultados das variáveis mais relevantes são apresentados na Tabela 6:

Tabela 6 - Estimação da Equação (11)

Variável	Coef	Desv Pad	Estat t
$\alpha$	1.2205	0.4798	2.5437
$\beta_1$	0.9074	0.0365	24.8750
$\beta_2$	0.0002	0.0003	0.4982
$\mu_1$	0.1964	0.0881	2.2285
$\mu_2$	0.0719	0.0368	1.9511
$\rho_1$	-0.1855	0.0846	-2.1925
$\rho_2$	-0.3556	0.0847	-4.2011
$\rho_{12}$	0.3623	0.0935	3.8734
$\rho_{24}$	0.2888	0.0747	3.8670

O valor de k que apresentou melhor resultado com base no correlograma e no Critério de Informação de Akaike foi na 24<sup>a</sup> diferença.

Para verificar a hipótese nula de raiz unitária ( $\beta_1 = 1$ ) temos:

$$(0.9074 - 1) / 0.0365 = -2.54$$

Ao consultar a tabela de Perron, para um  $\lambda = 0.5$  (pois o ponto de "mudança estrutural" em julho de 1994 está exatamente na metade da

série), verificamos o valor de -3.76 a 5% de significância. Portanto, não podemos rejeitar a hipótese nula de raiz unitária com "mudança estrutural".

Além disso, os parâmetros da tendência determinista ( $\beta_2$ ) e da dummy de "nível" ( $\mu_2$ ) não se mostraram significantes a 5%. Temos portanto:

$$y_t = \alpha + \beta_1 y_{t-1} + \mu_1 D_P + \varepsilon_t$$

Após verificar os resíduos desta equação, notamos que poderia haver a necessidade da diferenciação na frequência sazonal, pois o problema da autocorrelação não estava dirimido.

Realizamos o teste de raiz unitária sazonal tal como proposto por Dickey, Hasza e Fuller (1984) para a seguinte hipótese nula:

$$y_t^\# = \alpha y_{t-12}^\# + \sum_{i=1}^k \rho_i \Delta y_{t-12-i}^\# + \varepsilon_t$$

onde:

$y_t^\#$  = resíduo da equação de  $y_t$

Novamente, nesta equação o valor de  $k$  que apresentou melhor resultado com base no correlograma e no Critério de Informação de Akaike foi na 24ª diferença.

Para verificar a hipótese nula de raiz unitária ( $\alpha = 1$ ) temos:

$$(1.065444 - 1) / 0.109436 = 0.598$$

O valor da estatística  $t$  mencionado por Dickey, Hasza e Fuller para dados mensais com 180 observações a 5% de significância é de -2.02. Portanto não podemos rejeitar a hipótese nula de presença de raiz unitária sazonal.

Desta forma, transformamos a série original após a primeira e décima-segunda diferença,  $y_t^* = \Delta_1 \Delta_{12} y_t$ .



Após testar várias especificações, apresentamos os resultados de duas delas que pareceram ter melhor performance frente às demais:

$$y_t^* = a_2 y_{t-2}^* + a_{12} y_{t-12}^* + a_{24} y_{t-24}^* + \varepsilon_t \text{ (AR puro)}$$

$$y_t^* = \mu_1 D_P + a_1 y_{t-1}^* + a_2 y_{t-2}^* + a_{12} y_{t-12}^* + a_{24} y_{t-24}^* + \varepsilon_t \text{ (AR com dummy)}$$

Tabela 7

AR puro			AR com dummy		
variável	coefic	estat t	variavel	coef	estat t
A <sub>2</sub>	-0.2083	-2.8674	μ <sub>1</sub>	0.2526	3.1026
A <sub>12</sub>	-0.4796	-6.3489	a <sub>1</sub>	-0.1659	-2.3444
A <sub>24</sub>	-0.2914	-3.8695	a <sub>2</sub>	-0.2245	-3.1764
			a <sub>12</sub>	-0.4583	-6.2406
			a <sub>24</sub>	-0.2800	-3.8498

Com relação ao BSM, pelo fato de utilizarmos dummies no modelo AR e suspeitarmos da mudança na série após o Plano Real, apresentamos cinco modelos diferentes:

- BSM puro<sup>19</sup>;
- com intervenção no nível da tendência em agosto de 1994;
- com intervenção na declividade em agosto de 1994;
- com a dummy de "pulso" - D<sub>P</sub>;
- com a dummy de "nível" - D<sub>L</sub>.

Os valores das variâncias de cada um dos componentes do BSM dos cinco diferentes modelos estão na Tabela 8.

Tabela 8

componentes	BSM puro	intervenção no nível	intervenção na declividade	dummy de pulso	dummy de nível
irregular	9.39	12.13	9.53	10.16	12.13
nível	29.19	20.88	28.92	25.25	20.88
declividade	0	0	0	0	0
sazonal	0.05	0.06	0.05	0.06	0.06

obs: valores multiplicados por 10,000

<sup>19</sup> O Gráfico 3 da página 29 refere-se ao componente sazonal desta especificação.

Note que a variância do termo sazonal praticante não sofreu alteração nas cinco especificações. As mudanças (suaves) ocorreram nos termos irregular e do nível da tendência. A variância do componente sazonal corresponde a 0.5% do termo irregular e a 0.17% do nível da tendência no BSM puro.

Para consolidar a análise, apresentamos na Tabela 9 as estatísticas de teste das sete especificações utilizadas.

Tabela 9 - Resultado das Estatísticas de Teste

Estat	AR		BSM				
	puro	c/dummy	puro	int.nível	int.decliv.	d.pulso	d.nível
vep	0.0067	0.0062	0.0070	0.0063	0.0065	0.0059	0.0053
rmse	0.1020	0.0800	0.1077	0.0928	0.1311	0.1051	0.0928
mae	0.0921	0.0721	0.0906	0.0787	0.1143	0.0885	0.0787
mape	0.0065	0.0051	0.0064	0.0055	0.0081	0.0062	0.0055
u	0.0072	0.0056	0.0076	0.0065	0.0092	0.0074	0.0065
um	0.7041	0.5540	0.6752	0.6174	0.7610	0.6782	0.6174
us	0.0210	0.0310	0.0002	0.0000	0.0008	0.0001	0.0000
uc	0.2749	0.4150	0.3246	0.3826	0.2381	0.3217	0.3826
ur	0.0004	0.0003	0.0352	0.0314	0.0179	0.0264	0.0314
ud	0.2955	0.4457	0.2896	0.3513	0.2210	0.2954	0.3513

Na comparação dos modelos puros, o modelo AR tem uma performance ligeiramente superior ao BSM em três das cinco estatísticas que foram nosso foco de atenção no capítulo 4 - vep, erro quadrático médio e o Coeficiente de Desigualdade de Theil.

Em seguida, com as informações adicionais que colocamos no modelo, que estão refletidas nas dummies e nas intervenções, a estatística da vep mostra que as dummies melhoram o BSM significativamente e têm melhor poder de aderência aos dados que no modelo AR com dummy.

Aqui é necessário fazer uma ressalva. Na estimação da vep das especificações do BSM com as dummies de pulso e nível, o algoritmo para o

modelo estrutural só pode ser utilizado a partir da observação referente a agosto de 1994, ou seja, temos somente metade da amostra para este fim.

Isto deve-se ao fato que "uma constante (*no nosso caso, dummies com valor zero até julho de 1994*) não pode ser incluída no modelo quando  $\mu_t$  (tendência) estiver presente: isto causaria um problema similar à multicolinearidade na regressão"<sup>20</sup>.

O problema da estimação com uma constante (aqui a dummy de nível assume valores iguais a um no período de previsão) pode ser verificado também nas colunas do BSM com intervenção no nível e na declividade. A vep apresenta uma queda em comparação com o BSM puro.

No entanto para as previsões, nas estatísticas "tradicionais" e no coeficiente de desigualdade de Theil o modelo AR apresentou melhores resultados na comparação com quaisquer dos modelos BSM.

---

<sup>20</sup> Doornik et al. (1999), pág. 157.

## 6. EFEITO FERIADO

### 6.1 - Modelo geral

Nas especificações utilizadas neste trabalho que foram detalhadas no capítulo 2 - modelos AR e BSM - houve a preocupação de realizar um estudo de séries univariadas sem a presença de variáveis explicativas. Relaxamos este procedimento no capítulo anterior quando foram apresentados os modelos AR e BSM com "mudança estrutural".

Seguindo este raciocínio, um outro fator explicativo importante para o caso das séries brasileiras refere-se à variação anual de alguns feriados. Chamemos este fator de efeito feriado. Ele é diferente do efeito sazonal, pois este compreende os efeitos na série em determinado mês invariavelmente. Por exemplo, na grande maioria dos casos, as vendas no varejo apresentam os maiores valores anuais em maio e dezembro.

O efeito "feriado refere-se às flutuações na atividade econômica devido às mudanças anuais na composição do calendário"<sup>21</sup>.

Em nosso estudo, um efeito feriado de interesse é o carnaval. A terça-feira de carnaval ocorre sempre entre 3 de fevereiro (ou 4 de fevereiro em anos bissextos) e 9 de março<sup>22</sup>. Ou seja, o período do carnaval (entre o sábado e a terça-feira) pode estar restrito a cada um destes meses ou em ambos quando o feriado ocorre entre fevereiro ou o começo do mês seguinte.

---

<sup>21</sup> Bell e Hillmer (1983), pág. 527. O restante deste capítulo segue sobre o trabalho destes autores.

<sup>22</sup> Isto foi definido pelo Calendário Gregoriano em 1582. A terça-feira do carnaval ocorre 47 dias antes da Páscoa. Para maiores detalhes, consulte por exemplo, <http://gresupe.vilabol.uol.com.br/pascoa.htm>.

Para estudarmos o efeito feriado do carnaval seguimos o modelo proposto por Bell e Hillmer (1983). Considere:

$$x_t = \sum_{\tau=1}^{\tau} \bar{\alpha} h(\tau, t)$$

onde:

$x_t$  = efeito feriado em  $t$

$\bar{\alpha}$  = efeito diário durante o período do carnaval

$\tau$  = limite superior do efeito feriado

$h(\tau, t)$  = proporção de  $\tau$  dias do período de carnaval que ocorre no mês correspondente a  $t$

Note que  $\bar{\alpha}$  é constante para todos os dias durante o período de carnaval. Esta hipótese simplificadora é importante pois os dados disponíveis, geralmente mensais, não permitem a estimação dos diferentes efeitos diários, por exemplo, entre o sábado de carnaval e a quarta-feira de cinzas. No entanto, na seção 6.3, faremos um teste de hipótese para verificar se esta restrição é válida.

A estratégia proposta por Bell e Hillmer e que utilizaremos abaixo é verificar se para diferentes valores de  $\tau$ , os modelos AR e BSM com a incorporação de  $x_t$  apresentam melhor performance vis-à-vis os modelos puros.

Isto posto, os modelos BSM e AR do capítulo 2 são modificados com as seguintes especificações respectivamente:

$$y_t = \mu_t + \gamma_t + x_t + \varepsilon_t$$

$$y_t = \mu + (\phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p) y_t + x_t + \varepsilon_t$$

## 6.2 - Consultas ao crédito

Retomando as quatro equações escolhidas que foram utilizadas para a série de consultas ao SCPC e adicionando o efeito feriado, temos:

$$y_t^* = a_2 y_{t-2}^* + a_{12} y_{t-12}^* + a_{24} y_{t-24}^* + x_t + \varepsilon_t \text{ (AR puro)}$$

$$y_t^* = \mu_1 D_P + a_1 y_{t-1}^* + a_2 y_{t-2}^* + a_{12} y_{t-12}^* + a_{24} y_{t-24}^* + x_t + \varepsilon_t \text{ (AR com dummy)}$$

$$y_t = \mu_t + \gamma_t + x_t + \varepsilon_t \text{ (BSM puro)}$$

$$y_t = \mu_t + \gamma_t + D_L + x_t + \varepsilon_t \text{ (BSM com dummy)}$$

Lembre-se que:

$D_P$  - dummy de "pulso" para agosto de 1994 = 1 e 0 para os demais meses

$D_L$  - dummy de "nível" para todos os meses após agosto de 1994 (inclusive)

= 1 e 0 para os anteriores

Na tabela 10, apresentamos os estimadores da vep para diferentes valores para  $\tau$ . Não apresentamos os estimadores para os outros coeficientes no intuito de simplificar a apresentação.

Tabela 10

	<u>vep</u>		<u>vep</u>		<u>vep</u>		<u>vep</u>
AR puro	0.006699	AR c/ dummy	0.006153	BSM	0.007025	BSM c/dummy	0.005338
$\tau = 4$	0.006451	$\tau = 4$	0.005968	$\tau = 4$	0.006540	$\tau = 4$	0.005297
$\tau = 5$	0.006441	$\tau = 5$	0.005958	$\tau = 5$	0.006527	$\tau = 5$	0.005268
$\tau = 6$	0.006470	$\tau = 6$	0.005974	$\tau = 6$	0.006540	$\tau = 6$	0.005289
$\tau = 7$	0.006494	$\tau = 7$	0.005988	$\tau = 7$	0.006566	$\tau = 7$	0.005314
$\tau = 8$	0.006513	$\tau = 8$	0.005999	$\tau = 8$	0.006593	$\tau = 8$	0.005335
$\tau = 9$	0.006533	$\tau = 9$	0.006013	$\tau = 9$	0.006624	$\tau = 9$	0.005367

Note que a introdução do efeito feriado melhora a vep em todos os modelos e que, o melhor resultado para todos os modelos é  $\tau = 5$  (adicionando-se o efeito da quarta-feira de cinzas).

Bell e Hillmer argumentam que "se  $\tau$  for aproximadamente independente dos estimadores dos outros parâmetros, então podemos (...)

realizar inferências. Isto pode ser verificado ao examinar as estimativas dos parâmetros e seus erros-padrão<sup>23</sup>.

A rigor, os autores mencionam que uma variação aceitável para os erros-padrão seria no máximo de 10%, que foi o que notamos na Tabela 11. Podemos neste caso realizar a inferência sem maiores problemas.

Tabela 11

	AR Puro		c/ $\tau = 5$		%	%
	coefic	desv pad	coefic	desv pad	coef	desv pad
$y_{t-2}^*$	-0.2083	0.0726	-0.2060	0.0717	-1.10%	-1.24%
$y_{t-12}^*$	-0.4796	0.0755	-0.4895	0.0747	2.06%	-1.06%
$y_{t-24}^*$	-0.2914	0.0753	-0.2828	0.0744	-2.95%	-1.20%
$\tau = 5$	-	-	-0.0516	0.0233	-	-
	AR c/ dummy		c/ $\tau = 5$		%	%
	coefic	desv pad	coefic	desv pad	coef	desv pad
$D_p$	0.2526	0.0814	0.2494	0.0805	-1.27%	-1.11%
$y_{t-1}^*$	-0.1659	0.0707	-0.1495	0.0704	-9.89%	-0.42%
$y_{t-2}^*$	-0.2245	0.0707	-0.2203	0.0699	-1.87%	-1.13%
$y_{t-12}^*$	-0.4583	0.0734	-0.4693	0.0728	2.40%	-0.82%
$y_{t-24}^*$	-0.2800	0.0727	-0.2727	0.0720	-2.61%	-0.96%
$\tau = 5$	-	-	-0.0464	0.0226	-	-

Comparamos na Tabela 12, os quatro modelos iniciais - AR e BSM puros e AR e BSM com dummy - com a introdução de  $x_t$  para  $\tau = 5$ :

Tabela 12

Estat	AR		AR		BSM		BSM	
	puro	$\tau = 5$	c/dummy	$\tau = 5$	puro	$\tau = 5$	c/dummy	$\tau = 5$
vep	0.0067	0.0064	0.0062	0.0060	0.0070	0.0065	0.0053	0.0053
rmse	0.1020	0.0916	0.0800	0.0727	0.1077	0.1045	0.0928	0.0892
mae	0.0921	0.0840	0.0721	0.0668	0.0906	0.0901	0.0787	0.0776
mape	0.0065	0.0059	0.0051	0.0047	0.0064	0.0064	0.0055	0.0055
u	0.0072	0.0064	0.0056	0.0051	0.0076	0.0074	0.0065	0.0063
um	0.7041	0.7173	0.5540	0.5741	0.6752	0.7059	0.6174	0.6426
us	0.0210	0.0058	0.0310	0.0119	0.0002	0.0002	0.0000	0.0011
uc	0.2749	0.2769	0.4150	0.4139	0.3246	0.2938	0.3826	0.3564
ur	0.0004	0.0037	0.0003	0.0030	0.0352	0.0199	0.0314	0.0172
ud	0.2955	0.2790	0.4457	0.4229	0.2896	0.2742	0.3513	0.3402

<sup>23</sup> Bell e Hillmer (1983), pag. 532.

Veja que em todas as colunas com  $\tau = 5$ , como esperado, a introdução do efeito feriado melhora as estatísticas de teste nos quatro modelos utilizados.

### 6.3 – Efeito diário constante

Na seção 6.1, o efeito diário ( $\bar{\alpha}$ ) no período do carnaval foi considerado constante.

Bell e Hillmer apresentam uma estratégia para testar esta hipótese para a série de vendas do comércio anterior ao feriado da Páscoa nos EUA. O intuito do trabalho destes autores é verificar se diferentes intervalos de tempo podem ser modelados por um parâmetro constante.

A divisão dos intervalos é feita de forma arbitrária. Os autores escolheram o "efeito agrupado" dos dois dias anteriores à Páscoa como base ( $\alpha_1$ ). A partir desta base, comparam-se os grupos de quatro dias anteriores a esta base até  $\tau = 22$ . Eles concluem que para a série em estudo, não há evidência para rejeitar a hipótese de um efeito diário constante até  $\tau = 10$ . Ou seja, não há evidência para rejeitar  $\bar{\alpha}$  como um coeficiente constante para as vendas do comércio até dez dias antes do feriado da Páscoa.

No nosso caso como o intervalo é relativamente pequeno,  $\tau = 5$ , parece não ser necessário testar esta hipótese. Porém como exercício econométrico, utilizamos o procedimento de Bell e Hillmer.



Considere que temos diferentes efeitos diários ( $\alpha_i$  para  $i=1,2, \dots$ ). Na série de consultas ao SCPC, onde utilizamos  $\tau = 5$ , o efeito feriado pode ser desmembrado da seguinte forma:

$$x_t = \sum_{i=1}^5 \alpha_i h(i, t)$$

Para testarmos a hipótese de que os efeitos diários podem ser considerados constantes, em primeiro lugar, realizamos de forma seqüencial o seguinte teste de hipótese:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_5 = 0$$

$$H_A: \alpha_i \neq 0, \alpha_{i+1} = \dots = \alpha_5 = 0 \quad \text{para } i = 1, \dots, 5$$

Caso  $H_0$  seja rejeitada, temos uma nova hipótese nula contra  $H_A$ :

$$H_0': \alpha_1 = \dots = \alpha_i, \alpha_{i+1} = \dots = \alpha_5 = 0 \quad \text{para } i > 1$$

Para realizar estes testes, utilizamos a estatística F:

$$F = \frac{(\text{RSS}(H_0) - \text{RSS}(H_A)) / J_1}{\text{RSS}(H_A) / J_2}$$

onde:

RSS - soma dos quadrados dos resíduos

$J_1$  - igual a 1 em  $H_0$  e  $i-1$  para  $H_0'$

$J_2$  -  $T$  ( $n^\circ$  de observações) -  $K$  ( $n^\circ$  de parâmetros) -  $i = 184 - 12 - i$

Bell e Hillmer partiram de  $i = 1$  para o sábado anterior à Páscoa até  $i = 22$ . Aqui ocorre um problema no estudo da série do carnaval. Isto decorre do fato para se estabelecer a base. Escolhemos duas bases: a quarta-feira de cinzas e o sábado de carnaval. Veja o resultado na Tabela 13.

Tabela 13 - Teste de Hipóteses  
Série de consultas ao crédito  
Quarta-feira de Cinzas (Base)

	F p/ Ho	F p/ Ho'	F <sub>0.05</sub>	F <sub>0.01</sub>
1	4.96		3.84	6.63
2	5.56	0.63	3.84	6.63
3	5.52	0.31	3.00	4.61
4	5.49	0.21	2.60	3.78
5	2.19	-	2.37	3.32

Sábado (Base)

	F p/ Ho	F p/ Ho'	F <sub>0.05</sub>	F <sub>0.01</sub>
5	2.25		3.84	6.63
4				
3				
2				
1				

Note que como o teste de hipóteses é feito de forma seqüencial, a escolha da base ( $i = 1$  em  $H_0$  e  $H_0'$ ) é determinante na performance dos testes. A escolha da quarta-feira de cinzas como base tem melhor resultado que o sábado de carnaval para a hipótese de efeito diário constante no período estudado.

Mesmo com esta ressalva, os resultados que encontramos para  $\tau = 5$ , parecem bastante intuitivos, pois este intervalo capta o período relevante do carnaval para as séries econômicas brasileiras.

E como pudemos verificar nas estatísticas de teste da seção 6.2, a introdução do efeito feriado introduziu uma informação importante nas especificações que utilizamos no capítulo 5.

## CONSIDERAÇÕES FINAIS

Podemos destacar três resultados como os principais desta dissertação.

O primeiro e mais importante refere-se à melhor performance do BSM em comparação ao Modelo AR para séries que apresentam choques sazonais em menor magnitude que os choques de interesse primário como foi discutido nas seções 3.3 e 4.1.

O segundo resultado, que apresentamos como uma extensão do primeiro, diz respeito às especificações próximas a raiz unitária sazonal e com amostras de diferentes dimensões. O que notamos é que Modelo AR se mostra superior ao BSM, mas a diferença diminui com a aproximação à raiz unitária do parâmetro relacionado ao termo sazonal e com o aumento da amostra.

O terceiro resultado refere-se à introdução do efeito feriado como variável explicativa relevante para o caso de séries econômicas brasileiras. A série que utilizamos, consultas ao crédito, confirmou tal hipótese.

Como extensão no estudo da sazonalidade, para futuros trabalhos sugerimos quatro linhas de pesquisa.

A primeira seria verificar a performance dos Modelos AR com a restrição de raiz unitária sazonal em processos quase integrados tal como argumentado em Campbell e Perron (1991)<sup>24</sup>.

A segunda linha de pesquisa seria a comparação dos programas de ajustamento sazonal (em especial o X-12) e o BSM.

---

<sup>24</sup> Veja nota de rodapé 9 para o caso de raiz unitária na frequência zero.

A terceira linha segue do que foi colocado na apresentação desta dissertação, a comparação e elaboração de algoritmos para os diferentes testes de raízes unitárias sazonais<sup>25</sup>. Para isso, os algoritmos que utilizamos neste trabalho poderiam ser utilizados como ponto de partida.

E finalmente, sugerimos a verificação da significância do efeito feriado (particularmente o carnaval) como variável explicativa em outras séries econômicas brasileiras. Além disso, seria interessante testar se outros feriados móveis (por exemplo, a Páscoa) também se mostram relevantes.

---

<sup>25</sup> Mencionamos na introdução que o principal teste para raízes unitárias sazonais é o HEGY. Para o caso de séries mensais, há o trabalho de Beaulieu, H.J. e Miron, J.A. (1993) – “Seasonal unit roots in aggregate U.S. data”, *Journal of Econometrics*, nº55, pp. 305-328. No capítulo 5, utilizamos Dickey, Hazra e Fuller (1984).

## BIBLIOGRAFIA

Bell, W. R. e Hillmer, S.C. (1983) - "Modeling Time Series With Calendar Variation", Journal of the American Statistical Association, vol.78, Sep., nº 383, pp. 526-534.

Campbell, J. Y. e Perron, P. (1991) - "Pitfalls and Opportunities: What Macroeconomists should know about Unit Roots", www.nber.org, Technical Working Paper, nº 100.

Dickey, D. A., Hasza, D.P. e Fuller, W.A. (1984) - "Testing for Unit Roots in Seasonal Time Series", Journal of the American Statistical Association, vol.79, Jun., nº 386, pp. 355-367.

Enders, W. (1995) - Applied econometric time series, 1ed., New York, John Wiley & Sons Inc.

Findley, D.F., Monsell, B.C., Bell, W.R. Otto, M.C. e Chen, B.C. (1998) - "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program", Journal of Business & Economic Statistics, Apr, vol.16, nº 2, pp. 127-152.

Granger, C. W. G. e Newbold, P. (1977) - Forecasting Economic Time Series, New York, Academic Press.

Ghysels, E. e D.R. Osborn (2001) - The Econometric Analysis of Seasonal Time Series, 1ed., Cambridge, Cambridge University Press.

Hamilton, J. D. (1994) - Time Series Analysis, Princeton, Princeton University Press.

Harvey, A. C. (1989) - Forecasting, structural time series models and the Kalman filter, 1ed., Cambridge, Cambridge University Press.

Harvey, A. C. e Scott, A. (1994) - "Seasonality in dynamic regression models", The Economic Journal, 104, Nov., pp. 1324-1345.

Hylleberg, S., Engle, R.F., Granger, C.W.J. e Yoo, B.S. (1990) - "Seasonal Integration and Cointegration", Journal of Econometrics, 44, pp. 215-238.

Instituto Nacional de Estadística (2002) - www.ine.es

Instituto de Pesquisas Econômicas Aplicadas (2002) - www.ipeadata.gov.br

Koopman, S.J., Shephard, N., Doornik, J. A. (1999) - "Statistical algorithms for models in state space using Ssfpack 2.2", Econometrics Journal, vol.2, pp. 113-166.

Maravall, A. (1985) - "On Structural Time Series Models and the Characterization of Components ", Journal of Business & Economic Statistics, Oct, vol.3, n° 4, pp. 350-355.

Osborn, D.R. e Smith, J.R. (1989) - "The Performance of Periodic Autoregressive Models in Forecasting Seasonal U. K. Consumption", Journal of Business & Economic Statistics, Jan, vol.7, n° 1, pp. 117-127.

Perron, P. (1989) - "The great crash, the oil price shock, and the unit root hypothesis", Econometrica, vol.57, n°6, Nov., pp. 1361-1401.

Schwert, G. W. (1989) - "Tests for unit roots: a Monte Carlo investigation", Journal of Business & Economic Statistics, Apr, vol.7, n° 2, pp. 147-159.

Taylor, A.M.R. e Leybourne, S.J. (1999) - "Detecting seasonal unit roots: An approach based on the sample autocorrelation function", The Manchester School, Jun., vol.67, n° 3, pp. 261-286.

Theil, H. (1966) - Applied Economic Forecasting, Studies in Mathematical and Managerial Economics, vol. 4, Amsterdam, North Holland Publishing Company.

## APÊNDICE - ALGORITMOS

```

// dgp.ox
// Esta rotina gera DGPs com séries mensais
//  $y(t) = y(t-1) + s(t) + e(t)$  - equação principal
// onde  $s(t) = \beta s(t-12) + n(t)$  - equação da sazonalidade
//  $e(t) \sim N(0,1)$       $n(t) \sim N(0,S)$       $E(e(t),n(t)) = 0$  para todo t
// Autor: Carlos Assato

#include <oxstd.h>

main()
(
    decl a, r, desvp, s0, y0; // a = tamanho amostra    r= núm de repet
                               // desvp= fator do desv pad do choque da saz
(S^(1/2))
                               // s0 = vals inic da dif entre os meses
                               // y0 = valor inicial da serie yt

    decl yt, y_mod, y_teste; // yt = matriz dos dados
                               // y_mod = série a ser submetida aos
modelos
                               // y_teste = série para previsão

    decl et, nl, nt;           // et = choques da equação principal
                               // nt = choques da sazonalidade s(t) - nl
                               // ponderados pelo desvio padrão (desvp)

    decl dim_s0, st;          // st = matriz de termos sazonais

    decl i, j;                // contadores

    a = 636;
    r = 25000;
    s0 = <-15.8;7.8;24.5;23.2;9;3;46;13;-33;-17;-30;8>; // turismo da
Espanha
    y0 = 100;
    desvp = 0.01;
    ranseed(today());
    et = rann(a,r);
    nl = rann(a,r);
/* savemat("erroprinc.mat", et);
   savemat("errosazonal.mat", nl); */
    nt = nl * desvp;

    dim_s0 = sizer(s0);
    st = zeros(a,r);
    i = 0;

    while (i < dim_s0)
    (
        st[i][] = s0[i][0] + nt[i][]; // st é feita em duas etapas
        ++i;                          // (i) para os valores iniciais
    )

    i = dim_s0;

    while (i < a)
    (
        st[i][] = st[i-12][] + nt[i][]; // (ii) para os demais
        ++i;
    )

    yt = zeros(a,r);
    yt[0][] = y0 + st[0][] + et[0][];

```

```

    i = 1;

    while (i < a)
    {
        yt[i][j] = yt[i-1][j] + st[i][j] + et[i][j];
        ++i;
    }

    for (j=0; j < r; ++j)    // verifica se "inverno transformou-se em
verão"
    {
        for (i=7; i < a ; i+=12)
        {
            if (yt[i][j] < yt[i-7][j])
                println("há um problema na série - tipo 1");
        }
        for (i=0; i < a; ++i) // verifica se yt possui numeros
negativos
        {
            if (yt[i][j] < <0>)
                println("há um problema na série - tipo 2");
        }
    }

    y_mod = yt[24:(a-dim_s0-1)][j]; // descartam-se as primeiras
observações
    y_teste = yt[(a-dim_s0):][j]; // valores para os testes de previsão

    savemat("geral.mat", y_mod);
    savemat("y_teste.mat", y_teste);
}

// mod_geral.ox
// Este programa estima o BSM e Modelo AR separadamente
// para séries mensais com variância constante nas frequências sazonais
// as séries utilizadas foram construídas a partir do programa dgp.ox
// Autor: Carlos Assato

#include <oxstd.h>
#include <oxfloat.h>
#import <maximize>
#include <packages/ssfpack/ssfpack.h>

static decl s_mY, s_cT, s_mStsm, s_vVarCmp, dvar, s_dSigma, s_dVar, s_vPar;
static decl y_estrut, mforc, var_pred_est, valor_predicao, cont_num, z;

SetStsmModel(const vP)
{
    s_mStsm = < CMP_LEVEL,      1,  0, 0;
                CMP_SLOPE,     1,  0, 0;
                CMP_SEAS_TRIG,  1, 12, 0;
                CMP_IRREG,      1,  0, 0>;

    decl vr = exp(2.0 * vP);
    s_vVarCmp = // s_vVarCmp é a
diagonal(Omega)
    // level  slope  - monthly trig seasonal  - irreg
    vr[0] | vr[1] | vr[2] | vr[2] | vr[2] | vr[2] | vr[2] |
    vr[2] | vr[2] | vr[2] | vr[2] | vr[2] | vr[2] | vr[3];
}

LogLikStsm(const vY, const pdLik, const pdVar)
{
    decl mphi, momega, msigma, ret_val;
    GetSsfStsm(s_mStsm, &mphi, &momega, &msigma); // modelo de espaço
estado

```



```

        momega = diag(s_vVarCmp); // Ômega de s_vVarCmp
        ret_val = SsfLik(pdLik, pdVar, vY, mphi, momega, msigma);
        return ret_val; // 1 indica sucesso, 0
fracasso
}

LogLikScoStsm(const vY, const pdLik, const pvSco)
{
    decl mphi, momega, msigma, msco, ret_val;
    GetSsfStsm(s_mStsm, &mphi, &momega, &msigma);
    momega = diag(s_vVarCmp); // transforma o modelo de espaço estado
        // loglik e matriz score
    ret_val = SsfLikSco(pdLik, &dvar, &msco, vY, mphi, momega, msigma);
    if(dvar != 0)
    {
        decl vs = (diagonal(msco)' .* s_vVarCmp);
        pvSco[0][0][0] = vs[0];
        pvSco[0][1][0] = vs[1];
        pvSco[0][2][0] = vs[2] + vs[3] + vs[4] + vs[5] + vs[6] + vs[7] +
vs[8] +
        vs[9] + vs[10] + vs[11] + vs[12];
        pvSco[0][3][0] = vs[13];
        return ret_val; // 1 indica sucesso, 0 fracasso
    }
}

Likelihood2(const xP, const pdLik, const pvSco, const pmHes)
{
    decl ret_val;
    SetStsmModel(xP);
    ret_val = pvSco ? LogLikScoStsm(s_mY, pdLik, pvSco)
        : LogLikStsm(s_mY, pdLik, &s_dVar);
    pdLik[0] = s_cT ;
    pvSco[0][][] /= s_cT ;
    return ret_val;
}

InitialPar2()
{
    decl xp = <0; -200; -5; -1.5>; // vals inic para o processo de ML
        // tend / decl / saz / erro
    SetStsmModel(xp);
    decl dlik, dvar;
    LogLikStsm(s_mY, &dlik, &dvar);
    return xp + 0.5 * log(dvar);
}

Likelihood(const vP, const pdLik, const pvSco, const pmHes)
{
    decl ret_val;
    SetStsmModel(vP);
    if(dvar >= 1e-5)
    {
        ret_val = pvSco ? LogLikScoStsm(s_mY, pdLik, pvSco)
            : LogLikStsm(s_mY, pdLik, &s_dVar);
        pdLik[0] = s_cT ;
        pvSco[0][][] /= s_cT ;
        return ret_val; // 1 indica sucesso, 0 fracasso
    }
}

InitialPar()
{
    decl vp = < 0; -200; -5; -1.5>; // vals inic para o processo de ML
        // tend / decl / saz / erro
    SetStsmModel(vp);
    decl dlik;

```

```

    LogLikStsm(s_mY, &dlik, &dvar);
    return vp + 0.5 * log(dvar);
}

MaxLik()
{
    decl vp, dlik, ir, xp, yp;
    vp = InitialPar();
    MaxControl(200, 30, 1);
    MaxControlEps(1e-5, 1e-3);
    dvar=<1>;
    ir = MaxBFGS(Likelihood, &vp, &dlik, 0, FALSE);
    if(dvar <= 1e-5)
    {
        xp = InitialPar2();
        ir = MaxBFGS(Likelihood2, &xp, &dlik, 0, 1);
        cont_num[z][]=1;
    }

    println("\n", MaxConvergenceMsg(ir),
        " using analytical derivatives",
        "\nLog-likelihood = ", "%.8g", dlik * s_cT,
        "; n = ", s_cT, ";");
    println("variance parameters (* 10,000):",
        "%5.2f", 10000 * s_vVarCmp);
    s_vPar = vp;
}

Var_Pred(const mY)
{
    decl mphi, momega, msigma, mstate, mfor, m, dif;
    GetSsfStsm(s_mStsm, &mphi, &momega, &msigma);
    momega = diag(s_vVarCmp);
    m = columns(mphi);
    mstate = SsfMomentEst(ST_PRED, &mfor, mY, mphi, momega, msigma);
    dif = mY[1:(m+1):] - mfor[m][1:(m+1):];
    var_pred_est = varr(dif)*sizec(dif)/(sizec(dif)-1);
    valor_predicao = mfor[m][1];
}

Forc(const vY, const cForc)
{
    decl mphi, momega, msigma, mstate, mfor, m;
    GetSsfStsm(s_mStsm, &mphi, &momega, &msigma);
    momega = diag(s_vVarCmp);
    m = columns(mphi);
    mstate = SsfMomentEst(ST_PRED, 0, vY, mphi, momega, msigma);
    SsfMomentEst(ST_PRED, &mfor, constant(M_NAN, 1, cForc),
        mphi, momega, mstate);
    return (vY ~ mfor[m][1] | (zeros(vY) ~ sqrt(mfor[2 * m + 1][1]));
}

Mod_estrut()
{
    decl y_model, y_size, var_pred_estrut, valores_pred_estrut;
    y_model = loadmat("geral.mat");
    y_size = rows(y_model);
    var_pred_estrut = zeros(y_size,1);
    y_estrut = zeros(12,y_size);
    valores_pred_estrut = zeros(sizec(y_model),y_size);
    cont_num = zeros(y_size,1);
    z = 0;
    while (z < y_size)
    {
        s_mY = y_model[z][1];
        s_cT = columns(s_mY);
        MaxLik();
    }
}

```

```

        Var_Pred(s_mY);
        var_pred_estrut[z][] = var_pred_est[][];
        valores_pred_estrut[][z] = valor_predicao[][];
        mforc = Forc(s_mY, 12)';
        y_estrut[][z] = mforc[s_CT:][0];
        println("A coluna ", z, " esta ok!");
        println(" ");
        savemat("y_estrut.mat", y_estrut);
        savemat("var_pred_estrut.mat", var_pred_estrut);
        savemat("controle.mat", cont_num);
        ++z;
    }

/*   savemat("valores_pred_estrut.mat", valores_pred_estrut); */
}

Mod_AR()
{
    decl ar, dados1, dados, dados_nivel, n, col, max_def;
    decl yt, size_yt, beta, xt;
    decl ytil, ytil_nivel, erro_pred, var_erro_pred, yhat, yhat_nivel;
    decl q, i; // contadores
    decl ar12, ar24, ar36, ar48; // termos AR
    decl ar12_prev, ar24_prev, ar36_prev, ar48_prev; // previsão

    ar=<12;24;36;48>; // defasagens que serão utilizadas
    dados_nivel = loadmat("geral.mat");
    dados1 = diff0(dados_nivel, 1)[1:][];
    dados = diff0(dados1, 12)[12:][];

    n = sizer(dados); // tamanho da amostra
    col = sizec(dados); // número de séries
    max_def = ar[3]; // defasagem máxima do modelo AR

    yt = dados[max_def:][];
    size_yt = sizer(yt);
    ar12 = dados[(max_def-12):(n-13)];
    ar24 = dados[(max_def-24):(n-25)];
    ar36 = dados[(max_def-36):(n-37)];
    ar48 = dados[(max_def-48):(n-49)];

    q = 0; // contador
    beta = zeros(sizer(ar), col); // parâmetros do modelo AR
    while (q < col)
    {
        xt = ar12[][q] ~ ar24[][q] ~ ar36[][q] ~ ar48[][q];
        beta[][q] = invert(xt'*xt)*(xt'*yt[][q]);
        ++q;
    }

/*   savemat("beta.mat", beta); */

    // construção dos valores de predição (ytil)
    // e da variância (var_erro_pred) - método de um passo à frente
    ytil = zeros(size_yt, col);
    q = 0;
    while (q < col)
    {
        i = 0;
        while (i < size_yt)
        {
            ytil[i][q] = beta[0][q] * ar12[i][q] + beta[1][q] *
ar24[i][q]
ar48[i][q];
            + beta[2][q] * ar36[i][q] + beta[3][q] *
            ++i;
        }
        println(q);
    }
}

```

```

        ++q;
    }

    i = 1;
    ytil_nivel = zeros(size_yt+1,col);
    ytil_nivel[0][] = dados_nivel[max_def+12][];
    while (i < sizer(ytil_nivel))
    {
        ytil_nivel [i][] = dados_nivel[max_def+12+i-1][] + ytil[i-1][]
+
        dados1[max_def+i-1][];
        ++i;
    }
/* savemat("ytil.mat", ytil_nivel); */
erro_pred = dados_nivel[max_def+13][] - ytil_nivel[1:][];
// erro de predição
var_erro_pred = varc(erro_pred)*sizer(erro_pred)/(sizer(erro_pred)-
1);
savemat("var_pred_arma.mat", var_erro_pred');

// construção da matriz dos valores previstos (yhat)
// método de um passo à frente
ar12_prev = dados[(n-12):][];
ar24_prev = dados[(n-24):(n-13)];
ar36_prev = dados[(n-36):(n-25)];
ar48_prev = dados[(n-48):(n-37)];
q = 0;
yhat = zeros(12,col); // doze previsões para comparação
while (q < col)
{
    i = 0;
    while (i < 12)
    {
        yhat[i][q] = beta[0][q] * ar12_prev[i][q]+
        beta[1][q] * ar24_prev[i][q]+ beta[2][q] *
ar36_prev[i][q] +
        beta[3][q] * ar48_prev[i][q];
        ++i;
    }
    ++q;
}
yhat_nivel = zeros(12,col);
yhat_nivel[0][] = dados_nivel[n+12][] + yhat[0][] + dados1[n][];
i = 1;
while (i < 12)
{
    yhat_nivel[i][] = yhat_nivel[i-1][] + yhat[i][] +
dados1[n+i][];
    ++i;
}
savemat("y_arma.mat", yhat_nivel[][]);
}

main()
{
    Mod_estrut();
    Mod_AR();
}

// estat_final.ox
// construção das estatísticas de testes para comparação dos dois modelos
// comparacao com o modelo AR - os valores ficarão na 1a. linha
// comparação com o modelo estrutural - os valores ficarão na 2a. linha

// Autor: Carlos Assato

```

```

#include <oxstd.h>

decl y_real_tot, n, col; // matriz das séries para comparação
decl rmse, mae, mape, in_coef, um, us, uc, ur, ud;
// matrizes das estatísticas de teste
decl cont_rmse, cont_mae, cont_mape, cont_in_coef, cont_um,
    cont_us, cont_uc, cont_ur, cont_ud, cont_var_predicao; // contadores
decl y_arma_tot, e_n_arma, corr_coef_arma, r_arma; // var do modelo arma
decl y_estrut_tot, e_n_estrut, corr_coef_estrut, r_estrut; // var do modelo
estrut
decl var_pred_arma_tot, var_pred_estrut_tot; // matriz de var dos erros de
predicao
decl y_real, y_arma, y_estrut, var_pred_arma, var_pred_estrut;
decl estat_final;
decl i,j; // contador

main()
{
    y_real_tot = loadmat("y_teste.mat");
    y_arma_tot = loadmat("y_arma.mat");
    y_estrut_tot = loadmat("y_estrut.mat");
    var_pred_arma_tot = loadmat("var_pred_arma.mat");
    var_pred_estrut_tot = loadmat("var_pred_estrut.mat");

    cont_rmse = 0;
    cont_mae = 0;
    cont_mape = 0;
    cont_in_coef = 0;
    cont_um = 0;
    cont_us = 0;
    cont_uc = 0;
    cont_ur = 0;
    cont_ud = 0;
    cont_var_predicao = 0;

    for(j=0; j<25000; j+=100)
    {
        y_real = y_real_tot[][j:j+99];
        y_arma = y_arma_tot[][j:j+99];
        y_estrut = y_estrut_tot[][j:j+99];
        var_pred_arma = var_pred_arma_tot[j:j+99][];
        var_pred_estrut = var_pred_estrut_tot[j:j+99][];
        e_n_arma = y_real - y_arma;
        e_n_estrut = y_real - y_estrut;
        n = sizer(y_real);
        col = sizec(y_real);

        rmse = zeros(2,col);
        mae = zeros(2,col);
        mape = zeros(2,col);
        in_coef = zeros(2,col);
        um = zeros(2,col);
        us = zeros(2,col);
        uc = zeros(2,col);
        ur = zeros(2,col);
        ud = zeros(2,col);

        i = 0;
        while (i < col)
        {
            rmse[0][i] = ((1/n).*sumc(e_n_arma.^2)[][i]).^(1/2);
            mae[0][i] = (1/n).*sumc(fabs(e_n_arma)[][i]);
            mape[0][i] = (1/n).*sumc(fabs(e_n_arma./y_real)[][i]);
            in_coef[0][i] =
(sumc(e_n_arma.^2)[][i]/sumc(y_real[][i].^2)).^(1/2);
            um[0][i] = ((meanc(y_arma[][i]) - meanc(y_real[][i])).^2)/

```

```

                ((1/n).*sumc(e_n_arma.^2)[][i]);
        us[0][i] = (((sqrt(varc(y_arma[][i])))-
                (sqrt(varc(y_real[][i])))).^2) /
((1/n).*sumc(e_n_arma.^2)[][i]);
        corr_coef_arma = correlation(y_real[][i] ~ y_arma[][i]);
        r_arma = corr_coef_arma[0][1];
        uc[0][i] = 2*(1-r_arma)*sqrt(varc(y_arma[][i]))*
                sqrt(varc(y_real[][i]))
/((1/n).*sumc(e_n_arma.^2)[][i]);
        ur[0][i] = ((sqrt(varc(y_arma[][i]))-(r_arma *
                sqrt(varc(y_real[][i])))).^2)/
((1/n).*sumc(e_n_arma.^2)[][i]);
        ud[0][i] = (1-(r_arma^2)) * (varc(y_real[][i])) / ((1/n)
                .*sumc(e_n_arma.^2)[][i]);

        rmse[1][i] = ((1/n).*sumc(e_n_estrut.^2)[][i]).^(1/2);
        mae[1][i] = (1/n).*sumc(fabs(e_n_estrut)[][i]);
        mape[1][i] = (1/n).*sumc(fabs(e_n_estrut./y_real)[][i]);
        in_coef[1][i] = (sumc(e_n_estrut.^2)[][i] /
                sumc(y_real[][i].^2)).^(1/2);
        um[1][i] = ((meanc(y_estrut[][i]) - meanc(y_real[][i])).^2)/
                ((1/n).*sumc(e_n_estrut.^2)[][i]);
        us[1][i] = (((sqrt(varc(y_estrut[][i])) -
                (sqrt(varc(y_real[][i])))).^2) /
((1/n).*sumc(e_n_estrut.^2)[][i]);
        corr_coef_estrut = correlation(y_real[][i] ~ y_estrut[][i]);
        r_estrut = corr_coef_estrut[1][0];
        uc[1][i] = 2*(1-r_estrut) * sqrt(varc(y_estrut[][i]))
                *sqrt(varc(y_real[][i])) /
((1/n).*sumc(e_n_estrut.^2)[][i]);
        ur[1][i] = ((sqrt(varc(y_estrut[][i])) - (r_estrut *
                sqrt(varc(y_real[][i])))).^2)/((1/n).*sumc(e_n_estrut.^2)[][i]);
        ud[1][i] = (1 - (r_estrut^2)) * (varc(y_real[][i]))
                /((1/n).*sumc(e_n_estrut.^2)[][i]);

        if (rmse[0][i] > rmse[1][i])
            cont_rmse = cont_rmse + 1;
        if (mae[0][i] > mae[1][i])
            cont_mae=cont_mae+1;
        if (mape[0][i] > mape[1][i])
            cont_mape = cont_mape + 1;
        if (in_coef[0][i] > in_coef[1][i])
            cont_in_coef = cont_in_coef + 1;
        if (um[0][i] > um[1][i])
            cont_um = cont_um + 1;
        if (us[0][i] > us[1][i])
            cont_us = cont_us + 1;
        if (uc[0][i] > uc[1][i])
            cont_uc = cont_uc + 1;
        if (ur[0][i] > ur[1][i])
            cont_ur = cont_ur + 1;
        if (ud[0][i] > ud[1][i])
            cont_ud = cont_ud + 1;
        if (var_pred_arma[i][1] > var_pred_estrut[i][1])
            cont_var_predicao = cont_var_predicao + 1;
        ++i;
    }
}

estat_final = zeros(10,1);
estat_final = cont_var_predicao | cont_rmse | cont_mae | cont_mape |
                cont_in_coef | cont_um | cont_us | cont_uc | cont_ur |
                cont_ud;
estat_final = estat_final./25000;
println("valores ", estat_final);

```

```

        savemat("estat_final.mat", estat_final);
    }
    // verif_estat.ox
    // rotina para verificação das estatísticas da Desigualdade de Theil
    // este programa está dividido para agilizar o tempo de computação
    // autor: Carlos Assato

#include <oxstd.h>

decl y_teste, y_arma_tot, y_estrut_tot; // dados das matrizes "gerais"
decl y_real, n, col; // dados da matriz de dados
"reais" para comparacao
decl y_arma, e_n_arma, corr_coef_arma, r_arma; // variaveis do
modelo arma
decl y_estrut, e_n_estrut, corr_coef_estrut, r_estrut; // variaveis do
modelo estrut
decl i,j; // contadores
decl estatisticas, medias; // variáveis de interesse

main()
{
    y_teste = loadmat("y_teste.mat");
    y_arma_tot = loadmat("y_arma.mat");
    y_estrut_tot = loadmat("y_estrut.mat");
    medias = zeros(14,250);

    for(j=0; j<25000; j+=100)
    {
        y_real = y_teste[][j:j+99];
        y_arma = y_arma_tot[][j:j+99];
        y_estrut = y_estrut_tot[][j:j+99];

        n = sizer(y_real);
        col = sizec(y_real);
        estatisticas = zeros(14,col);

        // matrizes de erros dos modelos
        e_n_arma = y_real - y_arma;
        e_n_estrut = y_real - y_estrut;

        i = 0;
        while (i < col)
        {
            corr_coef_arma = correlation(y_real[][i] ~ y_arma[][i]);
            r_arma = corr_coef_arma[0][1];
            corr_coef_estrut = correlation(y_real[][i] ~
y_estrut[][i]);
            r_estrut = corr_coef_estrut[0][1];

            estatisticas[0][i] = ((meanc(y_arma[][i])-
meanc(y_real[][i])).^2)/((1/n).*sumc(e_n_arma.^2)[][i]);
            estatisticas[1][i] = (((sqrt(varc(y_arma[][i])))-
(sqrt(varc(y_real[][i])))).^2)/((1/n).*sumc(e_n_arma.^2)[][i]));
            estatisticas[2][i] = 2*(1-
r_arma)*sqrt(varc(y_arma[][i]))*sqrt(varc(y_real[][i]))/((1/n).*sumc(e_n_arm
a.^2)[][i]));
            estatisticas[3][i] = ((sqrt(varc(y_arma[][i]))-
(r_arma*sqrt(varc(y_real[][i])))).^2)/((1/n).*sumc(e_n_arma.^2)[][i]));
            estatisticas[4][i] = (1-
(r_arma^2))*varc(y_real[][i])/((1/n).*sumc(e_n_arma.^2)[][i]);
            estatisticas[5][i] = sqrt(varc(y_arma[][i]));
            estatisticas[6][i] = r_arma[0][1];
            estatisticas[7][i] = ((meanc(y_estrut[][i])-
meanc(y_real[][i])).^2)/((1/n).*sumc(e_n_estrut.^2)[][i]));

```

```

        estadisticas[8][i] = (((sqrt(varc(y_estruc[] [i])))-
(sqrt(varc(y_real[] [i])))).^2) /((1/n).*sumc(e_n_estruc.^2) [] [i]);
        estadisticas[9][i] = 2*(1-
r_estruc)*sqrt(varc(y_estruc[] [i]))*sqrt(varc(y_real[] [i])) /
((1/n).*sumc(e_n_estruc.^2) [] [i]);
        estadisticas[10][i] = ((sqrt(varc(y_estruc[] [i])))-
(r_estruc*sqrt(varc(y_real[] [i])))).^2)/((1/n).*sumc(e_n_estruc.^2) [] [i]);
        estadisticas[11][i] = (1-
(r_estruc^2))*varc(y_real[] [i])/((1/n).*sumc(e_n_estruc.^2) [] [i]);
        estadisticas[12][i] = sqrt(varc(y_estruc[] [i]));
        estadisticas[13][i] = r_estruc[] [];
        ++i;
    }

    i = 0;
    while (i < 14)
    {
        medias[i][j/100] = meanr(estadisticas[i] []);
        ++i;
    }
}
savemat("medias.mat", meanr(medias));

```