

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
DEPARTAMENTO DE ESTATÍSTICA

**MODELOS DE SÉRIES TEMPORAIS COM  
MUDANÇA DE REGIME MARKOVIANA**

Autor: Michel Charles Piper  
Orientador: Prof. Dr. Flávio A. Ziegelmann

Trabalho de conclusão apresentado para a obtenção  
do grau de Bacharel em Estatística

Porto Alegre, fevereiro de 2004

## AGRADECIMENTOS

Agradeço, em primeiro lugar a Deus, por ter me dado a vida.

Agradeço a todos aqueles que contribuíram de alguma forma para a realização deste trabalho e que me deram o apoio necessário durante o curso.

Ao professor e orientador Flávio, pela atenção oferecida para a conclusão deste trabalho.

À professora Jandyra, por ter me dado a primeira oportunidade de aprendizado prático, através de bolsa de trabalho no NAE.

Aos demais professores, pelos conhecimentos transmitidos e que serão de grande valia para o meu futuro profissional. Principalmente àqueles que ouviam aos chamados dos alunos, prontos para atendê-los.

Aos amigos de infância, Moacir, Fabrício e Rodrigo, que me acompanham até os dias de hoje. Agradeço também a todos os meus amigos do município de São Paulo das Missões.

Aos amigos do curso de estatística, principalmente aqueles que comigo compartilharam horas de estudo, discussões e lazer, e que sempre estiveram presentes quando foi necessário. Agradeço também aos demais amigos da Faculdade. Aos amigos da CEFAV, que como eu tiveram a oportunidade de morar numa Casa de Estudantes, lutando para mantê-la, possibilitando assim, o avanço dos estudos.

Aos amigos e colegas de serviço da Siqueira Campos, pelo apoio.

Aos tios e tias, que, sempre que foi necessário, fizeram o papel de pais.

À Raquel, pela vontade de ajudar, pela companhia, incentivo e carinho, que só ela é capaz de oferecer.

Em especial aos meus queridos pais, Darci e Francisco, e a minha irmã Francieli, por terem me dado as principais lições de vida, pelo carinho e pelo apoio dado perante as dificuldades enfrentadas no dia a dia.

## ÍNDICE

<b>RESUMO.....</b>	<b>v</b>
<b>1. INTRODUÇÃO.....</b>	<b>1</b>
<b>2. CADEIAS DE MARKOV EM MODELOS MS-VAR.....</b>	<b>4</b>
2.1. Definições das cadeias de Markov.....	4
2.2. Cadeias de Markov na forma de vetor autorregressivo .....	5
2.3. Cadeias de Markov Irredutíveis.....	7
2.4. Cadeias de Markov com Regimes Ergódicos.....	8
2.5. Tempo de Permanência nos Regimes.....	9
<b>3. INTRODUÇÃO AOS MODELOS ARIMA E METODOLOGIA DE BOX-JENKINS.....</b>	<b>10</b>
3.1. Modelos Estacionários.....	11
3.1.1. Modelos SARIMA.....	11
3.1.2. Modelos autorregressivos (AR).....	12
3.2. Função de Autocorrelação Amostral (FAC).....	13
3.3. Função de Autocorrelação Parcial Amostral (PACF).....	14
3.4. Modelos Não-estacionários.....	15
3.5. Metodologia Box-Jenkins.....	15
3.6. Critérios para a Escolha de Modelos.....	16
<b>4. METODOLOGIA MS-VAR.....</b>	<b>18</b>
4.1. Modelos de Mistura de Distribuições Normais .....	18
4.2. Modelos Autorregressivos.....	20
4.3. Modelos de Mudança de Regime na Variância .....	23
4.4. Estimação dos Parâmetros de Modelos Autorregressivos.....	24
4.5. Probabilidade Preditas, Filtradas e Suavizadas .....	25
4.5.1. Inicializando as Estimativas.....	28

4.5.2. Algoritmo Tipo EM.....	29
4.5.3. Inicializando o Algoritmo Tipo EM.....	30
4.6. Previsões .....	30
4.7. Estimativa de Máxima Verossimilhança dos Parâmetros.....	31
4.8. Número Adequado de Regimes .....	33
<b>5. SIMULAÇÃO UTILIZANDO O R.....</b>	<b>34</b>
5.1. Simulação de Modelos.....	34
5.1.1. Caso 1.....	35
5.1.2. Caso 2.....	39
5.1.3. Caso 3.....	45
<b>6. APLICAÇÃO A UM CASO REAL.....</b>	<b>50</b>
<b>7. CONCLUSÕES.....</b>	<b>55</b>
<b>8. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>57</b>
<b>9. ANEXOS: PROGRAMAS EM R.....</b>	<b>60</b>

## RESUMO

A análise de séries temporais se caracteriza pelo estudo de observações autocorrelacionadas ao longo do tempo. Neste trabalho abordaremos um caso particular de modelos de análise de séries temporais: Modelos com Mudança de Regime Markoviana. Estes modelos são aplicados a séries que apresentam comportamentos distintos ao longo do tempo, onde estes comportamentos podem ser descritos por regimes de uma Cadeia de Markov. Neste sentido, discutiremos a teoria relevante para os modelos que abordam mudança de regime, como as Cadeias de Markov e os modelos SARIMA, além da metodologia Box-Jenkins. Este trabalho será enriquecido pela realização de simulações de Monte Carlo de séries temporais com e sem mudança de regime, propiciando averiguar a adequação da metodologia. Para finalizar apresentamos a modelagem de uma série empírica com dados do Índice Geral de Produção Nacional dos E.U.A. e as conclusões pertinentes. Os modelos com mudança de regime foram ajustados a partir da implementação da rotina adequada no software *R*.

Palavras Chave: séries temporais, mudança de regime, cadeias de Markov, Box-Jenkins, SARIMA, MS-VAR, simulação.

## ABSTRACT

Time series analysis involves the study of observations along time. In this work we approach a particular times series model: the Markov Regime Switching Model. These models are applied to series that possibly do not have a constant pattern along time. Besides we assume that this switching behaviour may be described by a Markov chain. Therefore, relevant issues about Markov chains, SARIMA models and Box-Jenkins methods are discussed. Monte Carlo simulations are performed from models with and without regime switching, allowing one to check and compare the fitted models. We also analyse an empirical time series, the U.S.A. General National Production (GNP) series. The codes for estimation of Markov regime switching models were implemented in the *R* software.

Key words: time series, regime switching, Markov chains, Box-Jenkins, SARIMA, MS-VAR, simulation.

## 1. INTRODUÇÃO

O presente trabalho tem como primeira motivação a curiosidade intelectual e como objetivo principal a compreensão e implementação da metodologia que será abordada.

A semente para a escolha do tema apresentado foi plantada no estudo de modelos paramétricos de previsão de séries temporais conhecidos como modelos ARIMA(p,d,q) - (Auto-Regressivo Integrado de Médias Móveis), seguindo a metodologia sugerida por Box e Jenkins (1976). Tal metodologia foi estudada na disciplina de Séries Temporais no curso de graduação em Estatística desta universidade.

Uma série temporal pode ser definida como um conjunto de observações coletadas e ordenadas ao longo do tempo. Os modelos utilizados para descrever o mecanismo gerador destas séries temporais são processos estocásticos, ou seja, processos que evoluem segundo leis probabilísticas.

Modelos ARIMA (p,d,q) descrevem o comportamento de uma variável aleatória ao longo do tempo como função dos valores desta mesma variável aleatória no passado. Os parâmetros  $p$ ,  $d$ ,  $q$  referem-se respectivamente às partes autorregressiva, de diferenciação e de médias móveis presentes no modelo. Uma generalização de modelos ARIMA, que incluem uma componente sazonal (S) no modelo, são os modelos SARIMA. Estes modelos apresentam uma parte não sazonal, representada pelos parâmetros (p,d,q), e outra sazonal, representada pelos parâmetros (P,D,Q). A versão multiplicativa deste tipo de modelo recebe a nomenclatura de SARIMA(p,d,q)<sub>x</sub>(P,D,Q)<sub>s</sub> (ver Shummway, 1988).

Box-Jenkins é uma metodologia de ajustamento iterativo de modelos SARIMA em séries temporais, sendo que o objetivo principal é encontrar um modelo apropriado para os dados, de modo que os resíduos sejam tão pequenos quanto possível (respeitando o princípio de parcimônia) e não exibam padrão algum. A metodologia envolve as seguintes etapas: identificação do modelo, estimação dos parâmetros e checagem da adequação do modelo. Este ciclo é repetido (se necessário for) até que se encontre um modelo apropriado dentre os elegíveis. Salientamos que podem ser sugeridos vários modelos, onde dentre estes escolhemos aquele que simultaneamente se ajusta bem aos dados e não é demasiadamente complexo (princípio de parcimônia).

Neste trabalho decidimos estudar modelos de previsão mais sofisticados, ainda não incorporados a muitos livros texto tradicionais de séries temporais. O assunto escolhido foi Modelos Auto-Regressivos de Séries Temporais com Mudança de Regime Markoviana (MS-AR).

Interpretamos mudança de regime como uma passagem do processo de um estado para outro de comportamento distinto. Para entender melhor este conceito suponha um processo onde, com o passar do tempo, a variável de interesse passe a flutuar em diferentes faixas de valores. Cada faixa distinta poderia constituir um diferente regime ou estado do processo. Por exemplo, para a aplicação destes modelos em séries econômicas, é comum estabelecermos dois regimes, onde um regime seria de recessão ou de estagnação econômica e o outro de expansão econômica. Essas mudanças de estado ou de regime podem ter diversas explicações ou origens, como, por exemplo, serem resultantes de um período de guerra, crises econômicas ou até uma severa mudança na política governamental adotada em determinado país. Espera-se que a metodologia MS-AR identifique com a maior precisão possível os períodos em que ocorrem esses fenômenos.

Modelos de Séries Temporais contendo vetores Auto-Regressivos com Mudança de Regime Markoviana (MS-VAR – “Markov Switching Vector Autoregression”) foram propostos inicialmente por Hamilton no final da década de 80 (Hamilton, 1989), com o objetivo de estudar flutuações econômicas nos Estados Unidos, passando a ter maior importância ao longo da década de 90, principalmente nas áreas de economia e finanças. Devido a algumas crises financeiras mundiais ocorridas nos últimos anos, onde as principais variáveis financeiras e macroeconômicas passaram por relevantes alterações de seu comportamento em períodos de tempo relativamente pequenos, o interesse pelo estudo e utilização de modelos com mudança de regime têm aumentado consideravelmente. Estes modelos tentam captar de forma endógena as alterações de comportamento dessas variáveis, ajustando-se rapidamente à nova dinâmica da série. Na literatura recente encontramos alguns trabalhos aplicando a metodologia. Entre os quais podemos citar os trabalhos de Hamilton (1994) e Hamilton (2002).

Além de processos ARIMA, particularmente no estudo de modelos MS-AR, veremos que os processos chamados de Cadeias de Markov (CM) são de fundamental relevância, pois a natureza das mudanças de regime está baseada nestas cadeias. Os modelos MS-AR podem ser considerados como uma generalização de modelos ARIMA

(p, d, q), porém considerando-se apenas a parte auto-regressiva do modelo. Ou seja, seria uma generalização de modelos AR(p) ou ARIMA(p,0,0).

Modelos com mudança de regime apresentam variações entre os parâmetros de diferentes regimes. O processo  $y_t$  dependerá de uma variável não observável referente ao regime, a qual será denotada por  $s_t$ . Quando uma série temporal estiver sujeita a mudanças de regime, uma modelagem, por exemplo, do tipo AR(p) não é a mais adequada, uma vez que estes modelos apresentam parâmetros invariantes no tempo.

Ao longo deste trabalho serão discutidos os principais conceitos a respeito das CM e modelos ARIMA. Nos Capítulos 2 e 3 encontraremos toda a parte teórica relevante ao trabalho sobre estes tópicos. Porém o foco principal reside no modelo “misto” MS-AR e na metodologia de estimação dos parâmetros deste modelo, que baseia-se na conjugação do método de máxima verossimilhança com um processo iterativo tipo EM (Seção 22.3 de Hamilton, 1994). Este assunto será detalhado no Capítulo 4, onde serão apresentados alguns exemplos simples com o objetivo de facilitar a compreensão da metodologia por parte do leitor. Discutiremos também, no Capítulo 5, a implementação das rotinas de programação e simulação utilizando o software *R*, versão 1.7.1. Serão usadas simulações de Monte Carlo para verificar o desempenho da metodologia de estimação. Para finalizar o trabalho, analisaremos uma série temporal real, com dados do Produto Interno Bruto (GNP) dos E.U.A., coletados trimestralmente, desde o primeiro trimestre de 1947 até o primeiro trimestre de 1989.

Em anexo, apresentamos as rotinas criadas com a finalidade de obter as estimativas de máxima verossimilhança dos parâmetros de um modelo MS-AR, onde poderão ser encontradas todas as rotinas utilizadas para os três casos simulados e para a modelagem de um caso real.



## 2. CADEIAS DE MARKOV EM MODELOS MS-AR

Inicialmente, consideraremos o processo não observável  $s_t \{s_t, t \in Z^+\}$  como um processo a tempo discreto, onde o espaço dos estados  $S$  também é discreto. Assim, a variável latente representa o regime no qual o processo se encontra no tempo  $t$ . Assumiremos que tal processo é uma Cadeia de Markov (CM) de primeira ordem (neste caso conhecida como Cadeia de Markov à Parâmetro Discreto).

### 2.1. DEFINIÇÕES DE CADEIAS DE MARKOV

Seja  $S$  um conjunto finito que contenha todos os possíveis regimes do processo. Estes regimes são avaliados em determinados instantes temporais que formam um conjunto discreto no tempo. Chamaremos este conjunto de  $\tau$ .

Logo, temos que

$$S = \{ 1, 2, \dots, N \} \quad (2.1)$$

e

$$\tau = \{ 1, 2, \dots \}, \quad (2.2)$$

onde  $N$  representa o número de regimes.

Considere que a probabilidade do regime  $s_t$  assumir um particular valor  $j$ , onde  $j \in S$ , dependa apenas do valor assumido no instante  $t-1$ , ou seja, condicional à informação até o instante  $t-1$ , seja dada por:

$$P\{ s_t = j \mid s_{t-1} = i, s_{t-2} = k, \dots \} = P\{ s_t = j \mid s_{t-1} = i \} = p_t(i, j), \quad i, j \in S. \quad (2.3)$$

A equação (2.3) é chamada de propriedade markoviana, e assim  $s_t$  é uma CM (de primeira ordem). Em outras palavras, podemos dizer que, em processo markovianos, uma previsão para o instante seguinte depende apenas do presente, e não do passado, ou analogamente, uma previsão para o instante atual depende apenas do instante anterior ao atual. A probabilidade  $p_t(i, j)$  dada pela expressão acima é chamada de *probabilidade de transição em uma etapa do regime  $i$  para o regime  $j$  no tempo  $t$* . Neste trabalho, assumiremos esta probabilidade de transição estacionária ao longo do tempo, isto é,  $p_t(i, j) = p(i, j) = p_{ij}$ . Observe que

$$p_{i1} + p_{i2} + \dots + p_{iN} = 1 \quad (2.4)$$

Usualmente armazenamos as probabilidades de transição em uma matriz quadrada ( $N \times N$ ), conhecida como *matriz de transição* e denotada por  $P$ . Nesta matriz, a intersecção entre a linha  $j$  e a coluna  $i$  representa  $p_{ij}$ . A matriz  $P$  tem a seguinte forma:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}. \quad (2.5)$$

## 2.2. CADEIAS DE MARKOV NA FORMA DE VETOR AUTO-REGRESSIVO

Utilizaremos a notação  $\xi_t$  para representar o vetor aleatório de dimensão  $N \times 1$ , onde o  $j$ -ésimo elemento do vetor será igual a um se  $s_t = j$  e será igual a zero caso contrário. Sendo assim, a informação a respeito da realização dos estados da Cadeia de Markov será coletada em vetores,  $\xi_t$ ,  $t \in \tau$ , que se caracterizam por conter variáveis binárias que podem ser facilmente definidas por uma função indicadora assumindo apenas valores iguais a zero ou um:

$$\xi_t = \begin{bmatrix} I(s_t = 1) \\ \vdots \\ I(s_t = N) \end{bmatrix}, \quad \text{onde } I(s_t = j) = \begin{cases} 1, & \text{se } s_t = j \\ 0, & \text{em c. c.} \end{cases} \quad (2.6)$$

Esta representação também pode ser escrita da seguinte forma (Hamilton 1994, Seção 22.2):

$$\xi_t = \begin{cases} (1, 0, 0, \dots, 0) & \text{se } s_t = 1 \\ (0, 1, 0, \dots, 0) & \text{se } s_t = 2 \\ \vdots & \vdots \\ (0, 0, 0, \dots, 1) & \text{se } s_t = N \end{cases}. \quad (2.7)$$

Como consequência das definições acima podemos concluir que, dado que o processo encontra-se num regime qualquer  $i$  no tempo  $t$  ( $s_t = i$ ), então o  $j$ -ésimo elemento de  $\xi_{t+1}$  será uma variável aleatória que assume o valor um com probabilidade  $p_{ij}$  e assume o valor zero com probabilidade  $1 - p_{ij}$ . Portanto, o  $j$ -ésimo elemento de  $\xi_{t+1}$  será uma variável aleatória com esperança  $p_{ij}$ . Isto é, a esperança condicional de  $\xi_{t+1}$  dado que processo encontra-se no regime  $i$  no tempo  $t$  é dada por

$$E(\xi_{t+1} | s_t = i) = \begin{bmatrix} P_{i1} \\ P_{i2} \\ \vdots \\ P_{iN} \end{bmatrix}. \quad (2.8)$$

Como podemos claramente perceber, essa esperança condicional é exatamente igual a  $i$ -ésima coluna da matriz de transição  $P$ . Além disso, quando  $s_t = i$ , o vetor  $\xi_t$  pode ser representado pela  $i$ -ésima coluna da matriz identidade ( $I_N$ ), fazendo com que o vetor dado em (2.8) possa ser encontrado através do produto da matriz de transição  $P$  pelo vetor  $\xi_t$ , gerando a seguinte expressão:

$$E(\xi_{t+1} | \xi_t) = P\xi_t. \quad (2.9)$$

Note que pela propriedade markoviana temos que

$$E(\xi_{t+1} | \xi_t, \xi_{t-1}, \dots) = P\xi_t. \quad (2.10)$$

O resultado acima implica na possibilidade de podermos descrever uma Cadeia de Markov na forma

$$\xi_{t+1} = P\xi_t + v_{t+1}, \quad (2.11)$$

onde

$$v_{t+1} \equiv \xi_{t+1} - E(\xi_{t+1} | \xi_t, \xi_{t-1}, \dots). \quad (2.12)$$

A expressão (2.11) tem a forma de um vetor auto-regressivo de primeira ordem para  $\xi_t$ , conforme detalharemos na descrição de modelos ARIMA. Repare que (2.12) representa um erro aleatório de previsão e que pode ser obtido diretamente de (2.11). O vetor  $v_t$  pode assumir apenas um conjunto finito de valores, e sua esperança condicional é um vetor de zeros.

Da expressão (2.11) temos que

$$\xi_{t+m} = v_{t+m} + Pv_{t+m-1} + P^2v_{t+m-2} + \dots + P^{m-1}v_{t+1} + P^m\xi_t, \quad (2.13)$$

onde  $P^m$  indica a matriz de transição em  $m$  etapas, ou seja, nesta matriz encontraremos as probabilidades de transição de  $i$  para  $j$  após decorridas  $m$  unidades de tempo. Essa matriz pode ser encontrada pela multiplicação da matriz  $P$  por ela própria  $m$  vezes. Assim  $P\{s_{t+m} = j | s_t = i\}$  pode ser diretamente identificada através do elemento encontrado na  $j$ -ésima linha e  $i$ -ésima coluna da matriz  $P^m$ .

Portanto, uma previsão  $m$  passos a frente para uma Cadeia de Markov pode ser calculada através da seguinte expressão:

$$E(\xi_{t+m} | \xi_t, \xi_{t-1}, \dots) = P^m\xi_t. \quad (2.14)$$

O resultado da representação acima será um vetor de probabilidades de dimensão  $N$ , onde o  $j$ -ésimo elemento do vetor indicará a probabilidade de  $s_{t+m}$  estar no regime  $j$ , dado o atual regime  $s_t$ .

### 2.3. CADEIAS DE MARKOV IRREDUTÍVEIS

Uma Cadeia de Markov é dita *irredutível* quando quaisquer dois regimes de  $S$  se comunicam, isto é, quando  $S$  é uma classe comunicante. Dizemos que dois regimes quaisquer de  $S$ , digamos  $i$  e  $j$ , se comunicam se, e somente se,  $i$  conduz a  $j$  ( $i \rightarrow j$ ) e  $j$  conduz a  $i$  ( $j \rightarrow i$ ), ou seja,  $i \leftrightarrow j$ . Isto quer dizer que existe uma probabilidade positiva de que o processo passe do regime  $i$  para o regime  $j$  e vice-versa.

Para facilitar a compreensão deste conceito, suponha uma Cadeia de Markov simples com apenas dois regimes e com a seguinte matriz de transição:

$$P = \begin{bmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{bmatrix}. \quad (2.15)$$

Vemos que, se  $p_{11}=1$ , o processo jamais sairá do regime 1 uma vez tendo ingressado neste regime. Neste caso, concluímos que o regime 1 é absorvente e esta Cadeia de Markov não é irredutível (ou é redutível), pois os regimes 1 e 2 não se comunicam. Portanto, concluímos que um regime qualquer  $i$ ,  $i \in S$ , é considerado *absorvente* quando a probabilidade de transição do regime  $i$  para o regime  $j$  é igual a zero, ou seja,  $p_{ij} = 0$ ,  $i \neq j$ .

No caso em que um particular processo apresenta probabilidade um de retornar a um determinado regime  $i$   $\{ i \in S \}$  em algum instante futuro, dado que saiu de  $i$ , dizemos que o regime  $i$  é *recorrente*, ou seja, no caso de  $i$  ser recorrente, o processo certamente voltará a ingressar no regime  $i$ . Se houver probabilidade positiva de que o processo jamais possa retornar ao regime  $i$ , dizemos que o regime  $i$  é *transiente* (ou transitório).

Generalizando a idéia de uma Cadeia de Markov com 2 regimes para uma Cadeia de Markov com  $N$  regimes ( $N > 2$ ), verificamos que esta apenas será redutível quando a matriz de transição for triangular conforme apresentado abaixo:

$$P = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}. \quad (2.16)$$

## 2.4. CADEIAS DE MARKOV COM REGIMES ERGÓDICOS

Estes tipos de Cadeias de Markov são caracterizados pela presença de probabilidades ergódicas, que são probabilidades associadas a cada regime no estado de equilíbrio. Mais adiante veremos como são calculadas essas probabilidades, utilizando autovalores e autovetores.

Um regime qualquer é considerado ergódico quando seu tempo médio de recorrência é finito e quando seu *período* é igual a um. Um regime  $i$  que apresenta tempo médio de recorrência finito é chamado de regime *recorrente positivo*. Caso o regime  $i$  não apresente tempo médio de recorrência finito, mas apresenta probabilidade um de retornar ao regime  $i$  em algum instante futuro, será chamado de regime *recorrente nulo*. Entendemos por *tempo médio de recorrência* o tempo esperado (ou médio) que o processo leva para retornar a um determinado regime  $i$ . Este tempo médio é representado por  $\mu_i$  e é calculado pela seguinte fórmula:

$$\mu_i = \sum_{n=1}^{\infty} n \cdot P^n(i, i), \quad (2.17)$$

onde  $n$  representa o número de etapas do processo e  $P^n(i, i)$  representa a probabilidade de o processo retornar ao regime  $i$  após decorridas  $n$  etapas.

O período de um regime qualquer  $i$  é representado por

$$\text{per}(i) = \text{m.d.c.} \{ n > 0 : P^n(i, i) > 0 \}, \quad (2.18)$$

que nada mais é do que o máximo divisor comum entre todos os possíveis tempos de retorno ao regime  $i$ . Quando o período do regime  $i$  for igual a um, dizemos que este regime é *aperiódico*. Maiores detalhes a respeito de Cadeias de Markov com probabilidades ergódicas podem ser encontradas em Hamilton (1994, Seção 22.2).

Considerando uma Cadeia de Markov com apenas dois regimes, temos que os autovalores da matriz de transição serão obtidos pela solução de  $|P - \lambda I| = 0$ , onde  $I$  é a matriz identidade de dimensão  $2 \times 2$ . Resolvendo a equação, teremos que

$$\begin{vmatrix} p_{11} - \lambda & 1 - p_{22} \\ 1 - p_{11} & p_{22} - \lambda \end{vmatrix} = 0,$$

e, realizando-se as operações adequadas, chegaremos ao seguinte resultado:

$$(\lambda - 1)(\lambda - p_{11} - p_{22} + 1) = 0. \quad (2.19)$$

Assim, encontramos os autovalores associados à matriz de transição, dados por

$$\lambda_1 = 1 \text{ e } \lambda_2 = p_{11} + p_{22} - 1. \quad (2.20)$$

Do resultado obtido em (2.20) encontramos que os autovetores associados a  $\lambda_1$  e  $\lambda_2$  são, respectivamente,

$$\begin{bmatrix} \frac{(1-p_{22})}{2-p_{11}-p_{22}} \\ \frac{(1-p_{11})}{2-p_{11}-p_{22}} \end{bmatrix} \text{ e } \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \quad (2.21)$$

Os autovetores de (2.21) são obtidos resolvendo-se a equação  $Px = \lambda x$ . Maiores detalhes para obtenção de autovalores e autovetores podem ser obtidos em Lay (1999).

Portanto, para um modelo com dois regimes, teríamos que as probabilidades ergódicas de cada regime seriam respectivamente dadas por

$$p_1 = \frac{(1-p_{22})}{2-p_{11}-p_{22}} \text{ e } p_2 = \frac{(1-p_{11})}{2-p_{11}-p_{22}}, \quad (2.22)$$

onde  $p_1$  representa  $P\{s_t=1\}$  e  $p_2$  representa  $P\{s_t=2\}$ , ou seja, as probabilidades de uma determinada observação estar respectivamente nos regimes 1 ou 2.

## 2.5. TEMPO DE PERMANÊNCIA NOS REGIMES

Os modelos com Mudança de Regime também podem fornecer outra importante informação, que chamaremos de tempo médio de permanência em cada regime (ou duração média em cada regime, ou ainda, proporção de tempo que o processo ocupa determinado regime). Denotaremos por  $D$  o tempo de permanência em cada regime, o qual será definido por:

$$\begin{aligned} D = 1, \text{ se } s_t = i \text{ e } s_{t+1} = j; P[D=1] &= (1 - p_{ii}), \\ D = 2, \text{ se } s_t = s_{t+1} = i \text{ e } s_{t+2} = j; P[D=2] &= p_{ii} (1 - p_{ii}), \\ D = 3, \text{ se } s_t = s_{t+1} = s_{t+2} = i \text{ e } s_{t+3} = j; P[D=3] &= p_{ii}^2 (1 - p_{ii}), \\ &\vdots \end{aligned} \quad (2.23)$$

onde  $i \neq j$ .

Verifique que as probabilidades  $P[D=i]$  apresentam uma distribuição de probabilidades Geométrica. Então, para encontrarmos o tempo de permanência esperado em cada regime, temos que

$$E(D) = \sum_{i=1}^{\infty} iP[D=i] = \frac{1}{1-p_{ii}}. \quad (2.24)$$

### 3. INTRODUÇÃO AOS MODELOS ARIMA E METODOLOGIA DE BOX-JENKINS

Neste capítulo, enfatizaremos a descrição de modelos para séries temporais chamados ARIMA (Auto-Regressivo e Médias Móveis Integrado) e a metodologia de identificação destes modelos sugerida por Box e Jenkins (1976). Porém, o foco principal em estudo neste capítulo serão os modelos AR, que podem ser considerados como uma derivação dos modelos ARIMA contendo apenas os parâmetros auto-regressivos do modelo, conforme será explicitado adiante.

Uma parcela considerável de pesquisas classificadas como metodologia no domínio do tempo tem por objetivo final a construção de estruturas adequadas para que se possa realizar previsões de valores futuros de uma determinada série temporal. Sendo assim, o interesse principal consiste em encontrar algum modelo que seja capaz de relacionar os valores atuais de uma série temporal a um conjunto de valores do passado.

Conforme Morettin e Toloí (1981), séries temporais são compostas basicamente por quatro componentes (não necessariamente todos estes componentes aparecem numa mesma série temporal), que estão definidos a seguir:

- *tendência*: verifica o sentido do deslocamento da série ao longo do tempo;
- *ciclo*: caracteriza-se por um movimento ondulatório que tende a ser periódico com o passar do tempo;
- *sazonalidade*: representada por movimentos ondulatórios de curta duração;
- *ruído aleatório ou erro*: compreende a parte da série que não pode ser modelada e representa a variabilidade intrínseca aos dados.

Pelo fato de uma série temporal apresentar observações coletadas seqüencialmente ao longo do tempo, esperamos que estas observações sejam correlacionadas. Os modelos ARIMA são modelos matemáticos que têm por objetivo captar o comportamento desta correlação ou autocorrelação entre os valores de uma determinada série temporal, e realizar previsões com base neste comportamento. Nos casos em que essa estrutura de correlação for bem modelada, fornecerá boas previsões.

Modelos ARIMA podem ser considerados como resultantes da combinação de três componentes (ou filtros): um componente auto-regressivo (AR), um filtro de

integração (I) e um componente de médias móveis (MA). Salientamos que uma série temporal não precisa ser modelada por todos os três componentes citados acima, podendo resultar em vários modelos. Como citado anteriormente, o foco deste trabalho será dado a parte AR do modelo.

Segundo a metodologia de Box-Jenkins, ajustamos modelos ARIMA seguindo passos que podem ser repetidos, se for necessário, durante o processo de análise dos dados. Esta metodologia modela um determinado conjunto de dados captando o máximo possível dos padrões nele existentes. Posteriormente descreveremos estes passos.

### 3.1. MODELOS ESTACIONÁRIOS

Processos que se encontram em estado de “equilíbrio” são ajustados por modelos denominados *modelos estacionários*. Classificaremos estes modelos em fracamente estacionários e fortemente estacionários.

Quando nem a média e nem a função de autocovariância dependem do tempo  $t$ , o processo  $y_t$  é chamado de fracamente estacionário. Isto ocorre quando obtemos

$$E(y_t) = \mu, \quad (3.1)$$

$$E(y_t - \mu)(y_{t-j} - \mu) = \gamma_j, \text{ para todo } t \text{ e um } j \text{ qualquer}, \quad (3.2)$$

onde  $\gamma$  representa a função de autocovariância.

Um processo é dito fortemente ou estritamente estacionário quando, para qualquer valor de  $j_1, j_2, \dots, j_n$ , a distribuição conjunta de  $(y_t, y_{t+j_1}, \dots, y_{t+j_n})$  depende apenas do intervalo que separa as datas  $(j_1, j_2, \dots, j_n)$  e não do tempo  $t$ , isto é, o vetor  $(y_t, y_{t+j_1}, \dots, y_{t+j_n})$  tem a mesma distribuição conjunta de  $(y_s, y_{s+j_1}, \dots, y_{s+j_n})$ , para todo  $s$  e  $t$ .

Maiores detalhes sobre estacionariedade e as funções de autocovariância podem ser encontrados em Shummway (1988).

#### 3.1.1. MODELOS SARIMA

Freqüentemente, existem componentes periódicos em uma série que tendem a se repetir com um período de  $S$  unidades de tempo. Isto significa que o comportamento



estrutural subjacente do processo pode ser descrito através de uma série indexada em incrementos de algum período sazonal subjacente  $S$ . Nestes casos poderíamos desejar ajustar as equações de diferenças em múltiplos desse comportamento sazonal. Estas equações poderiam incorporar o operador defasagem sazonal que é dado por

$$B^S y_t = y_{t-S}, \quad (3.3)$$

ou o operador diferença sazonal

$$\Delta_S^D = (1 - B^S)^D. \quad (3.4)$$

Também podemos definir os componentes sazonais de ordem  $P$  e as médias móveis sazonais de ordem  $Q$ , respectivamente, através de

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_{PS} B^{PS}, \text{ e} \quad (3.5)$$

$$\Theta(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{2S} + \dots + \Theta_Q B^{QS}. \quad (3.6)$$

Portanto, um modelo completo (autorregressivo e médias móveis com Sazonalidade) pode ser dado por

$$\Phi(B^S)\theta(B)\Delta_S^D\Delta^d y_t = c + \Theta_Q(B^S)\theta(B)\varepsilon_t \quad (3.7)$$

Representamos o modelo multiplicativo apresentado em (3.7) da forma SARIMA(p,d,q)<sub>X</sub>(P,D,Q)<sub>S</sub>.

Quando um processo  $\varepsilon_t$  apresentar média  $\mu$  e variância  $\sigma_\varepsilon^2$ , para todo  $t$ , e covariância igual a zero entre  $\varepsilon_s$  e  $\varepsilon_t$ , para todo  $s \neq t$ , dizemos que este processo é um ruído branco (RB). A seguir, estudaremos um caso particular de modelos SARIMA(p,d,q)<sub>X</sub>(P,D,Q)<sub>S</sub>, os modelos AR(p).

### 3.1.2. MODELO AUTORREGRESSIVO – AR

Nos modelos autorregressivos, a série de dados históricos  $y_t$  é descrita através de seus valores passados e pelo erro (ruído) aleatório  $\varepsilon_t$ . Sendo assim, um modelo autorregressivo de ordem  $p$  – AR(p) é escrito na forma:

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \dots + \phi_p \tilde{y}_{t-p} + \varepsilon_t, \quad (3.8)$$

onde  $\tilde{y}_t$  representa a diferença entre o valor do processo no instante  $t$ ,  $y_t$ , e sua média,  $\mu$ ,  $\phi_i$  descreve a taxa de variação de  $\tilde{y}_t$  por unidade de  $\tilde{y}_{t-i}$ ,  $i = 1, 2, p$ , e  $\varepsilon_t$  é um ruído branco com  $E[\varepsilon_t] = 0$  e  $E[\varepsilon_t^2] = \sigma^2$ .

Há ainda uma outra maneira de descrevermos o modelo apresentado em (3.8), fazendo-se uso de um operador, chamado de operador de defasagem e representado pelo símbolo  $L$ . Através do operador de defasagem, poderemos reescrever o modelo apresentado em (3.8) na forma

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) \tilde{y}_t = \phi(L) \tilde{y}_t = \varepsilon_t, \quad (3.9)$$

onde

$$L^m \tilde{y}_t = \tilde{y}_{t-m}. \quad (3.10)$$

Um modelo AR(p) é igual a um modelo ARIMA(p,0,0), onde inexistem os componentes de médias móveis (MA) e de integração (I).

Para compreender os modelos AR(p), suponha o modelo mais simples desta classe de modelos: o AR(1). Conforme apresentado em (3.8), um modelo AR(1) pode ser escrito como:

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \varepsilon_t. \quad (3.11)$$

A condição necessária e suficiente para o modelo acima é  $|\phi_1| < 1$ .

No caso do modelo AR(1), as autocovariâncias serão dadas por

$$\gamma_j = \phi_1^j \gamma_0, \quad (3.12)$$

enquanto as autocorrelações serão dadas por

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi_1^j. \quad (3.13)$$

Quando  $\phi_1$  apresentar um valor positivo, a função de autocorrelação decairá exponencialmente. Por sua vez, quando  $\phi_1$  apresentar um valor negativo, a função de autocorrelação também decairá exponencialmente, porém, apresentará alternância de sinais positivos e negativos.

### 3.2. FUNÇÃO DE AUTOCORRELAÇÃO AMOSTRAL (ACF)

Considere uma série temporal que contenha os valores  $y_1, y_2, \dots, y_T$ . Para esta série, teremos que a função de autocorrelação no lag  $m$  será dada por:

$$r_m = \frac{\sum_{t=1}^{T-m} (y_t - \bar{y})(y_{t+m} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (3.14)$$

onde  $r_m$  varia entre  $-1$  e  $1$ , e  $\bar{y}$  é representado por:

$$\bar{y} = \frac{\sum_{t=1}^T y_t}{T}. \quad (3.15)$$

Verifique que o valor encontrado para  $r_m$  mede o relacionamento linear entre as observações da série separadas por uma diferença (*lag*) de  $m$  unidades de tempo. À medida que o valor de  $r_m$  se aproxima de  $1$ , dizemos que o relacionamento entre as observações é forte positivo. Porém, quando se aproxima de  $-1$ , dizemos que o relacionamento entre as observações é forte negativo. Valores de  $r_m$  próximos de zero indicam que não há relacionamento entre as observações.

Se  $y_t$  é um ruído branco, então  $r_m$  é assintoticamente normal com média zero e  $Var[r_m] \equiv T^{-1}$ , para  $m \neq 0$  (Anderson 1971, p. 478). Assim a significância de picos na autocorrelação amostral pode ser determinada verificando se  $r_m$  está fora do intervalo compreendido por

$$\pm Z_{\alpha/2} / \sqrt{T}, \quad (3.16)$$

onde  $Z_{\alpha/2}$  é o valor de  $Z$  da distribuição normal padrão tal que

$$P(|Z| > Z_{\alpha/2}) = \alpha. \quad (3.17)$$

### 3.3. FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL AMOSTRAL (PACF)

A função de autocorrelação parcial é analisada e interpretada de forma análoga a função de autocorrelação descrita na seção anterior.

Procederemos da seguinte maneira para encontrar a função de autocorrelação parcial (PACF):

$$r_{mm} = \begin{cases} r_1 & \text{se } m = 1 \\ \frac{r_m - \sum_{j=1}^{m-1} r_{m-1,j} r_{m-j}}{1 - \sum_{j=1}^{m-1} r_{m-1,j} r_{m-j}} & \text{se } m = 2, 3, \dots \end{cases} \quad (3.18)$$

onde  $r_{mj} = r_{m-1,j} - r_{mm} r_{m-1,m-j}$ , para  $m = 1, 2, \dots, m-1$ .

Se  $y_t$  é um AR(p), então  $r_{mm}$ ,  $m > p$ , é assintoticamente normal com média zero e  $Var[r_{mm}] \equiv T^{-1}$ , para  $m \neq 0$ . Assim a significância de picos na autocorrelação parcial pode ser determinada verificando se  $r_{mm}$  está fora do intervalo compreendido por

$$\pm Z_{\alpha/2} / \sqrt{T}, \quad (3.19)$$

onde  $Z_{\alpha/2}$  é o valor de  $Z$  da distribuição normal padrão tal que

$$P(|Z| > Z_{\alpha/2}) = \alpha. \quad (3.20)$$

Bowerman & O'Connell (1993, Seção 9.2) apresentam maiores detalhes a respeito das funções de autocorrelação parcial.

### 3.4. MODELOS NÃO-ESTACIONÁRIOS

Quando uma série temporal apresenta média e/ou autocovariância dependentes do tempo, dizemos que a mesma é *não-estacionária*. Quando a causa da não-estacionariedade é a heterocedasticidade, transformações tipo Box-Cox são sugeridas à série original. Já quando uma série temporal não é estacionária devido a média, Box e Jenkins (1976) sugerem aplicar sucessivas diferenças até que a série se torne estacionária. Em geral, uma ou duas diferenças já são suficientes para alcançar esse objetivo.

Há também testes para a raiz unitária, como por exemplo Dickey-Fuller (1979). Estes testes verificam se o processo gerador é um passeio aleatório, isto é, se o processo pode ser descrito por

$$y_t = y_{t-1} + \varepsilon_t, \quad (3.21)$$

como pode ser visto em Enders 1995, Capítulo 4. Outros tipos de modelos vastamente estudados na literatura recente são os modelos ARFIMA (ver Reisen, 2000, por exemplo).

### 3.5. METODOLOGIA BOX-JENKINS

O ajustamento de modelos SARIMA segundo Box Jenkins (1976) baseia-se num ciclo iterativo, onde, a partir dos dados, é feita a escolha do modelo apropriado. Box &

Jenkins (1976) apresentam três etapas distintas para a construção do modelo: *identificação, estimação e verificação*.

Na etapa de identificação, deseja-se descobrir qual dentre os possíveis modelos melhor descreve os dados. A identificação do modelo se faz a partir da análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) amostrais. Comparam-se estas funções àquelas teóricas para possíveis modelos SARIMA.

Já na fase de estimação, estaremos interessados em estimar os parâmetros do modelo SARIMA. Também será preciso estimar a variância do erro aleatório  $\varepsilon_t$ . Os principais métodos utilizados são Mínimos Quadrados e Máxima Verossimilhança.

A última etapa, de verificação, consiste em avaliar a adequabilidade do modelo ajustado, ou seja, verificar se o modelo estimado é adequado para descrever o comportamento existente nos dados. Isto ocorre através da análise da ACF e PACF dos resíduos. Caso aceite-se estatisticamente que estas funções podem ser as de um ruído branco, então o modelo é considerado bem ajustado.

Caso se julgue necessário, o ciclo acima descrito pode ser repetido, voltando-se à fase inicial, de identificação. Só faz sentido repetirmos este ciclo se o modelo ajustado não for adequado, ou se outros modelos pudessem ter sido sugeridos em alguma fase do processo de identificação. Um procedimento bastante utilizado consiste na estimação e verificação de vários modelos. Quando for escolhido um modelo considerado satisfatório, através de critérios como AIC e BIC (ver Shummway, 1988), por exemplo, podemos então usá-lo para *previsão*. Esta etapa é muito importante, pois consiste na obtenção de previsões para valores futuros de uma determinada série temporal.

### 3.6. CRITÉRIOS PARA A ESCOLHA DE MODELOS

Quando desejamos verificar qual dos modelos dentre aqueles elegíveis se ajusta melhor aos dados, podemos fazer uma comparação com base no critério de Akaike (AIC – *Akaike's Information Criterion*) ou no critério de Schwarz (BIC – *Bayesian Information Criterion*), que pode ser encontrado com base no estimador de máxima verossimilhança da variância residual. A estatística de AIC pode ser encontrada por

$$AIC(K) = \ln(\hat{\sigma}_\varepsilon^2) + 2K/T, \quad (3.22)$$

onde  $\hat{\sigma}_\varepsilon^2$  representa a variância dos resíduos,  $K$  é o número de parâmetros do modelo e  $T$  é o tamanho da amostra (ver Shummway 1988, pág 154).

A estatística BIC pode ser encontrada por

$$BIC(K) = \ln(\hat{\sigma}_\varepsilon^2) + KT^{-1} \ln T \quad (3.23)$$

Quando comparamos dois modelos através dos critérios AIC ou BIC, consideramos como o melhor modelo aquele que apresentar o menor AIC ou BIC. Como estes critérios consideram o princípio de parcimônia, acabam penalizando modelos que apresentam muitos parâmetros. Assim, vemos em (4.22) e (4.23) que os valores de AIC e BIC aumentam a medida que aumenta o número de parâmetros do modelo.

#### 4. METODOLOGIA MS-VAR

Os modelos MS-VAR podem ser considerados como combinação dos modelos VAR(p) (*Vetor auto-regressivo de ordem p*) com a idéia de mudança de regime baseada nas Cadeias de Markov.

Entretanto, neste trabalho abordaremos apenas a modelagem univariada, isto é, para uma única série de dados observados ao longo do tempo, e não um vetor de séries, o que simplificará o modelo, podendo ser chamado de modelo MS-AR. Isto reduzirá a dimensão do vetor auto-regressivo de  $K$  (que representa o número de séries temporais sob análise) para um, pois não estaremos mais interessados em analisar simultaneamente diferentes séries de tempo, mas apenas uma.

Neste capítulo apresentaremos uma descrição teórica sobre os modelos auto-regressivos com mudança markoviana de regime (MS-AR), abordando como caso particular um modelo de mistura de distribuições normais. Também descreveremos muito brevemente modelos de mudança de regime na variância. Além disso, para os modelos autoregressivos estudaremos a estimação dos parâmetros. Nesse sentido, discutiremos a obtenção dos *estimadores de máxima verossimilhança* dentro do procedimento iterativo do tipo *EM*.

##### 4.1. MODELOS DE MISTURA DE DISTRIBUIÇÕES NORMAIS

Esta seção apresentará o caso mais simples dos modelos com mudança de regime, o de uma mistura de distribuições normais independentes e identicamente distribuídas. O conhecimento destes conceitos será utilizado posteriormente.

Modelos de mistura de normais apresentam mudança de regime na média. Portanto, nestes modelos, a série  $y_t$  é modelada através da seguinte equação:

$$y_t = \mu_{s_t} + \varepsilon_t, \quad (4.1)$$

onde  $\mu_{s_t}$  informa a média da série em cada regime e  $\varepsilon_t$  é o erro aleatório do modelo, para o qual assume-se distribuição normal com média igual a zero e variância  $\sigma_{s_t}^2$  (variância em cada regime).

Supondo que a série apresente  $N$  regimes, então (4.1) indica que, por exemplo,  $y_t \sim N(\mu_1, \sigma_1^2)$  caso o processo se encontre no regime 1, e assim sucessivamente até o regime  $N$ , onde  $y_t \sim N(\mu_N, \sigma_N^2)$ .

De acordo com Hamilton (1994, Seção 22.3) uma mistura de distribuições normais pode produzir distribuições “semelhantes” a uma distribuição normal ou distribuições com densidades mais complexas com mais de uma moda, diferentes assimetrias e curtoses.

A densidade condicional de  $y_t$  dado o regime  $s_t = j$  é

$$f(y_t | s_t = j; \theta) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\}, \quad (4.2)$$

para  $j = 1, 2, \dots, N$  (pois temos  $N$  regimes), onde  $\theta$  é um vetor que contém os parâmetros de todos os regimes da população  $\mu_1, \dots, \mu_N$  e  $\sigma_1^2, \dots, \sigma_N^2$ , além das probabilidades ergódicas  $p_1, \dots, p_N$ , definidas por:

$$P(s_t = j; \theta) = p_j. \quad (4.3)$$

As probabilidades apresentadas na expressão (4.3) são as mesmas probabilidades ergódicas que foram apresentadas na Seção 2.4.

Sendo assim, temos que

$$g(y_t, s_t; \theta) = f(y_t | s_t = j; \theta) \cdot P(s_t = j; \theta), \quad (4.4)$$

representa a função de probabilidade conjunta de  $y_t$  e  $s_t$ , a qual, substituindo-se as expressões (4.2) e (4.3) na expressão (4.4), é dada por

$$g(y_t, s_t; \theta) = \frac{p_j}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\}. \quad (4.5)$$

Quando realizarmos a soma da expressão (4.5) para todos os possíveis regimes, encontraremos a função densidade marginal de  $y_t$ , dada por

$$f(y_t; \theta) = \sum_{j=1}^N P(y_t, s_t = j; \theta) = \sum_{j=1}^N \frac{p_j}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}\right\}. \quad (4.6)$$

A partir de (4.6) poderemos facilmente encontrar a função de log-verossimilhança, cuja expressão é

$$L(\theta) = \sum_{t=1}^T \log(f(y_t; \theta)), \quad (4.7)$$

onde  $T$  é o total de pontos da série.



Obteremos os estimadores de máxima verossimilhança de  $\theta$  através da maximização, em relação a  $\theta$ , de  $L(\theta)$ . Os procedimentos para a maximização serão discutidos adiante, pois tais procedimentos baseiam-se em métodos numéricos.

## 4.2. MODELOS AUTOREGRESSIVOS

Começaremos a abordagem através de modelos simples, que posteriormente poderão ser estendidos e aprofundados.

Inicialmente considere o modelo auto-regressivo de primeira ordem apresentado a seguir, onde o valor da série para um determinado instante depende apenas da observação no instante anterior:

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t, \quad (4.8)$$

onde  $\varepsilon_t$  é normalmente distribuído com média zero e variância  $\sigma^2$ ,  $\phi_{s_t}$  é o parâmetro do componente auto-regressivo no regime  $s_t$ ,  $c_{s_t}$  é a constante no regime  $s_t$ , e  $\{s_t, t \in Z^+\}$  representa um processo estocástico latente (não observável) com espaço de estados (ou regimes) discreto  $S = \{1, 2, \dots, N\}$ . Note que teremos uma equação para cada regime da série, ou seja, teremos  $N$  equações.

De acordo com o modelo proposto em (4.8) a série apresenta mudança de regime. Dependendo do regime atual  $s_t$ , os parâmetros do modelo auto-regressivo serão distintos (eventualmente poderão ser iguais). Isto é, serão variantes no tempo. Porém, quando condicionados à variável não observável  $s_t$ , passarão a ser invariantes no tempo, pois  $s_t$  indica o regime em que a série se encontra em determinado instante. Com isto, como dito anteriormente, passaremos a ter uma diferente equação do nosso modelo para cada regime  $s_t$ . Logo, se tivermos  $N$  regimes, teremos a seguinte densidade de probabilidade condicional para  $y_t$ , dado o regime:

$$f(y_t | y_{t-1}, s_t) = \begin{cases} f(y_t | y_{t-1}, \theta_1), se s_t = 1 \\ \vdots \\ f(y_t | y_{t-1}, \theta_N), se s_t = N \end{cases}, \quad (4.9)$$

onde  $\theta_n$  é um vetor contendo os parâmetros do modelo AR em cada um dos regimes  $n = 1, 2, \dots, N$ .

Uma vez que já definimos que os parâmetros do modelo dependem do regime (aleatório e não observável), precisamos agora elaborar uma hipótese a respeito do comportamento estocástico de  $s_t$ . Deste modo poderemos derivar a densidade marginal de  $y_t$  e, conseqüentemente, a função log-verossimilhança para a estimação dos parâmetros.

Em modelos MS-AR assume-se que o processo  $s_t$  segue uma Cadeia de Markov Ergódica Irredutível a tempo e espaço de estados discretos, onde as probabilidades de transição serão dadas por  $p_{ij}$ . As definições a respeito das cadeias de Markov podem ser vistas no Capítulo 2.

Vamos imaginar que uma determinada série em estudo apresente  $N$  regimes. Se, num determinado instante  $t$ , observarmos que  $s_t = j$ , a densidade condicional de  $y_t$  será dada por

$$f(y_t | s_t = j, y_{t-1}; \lambda) \quad (4.10)$$

onde  $\lambda = (c_1, \dots, c_N, \phi_1, \dots, \phi_N, \sigma^2, p_{ij})$  é um vetor de parâmetros desconhecidos que caracterizam a densidade condicional. Se a série apresentar  $N$  regimes distintos, teremos então  $N$  diferentes densidades como definido em (4.10), sendo uma densidade para cada regime. Estas  $N$  densidades farão parte de um vetor de dimensão  $N \times 1$ , representado por  $\eta_t$ .

Assumimos em (4.10) que a densidade condicional depende apenas do atual regime  $s_t$ . Caso a série apresente dois regimes ( $N = 2$ ), o vetor com as densidades condicionais de  $y_t$  será dado por

$$\eta_t = \begin{bmatrix} f(y_t | s_t = 1, y_{t-1}; \theta) \\ f(y_t | s_t = 2, y_{t-1}; \theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-(y_t - c_1 - \phi_1 y_{t-1})^2}{2\sigma^2} \right\} \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-(y_t - c_2 - \phi_2 y_{t-1})^2}{2\sigma^2} \right\} \end{bmatrix}. \quad (4.11)$$

Explorando mais este exemplo, supondo que  $y_t$  depende de  $s_t^*$  e  $s_{t-1}^*$ , através da especificação do modelo

$$y_t - \mu_{s_t^*} = \phi_{s_t^*} (y_{t-1} - \mu_{s_{t-1}^*}) + \varepsilon_t, \quad (4.12)$$

onde  $s_t^*$  é uma cadeia de Markov com dois estados, teremos a função densidade condicional apresentada abaixo:

$$f(y_t | I_t, s_t^* = j, s_{t-1}^* = i, s_{t-2}^* = l, \dots; \theta) = f(y_t | I_t, s_t = j, s_{t-1} = i; \theta), \quad (4.13)$$

onde  $I_t$  representa toda a informação acumulada até o instante  $t$ .

De modo a termos uma representação como em (4.10), podemos definir

$$s_t = \begin{cases} 1 & \text{se } s_t^* = 1 \text{ e } s_{t-1}^* = 1 \\ 2 & \text{se } s_t^* = 2 \text{ e } s_{t-1}^* = 1 \\ 3 & \text{se } s_t^* = 1 \text{ e } s_{t-1}^* = 2 \\ 4 & \text{se } s_t^* = 2 \text{ e } s_{t-1}^* = 2 \end{cases} \quad (4.14)$$

Assim passaremos a ter 4 regimes para  $s_t$ . Portanto, a nova matriz de transição para  $s_t$ , com quatro regimes, é dada por

$$P = \begin{bmatrix} p_{11}^* & 0 & p_{11}^* & 0 \\ p_{12}^* & 0 & p_{12}^* & 0 \\ 0 & p_{21}^* & 0 & p_{21}^* \\ 0 & p_{22}^* & 0 & p_{22}^* \end{bmatrix}, \quad (4.15)$$

onde  $P\{s_t^* = j | s_{t-1}^* = i\} = p_{ij}^*$ .

Agora, podemos reescrever a densidade condicional encontrada em (4.13) em quatro densidades:

$$\begin{aligned} f(y_t | y_{t-1}, s_t = 1; \theta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[(y_t - \mu_1) - \phi_1(y_{t-1} - \mu_1)]^2}{2\sigma^2}\right\} \\ f(y_t | y_{t-1}, s_t = 2; \theta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[(y_t - \mu_2) - \phi_2(y_{t-1} - \mu_1)]^2}{2\sigma^2}\right\} \\ f(y_t | y_{t-1}, s_t = 3; \theta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[(y_t - \mu_1) - \phi_3(y_{t-1} - \mu_2)]^2}{2\sigma^2}\right\} \\ f(y_t | y_{t-1}, s_t = 4; \theta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[(y_t - \mu_2) - \phi_4(y_{t-1} - \mu_2)]^2}{2\sigma^2}\right\} \end{aligned} \quad (4.16)$$

Portanto, o modelo (4.12) pode ser reescrito da seguinte maneira:

$$y_t - \mu_{s_t} = \phi_{s_t} (y_{t-1} - \mu_{s_t}) + \varepsilon_t, \quad (4.17)$$

onde  $\mu_{s_t}$  representa a média no regime  $s_t$ .

Vemos que em (4.16) temos um modelo MS-AR ajustado à média. Note que a média e os parâmetros auto-regressivos do modelo estão condicionados ao regime estocástico não observável  $s_t$ .

### 4.3. MODELOS DE MUDANÇA DE REGIME NA VARIÂNCIA

Além dos modelos de mistura de normais e dos modelos auto-regressivos, existe ainda uma outra forma de tentar controlar as quebras estruturais, que são os modelos de mudança de regime markoviana para a variância condicional (ver, por exemplo, Almeida e Pereira, 2002). Nesse caminho, uma possível abordagem é considerar uma extensão dos modelos ARCH (*Autoregressive Conditional Heteroskedasticity*) e GARCH (*Generalized ARCH*). Tais modelos tratam a variância condicional como uma variável aleatória. Bollerslev, Engle e Nelson (1994) apresentam uma excelente revisão sobre estes modelos. Ziegelmann (2002) também revisa estes modelos, ajustando-os às séries financeiras BOVESPA e S&P500.

Representamos um modelo GARCH por

$$y_t = Z_t' b + u_t, \quad (4.18)$$

onde  $Z_t$  é o vetor de regressores da série  $y_t$  e  $b$  é um vetor contendo os coeficientes da regressão. A variável aleatória  $u_t$  é dada por

$$u_t = \sqrt{h_t} v_t, \quad (4.19)$$

onde  $v_t$  é i.i.d. e  $h_t$  será dado por

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \quad (4.20)$$

para  $\alpha_0 > 0$ ,  $\alpha_i, \beta_j \geq 0$ ,  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ .

De acordo com o modelo acima, no caso em que  $\beta_j = 0, j = 1, 2, \dots, q$ , teremos um modelo ARCH(p). No caso geral, o modelo terá a forma de um GARCH(p, q).

Generalizando-se o modelo descrito de (4.18) a (4.20) para modelos com mudança de regime na volatilidade condicional, teremos que

$$y_t = Z_t' \beta + u_t, \quad (4.21)$$

$$u_t = \sqrt{g_{s_t}} \tilde{u}_t, \quad (4.22)$$

$$\tilde{u}_t = \sqrt{h_t} \varepsilon_t, \quad (4.23)$$

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \tilde{u}_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \quad (4.24)$$

originando diferentes regimes para a volatilidade condicional. Quando aplicamos estes modelos, estamos interessados em modelar as mudanças de regime como alterações na

escala do processo da variância condicional. Chamaremos o modelo descrito de (4.21) a (4.24) de SWGARCH( $N, p, q$ ), onde  $N$  representa o número de regimes do modelo.

#### 4.4. ESTIMAÇÃO DOS PARÂMETROS DE MODELOS AUTORREGRESSIVOS

Inicialmente nos preocuparemos em encontrar a densidade condicional de  $y_t$ . Seja  $\xi_t$  como definido na Seção 2.2. O modelo para a série temporal apresenta um erro aleatório  $\varepsilon_t$ , que consideramos ter distribuição Normal. Nestas condições  $y_t$  terá a seguinte densidade condicional:

$$f(y_t | \xi_t = \iota_j, y_{t-1}; \theta) = \ln(2\pi)^{-\frac{1}{2}} \ln(\sigma_j)^{-1} \exp\left\{ -\frac{1}{2} (y_t - \mu_j) \sigma_j^{-1} (y_t - \mu_j) \right\}, \quad (4.25)$$

onde  $\iota_j$  é a  $j$ -ésima coluna da matriz identidade de dimensão ( $N \times N$ ), pois o processo encontra-se no regime  $j$ ,  $\mu_j = E[y_t | \xi_t = \iota_j]$  representa a esperança condicional de  $y_t$  dado que o processo está no regime  $j$ , e  $\theta$  contém todos os parâmetros da população num único vetor, incluindo-se os parâmetros do modelo auto-regressivo e as probabilidades de transição que regem a Cadeia de Markov.

De forma semelhante à (4.11), representaremos todas as densidades condicionais num vetor de dimensão ( $N \times 1$ ), para o caso geral, onde temos  $N$  regimes. Este vetor está representado a seguir:

$$\eta_t = \begin{bmatrix} f(y_t | \xi_t = \iota_1, y_{t-1}) \\ \vdots \\ f(y_t | \xi_t = \iota_N, y_{t-1}) \end{bmatrix}. \quad (4.26)$$

Assim, para encontrarmos a densidade marginal de  $(y_t, s_t)$ , utilizaremos a distribuição conjunta de  $y_t$  e  $\xi_t$ , integrando-se em relação a  $\xi_t$ . Ilustraremos a seguir como derivar a distribuição marginal de  $y_t$ , através de um caso simples, o de mistura de distribuições *i.i.d.*

Primeiramente, veja que a distribuição conjunta de  $y_t$  e  $s_t$  é dada por

$$f(y_t, s_t = j; \theta) = f(y_t | s_t = j; \theta) \cdot P(s_t = j; \theta). \quad (4.27)$$

Visto que nesta ilustração estamos tratando de um modelo de mistura de distribuições, o vetor de parâmetros  $\theta$  será diferente daquele de um modelo auto-regressivo, pois este não conterá os parâmetros da auto-regressão.

Uma vez tendo encontrado a distribuição conjunta de  $y_t$ , obtém-se a função densidade marginal de  $y_t$  através da seguinte soma:

$$f(y_t; \theta) = \sum_{j=1}^N f(y_t, s_t = j; \theta) = \sum_{j=1}^N f(y_t | s_t = j; \theta) \cdot P(s_t = j; \theta) \quad (4.28)$$

Como podemos perceber, precisamos de  $P\{s_t = j; \theta\}$  para podermos derivar a densidade marginal de  $y_t$  e a função de verossimilhança.

#### 4.5. PROBABILIDADES PREDITAS, FILTRADAS E SUAIVIZADAS

Tendo uma estimativa inicial para o vetor de parâmetros  $\theta$ , poderemos fazer alguma inferência a respeito de qual regime tenha sido o mais provável responsável por determinada observação num instante qualquer  $t$ . Utilizaremos as definições de probabilidade condicional para fazer as inferências cabíveis a respeito de tais regimes, conforme abaixo:

$$P\{s_t = j | y_t; \theta\} = \frac{f(y_t, s_t = j; \theta)}{f(y_t; \theta)} = \frac{f(y_t | s_t = j; \theta) \cdot P(s_t = j; \theta)}{\sum_{j=1}^N f(y_t | s_t = j; \theta) \cdot P(s_t = j; \theta)} \quad (4.29)$$

**Obs.:** o cálculo da probabilidade apresentada acima apenas poderá ser utilizado no caso *i.i.d.*, pois neste caso, as inferências a respeito de  $s_t$  dependem apenas de  $y_t$ .

Considerando um caso mais geral poderemos estender o raciocínio acima, uma vez que a inferência a respeito do regime  $s_t$  depende de todas as observações disponíveis, e não apenas de uma única observação num determinado instante. Portanto, para generalizar a equação (4.29) utilizaremos um filtro e suavizador conhecido por BLHK, que recebeu este nome por ter sido sugerido conjuntamente por Baum, Lindgren, Hamilton e Kim. Através deste filtro podemos fazer inferências a respeito dos regimes da série por meio da estimação de probabilidades previstas, filtradas e suavizadas, que serão discutidas mais detalhadamente a seguir.

Um dos objetivos, seguindo a metodologia de estimação de modelos MS-AR, será estimar os parâmetros que compõem  $\theta$  baseados nas observações da série. Porém, uma estimativa inicial para  $\theta$  é necessária para estimar as probabilidades do processo  $s_t$  estar em cada regime ao longo do tempo. Mais tarde discutiremos alternativas para esta estimativa inicial de  $\theta$ .

Inicialmente, esclareceremos a notação a ser usada para probabilidades preditas, filtradas e suavizadas estimadas, dadas conforme o quadro a seguir:

Notação	Condição	Significado
$\hat{\zeta}_{t \tau}^{\hat{\tau}}$	$\tau \leq t$	Probabilidades Preditas Estimadas
$\hat{\zeta}_{t \tau}^{\tau}$	$\tau = t$	Probabilidades Filtradas Estimadas
$\hat{\zeta}_{t \tau}^{\hat{\tau}}$	$t \leq \tau \leq T$	Probabilidades Suavizadas Estimadas

Quadro 4.1: Notação para as estimativas das probabilidades preditas, filtradas e suavizadas.

Pelo quadro acima, vemos que as probabilidades preditas são usadas para fazermos previsões a respeito dos regimes para algum período futuro. Já as probabilidades filtradas são usadas para fazermos estimativas a respeito dos regimes para o instante atual. Enfim, as probabilidades suavizadas são usadas para fazermos estimativas do passado a respeito dos regimes, ou seja, fazer estimativas para uma determinada data  $t$  condicional a todas as observações da amostra.

Como podemos perceber, as probabilidades filtradas podem ser consideradas como um caso particular das probabilidades preditas, uma vez que a condição desta última engloba a condição de probabilidade filtrada.

Faremos agora uma generalização da equação (4.29), como havia sido dito anteriormente. Para tanto utilizaremos notação matricial, onde buscaremos a inferência ótima a respeito dos regimes e as previsões para os mesmos a cada instante  $t$  da amostra, que serão encontradas iteragindo o par de equações apresentado abaixo:

$$\hat{\zeta}_{t|\tau}^{\hat{\tau}} = \frac{\left( \hat{\zeta}_{t|\tau-1}^{\hat{\tau}} \otimes \eta_t \right)}{1' \left( \hat{\zeta}_{t|\tau-1}^{\hat{\tau}} \otimes \eta_t \right)}, \quad (4.30)$$

$$\hat{\zeta}_{t+1|\tau}^{\hat{\tau}} = P \cdot \hat{\zeta}_{t|\tau}^{\hat{\tau}}, \quad (4.31)$$

onde  $\eta_t$  representa o vetor das densidades condicionais tal como em (4.26),  $P$  é a matriz de transição como definida em (2.5),  $\otimes$  denota a multiplicação elemento por elemento dos vetores e  $1'$  é um vetor unitário transposto de dimensão  $N$ .

Como pode ser verificado no Quadro 4.1, a equação (4.30) determina o cálculo das probabilidades filtradas estimadas, enquanto a equação (4.31) apresenta o cálculo das probabilidades preditas estimadas. Além disso, ao iteragir estas duas equações, verificamos que teremos vetores de probabilidades estimadas (filtradas e preditas) para

cada instante  $t$  da amostra, onde  $t = 1, 2, \dots, T$ . Repare que, em (4.29), para efetuarmos a estimativa das probabilidades filtradas, necessitamos de um vetor com as probabilidades preditas estimadas, e em (4.30) acontece o contrário, precisamos de um vetor com as probabilidades filtradas estimadas para encontrarmos uma estimativa para as probabilidades preditas. E assim segue o processo iterativo até encontrarmos todas as estimativas para os vetores de probabilidades.

Este processo se inicia a partir do momento em que tivermos valores iniciais para  $\hat{\xi}_{1|0}$  e para o vetor de parâmetros da população  $\theta$ . Assim, iteragindo as equações (4.30) e (4.31) para  $t = 1, \dots, T$ , obtemos  $\hat{\xi}_{t|t}$  e  $\hat{\xi}_{t+1|t}$  para todo instante  $t$  que aparece na amostra.

Através das informações obtidas até o presente, podemos encontrar a função de log-verossimilhança, considerando os dados da amostra e os valores dos parâmetros do vetor  $\theta$  para a realização das iterações. Essa função pode ser calculada como

$$L(\theta) = \sum_{t=1}^T \log(f(y_t | y_{t-1}; \theta)), \quad (4.32)$$

onde

$$f(y_t | y_{t-1}; \theta) = 1'(\hat{\xi}_{t|t-1} \otimes \eta_t). \quad (4.33)$$

Maiores detalhes a respeito das expressões (4.30) a (4.33) podem ser encontradas em Hamilton (1994, Seção 22.4).

Em (4.31), vimos a possibilidade de fazermos previsões a respeito dos regimes *um passo a frente*, ou seja, previsão para um instante  $t + 1$  dado que já coletamos as informações até o instante  $t$ . Seguindo o mesmo raciocínio, poderemos fazer previsões para os regimes *m passos a frente*, ou seja, previsão para um instante  $t + m$  dado que já coletamos as informações até o instante  $t$ . Uma previsão ótima para  $\xi_{t+m}$  poderá ser feita encontrando-se a esperança apresentada em (2.13) condicional ao instante  $t$ , conforme proposto por Hamilton (1994, Seção 22.4):

$$E(\xi_{t+m} | y_t) = P^m \cdot E(\xi_t | y_t), \quad (4.34)$$

que pode ser escrita de maneira similar como:

$$\hat{\xi}_{t+m|t} = P^m \cdot \hat{\xi}_{t|t}. \quad (4.35)$$

Há ainda outra forma de fazermos inferências a respeito dos regimes: através das probabilidades suavizadas. Estas probabilidades podem ser encontradas estendendo-se a idéia contida nas expressões de (4.30) e (4.31). Uma inferência para um regime da série



no instante  $t$  baseada em dados coletados até um instante posterior a  $t$  é obtida através de uma inferência suavizada. Portanto, uma inferência desse tipo pode ser encontrada utilizando-se toda a informação contida na amostra.

Obteremos as inferências suavizadas fazendo uso do algoritmo de Kim (1993), que faz parte do filtro suavizador *BLHK*, anteriormente citado. Em sua forma vetorial, este algoritmo pode ser escrito como

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \otimes \left\{ P \cdot \left[ \hat{\xi}_{t+1|T} (\div) \hat{\xi}_{t+1|t} \right] \right\}, \quad (4.36)$$

onde  $\otimes$  representa a multiplicação e  $(\div)$  representa a divisão elemento por elemento dos vetores.

As probabilidades suavizadas  $\hat{\xi}_{t|T}$  de (4.36) são obtidas iteragindo (4.36) com (4.30) e (4.31) para trás fazendo-se  $t = T-1, T-2, \dots, 1$ . Porém, salientamos que este processo iterativo de estimação será iniciado fazendo-se  $t=T$ , obtendo-se assim  $\hat{\xi}_{T|T}$  a partir da expressão (4.30).

Como podemos observar, as probabilidades suavizadas são obtidas de forma inversa às probabilidades preditas e filtradas, pois suas estimativas são iniciadas quando  $t=T$  (a última observação da amostra) e finalizadas quando  $t=1$  (a primeira observação da amostra). É exatamente por este motivo que devemos primeiro estimar as probabilidades preditas e filtradas para então podermos estimar as probabilidades suavizadas.

O algoritmo de Kim só pode ser aplicado se algumas exigências forem cumpridas, tais como:

- $s_t$  segue uma Cadeia de Markov de primeira ordem;
- a densidade condicional de  $y_t$  depende apenas de  $s_t$ ;

No Apêndice 22.A de Hamilton (1994) podem ser obtidas maiores informações a respeito do algoritmo de Kim.

#### 4.5.1. INICIALIZANDO AS ESTIMATIVAS

Dados os valores iniciais para o vetor de probabilidades preditas  $\hat{\xi}_{1|0}$ , poderemos facilmente calcular  $\hat{\xi}_{t|t}$  para qualquer instante  $t$  através da iteração do par de equações (4.30) e (4.31).

Para a escolha dos valores iniciais das probabilidades preditas citadas acima, sugerimos duas opções:

1. atribuir pesos iguais para as probabilidades do vetor, ou seja,  $N^{-1}$  pois temos  $N$  regimes;
2. utilizar o conhecimento a priori do processo.

Devemos no entanto tomar cuidado para que nenhuma posição do vetor das estimativas iniciais das probabilidades preditas assuma o valor zero, o que acarretará problemas de convergência.

#### 4.5.2. ALGORITMO TIPO *EM*

Quando iterarmos o par de equações (4.30) e (4.31), estamos assumindo que o vetor dos parâmetros da população denotado por  $\theta$  é fixo e conhecido. Com relação ao vetor  $\theta$ , uma vez tendo sido completada a inferência a respeito dos regimes do processo na amostra, o valor da log-verossimilhança pode ser obtido da expressão (4.32). No entanto, uma estimativa para  $\theta$  capaz de maximizar a função log-verossimilhança só é possível através de métodos numéricos, uma vez que as derivadas da expressão (4.32) não possuem forma explícita.

O algoritmo (iterativo) EM, assim conhecido por se tratar de uma abreviatura de Maximização da Esperança (do inglês *Expectation-Maximization – EM Algorithm*) permite obter as estimativas de máxima verossimilhança dos parâmetros do modelo.

O algoritmo recebeu este nome pelo fato de englobar duas fases no processo de estimação da máxima verossimilhança:

- *esperança*: onde os regimes não observados são estimados através das probabilidades suavizadas  $P\{s_t|y_T; \theta^{j-1}\}$ , onde  $\theta^{j-1}$  representa o vetor de parâmetros estimados na etapa de estimação anterior a  $j$ ;
- *maximização*: quando obtemos uma estimativa para o vetor dos parâmetros da população que satisfaça às condições de primeira ordem relacionadas com a função de verossimilhança.

Num primeiro momento serão calculadas todas as probabilidades condicionais através das recursões filtradas e suavizadas, utilizando-se as estimativas do vetor dos

parâmetros  $\theta^{j-1}$  encontradas na fase de maximização imediatamente anterior, em substituição ao verdadeiro vetor (desconhecido) de parâmetros  $\theta$ .

Já na fase de maximização, as probabilidades condicionais dos regimes desconhecidos serão substituídas pelas probabilidades suavizadas encontradas na fase anterior.

Procedendo desta maneira, cada vez que realizarmos essa iteração do algoritmo EM, teremos uma nova estimativa para o vetor de parâmetros populacionais. O processo de estimação segue assim até que haja uma convergência, segundo algum critério definido pelo pesquisador.

#### 4.5.3. INICIALIZANDO O ALGORITMO EM

Tendo uma estimativa inicial arbitrária para o vetor de parâmetros da população  $\theta$ , que chamaremos de  $\theta^0$ , podemos calcular o valor de  $P\{s_t=j|y_t; \theta^0\}$  através do filtro e suavizador *BLHK*. Faz-se uso das condições de maximização da função log-verossimilhança para utilizar  $\theta^0$  no lugar de  $\hat{\theta}$ . Com isto, estaremos gerando uma nova estimativa para o vetor de parâmetros  $\theta$ , que chamaremos de  $\theta^1$ . Como encontramos novos valores para o vetor de parâmetros  $\theta$ , podemos calcular  $P\{s_t=j|y_t; \theta^1\}$ . Este cálculo de probabilidades é efetuado a cada nova estimativa que obtivermos para  $\theta$ . O procedimento é agora repetido para calcular uma nova estimativa para  $\theta$ , fazendo-se as substituições cabíveis, que agora será chamado de  $\theta^2$ . Procedemos desta maneira até obtermos a convergência.

Logo, a cada iteração feita através do referido algoritmo, necessariamente precisaremos passar por um processo de filtragem e suavizamento, seguindo-se à resolução das condições de primeira ordem para podermos estimar  $\theta$ . Assim, estaremos maximizando a função de verossimilhança.

#### 4.6. PREVISÕES PARA A SÉRIE $y_t$

Da densidade condicional apresentada na expressão (4.10), é possível obtermos diretamente uma previsão para  $y_{t+1}$ , uma vez que conhecemos  $y_t$  e  $s_{t+1}$ . Ou seja, uma previsão para  $y_{t+1}$  está condicionada à  $y_t$  e  $s_{t+1}$ .

Considere o seguinte modelo AR(1):

$$y_{t+1} = c_{s_{t+1}} + \phi_{s_{t+1}} y_t + \varepsilon_{t+1}. \quad (4.37)$$

Uma previsão para o modelo apresentado em (4.37) será dada por

$$E(y_{t+1} | s_{t+1} = j, y_t; \theta) = c_j + \phi_j y_t. \quad (4.38)$$

Para cada valor que a variável não observável  $s_{t+1}$  pode assumir ao longo da série, teremos uma diferente previsão associada. Portanto, se a série apresentar  $N$  regimes, teremos  $N$  possíveis previsões.

Em (4.38) chegamos a  $N$  possíveis previsões para a série, dependendo do regime em que a série se encontra. Essas previsões podem ser coletadas num vetor de dimensão  $N$ , que será denotado por  $h_t'$ . Por meio de  $h_t'$  verificamos que a previsão incondicional da série baseada na observação atual está relacionada com a previsão condicional por

$$E(y_{t+1} | y_t; \theta) = h_t' \hat{\xi}_{t+1|t}, \quad (4.39)$$

onde  $\hat{\xi}_{t+1|t}$  representa o vetor contendo as probabilidades previstas para previsões um passo a frente.

#### 4.7. ESTIMATIVA DE MÁXIMA VEROSSIMILHANÇA DOS PARÂMETROS

Começaremos pelas estimativas das probabilidades de transição  $p_{ij}$ . Assumindo que estas probabilidades possam assumir valores entre zero e um e  $\sum_{j=1}^N p_{ij} = 1, \forall i \text{ e } j$ , e que as probabilidades iniciais do vetor  $\hat{\xi}_{1|0}$  assumem algum valor fixo não correlacionado com os demais parâmetros, teremos que a estimativa de máxima verossimilhança para as probabilidades de transição do processo podem ser encontradas por

$$\hat{p}_{ij} = \frac{\sum_{t=2}^T P\{s_t = j, s_{t-1} = i | y_T; \hat{\theta}\}}{\sum_{t=2}^T P\{s_{t-1} = i | y_T; \hat{\theta}\}}. \quad (4.40)$$

Perceba que na equação acima o intervalo de variação para  $t$  é delimitado por 2 e  $T$ , pois se  $t$  pudesse assumir o valor 1, entraria no cálculo um valor para o regime  $s_0$ , o

que impossibilitaria a estimativa das probabilidades de transição. Perceba também que os cálculos são realizados com base nas probabilidades suavizadas.

Se o vetor das probabilidades iniciais puder ser considerado como um vetor de parâmetros limitado apenas pelas exigências de que a soma das probabilidades seja um e que cada valor seja superior a zero, então a estimativa de máxima verossimilhança para este vetor pode ser obtida através de uma inferência suavizada para o regime inicial do tipo:

$$\hat{\rho} = \hat{\xi}_{1|T}. \quad (4.41)$$

Agora especificaremos as estimativas de máxima verossimilhança para os parâmetros auto-regressivos do modelo. Considere o modelo:

$$y_t = z_t' \beta_{s_t} + \varepsilon_t, \quad (4.42)$$

onde  $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$ ,  $z_t$  é um vetor contendo as defasagens da variável  $y_t$ , ou seja, se o modelo auto-regressivo for de primeira ordem, o vetor  $z_t$  será representado por  $[1, y_{t-1}]$ , se for de segunda ordem por  $[1, y_{t-1}, y_{t-2}]$ , e assim sucessivamente, e  $\beta_{s_t}$  é o vetor com os coeficientes da regressão para cada regime.

No modelo proposto, o vetor contendo as densidades condicionais  $\eta_t$  será dado por:

$$\eta_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(y_t - z_t' \beta_1)^2}{2\sigma^2}\right\} \\ \vdots \\ \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(y_t - z_t' \beta_N)^2}{2\sigma^2}\right\} \end{bmatrix}. \quad (4.43)$$

Pelo modelo proposto em (4.42), teremos de encontrar as estimativas de máxima verossimilhança para os parâmetros da auto-regressão  $\beta_1, \dots, \beta_N$  e  $\sigma^2$ . Estes parâmetros serão agora denominados por  $\alpha$ . Para  $\alpha = (\beta_1', \beta_2', \dots, \beta_N', \sigma^2)'$  teremos que

$$\sum_{t=1}^T (y_t - z_t' \hat{\beta}_j) z_t \cdot P\{s_t = j | y_T; \hat{\theta}\} = 0. \quad (4.44)$$

Pela expressão (4.44) verificamos que os vetores  $\hat{\beta}_j$  satisfazem as condições OLS de ortogonalidade, onde cada observação é ponderada pela probabilidade de que ela tenha sido observada no regime  $j$ .

As estimativas de máxima verossimilhança serão obtidas através das seguintes expressões:

$$\hat{\beta}_j = \left[ \sum_{t=1}^T [\bar{z}_t(j)] [\bar{z}_t(j)]' \right]^{-1} \left[ \sum_{t=1}^T [\bar{z}_t(j)] \bar{y}_t(j) \right], \quad (4.45)$$

onde  $\bar{y}_t(j)$  e  $\bar{z}_t(j)$  serão dados respectivamente por

$$\bar{y}_t(j) = y_t \cdot \sqrt{P\{s_t = j | y_T; \hat{\theta}\}} \text{ e } \bar{z}_t(j) = z_t \cdot \sqrt{P\{s_t = j | y_T; \hat{\theta}\}}. \quad (4.46)$$

E a variância  $\sigma^2$ , por sua vez, será estimada por

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \sum_{j=1}^N (y_t - z_t \hat{\beta}_j)^2 \cdot P\{s_t = j | y_T; \hat{\theta}\}}{T}. \quad (4.47)$$

Note que as probabilidades utilizadas no par de equações (4.46) são suavizadas. Perceba também que as estimativas para  $\hat{\beta}_j$  podem ser encontradas fazendo-se uma regressão OLS de  $\bar{y}_t(j)$  sobre  $\bar{z}_t(j)$ , como aparece em (4.45).

Os parâmetros do modelo acima poderão agora ser estimados implementando-se o algoritmo EM, procedendo-se aos passos discutidos na Seção 4.5.3.

#### 4.8. NÚMERO ADEQUADO DE REGIMES

Uma das tarefas a serem feitas na metodologia MS-AR é determinar o número adequado de regimes que governarão o processo. Em geral, o que fazemos é a realização de testes comparando modelos com números distintos de regimes.

Porém estes testes não podem ser realizados através de testes de verossimilhança usuais. Uma das condições de regularidade para que esses testes de verossimilhança tenham uma distribuição  $\chi^2$  assintótica é que a matriz  $P$  seja não singular.

Há uma vasta bibliografia a respeito de testes envolvendo os problemas de regularidade, onde podemos citar Davies (1977), Hansen (1993), Andrews & Ploberger (1992) e Stinchcombe & White (1993). Segundo Hamilton (1994), podemos fazer uma aproximação utilizando na hipótese nula um modelo com  $N-1$  regimes e realizando uma série de testes sobre a validação das especificações como uma forma de verificar se realmente é necessário ajustarmos um modelo com  $N$  regimes. Em Hamilton (1996) encontraremos alguns destes testes.

## 5. SIMULAÇÕES UTILIZANDO O *R*

Ao longo deste capítulo, implementaremos algumas simulações de Monte Carlo, enfocando a obtenção das estimativas de máxima verossimilhança para os parâmetros de diferentes modelos, com e sem mudança de regime. Para os diferentes modelos simulados, ajustamos modelos AR, com e sem mudança de regime, comparando os casos simulados. Todos os procedimentos simulação e estimação foram obtidos utilizando-se a versão 1.7.1 software estatístico *R* (ver anexo).

Este software é um programa que está sendo utilizado cada vez mais por pesquisadores que fazem uso de análises estatísticas. O crescente aumento da utilização deste software deve-se, entre outros, à confiabilidade dos resultados por ele fornecidos, à flexibilidade e diversidade de análises estatísticas, à boa estrutura de programação e também ao fato de ser gratuito. Também podem ser obtidas diversas saídas gráficas a partir do *R*.

Aqueles usuários que utilizam as linguagens de programação C, C++ e FORTRAN, não terão dificuldades em utilizar o *R*.

### 5.1. SIMULAÇÃO DE MODELOS

Passaremos agora a simular modelos com o objetivo de mostrar que, em séries temporais que apresentam comportamentos distintos ao longo do tempo, faz sentido e é vantajoso ajustar modelos com mudança de regime.

Nosso processo de simulação contemplará os seguintes passos:

- gerar séries temporais de modelos do tipo RB e AR com e sem mudança de regime via cadeias de Markov;
- ajustar modelos AR e MS-AR e estimar os parâmetros dos mesmos, com base nas séries temporais geradas;
- comparar os modelos ajustados com mudança de regime e sem mudança de regime através dos critérios de informação AIC e BIC.

Discutiremos a seguir as análises realizadas para três casos particulares.

### 5.1.1. CASO 1

Este caso introdutório será tratado de maneira diferente dos demais, pois simularemos uma série que não apresenta mudança de regime. Foi simulada uma única série temporal de tamanho 200 de um processo AR(1).

Os parâmetros do modelo simulado foram:

---

$c = 2;$	$\phi = 0,5; e$	$\sigma_e^2 = 0,5.$
----------	-----------------	---------------------

---

Quadro 5.1: Parâmetros do processo gerador.

A série temporal gerada apresenta o seguinte gráfico:

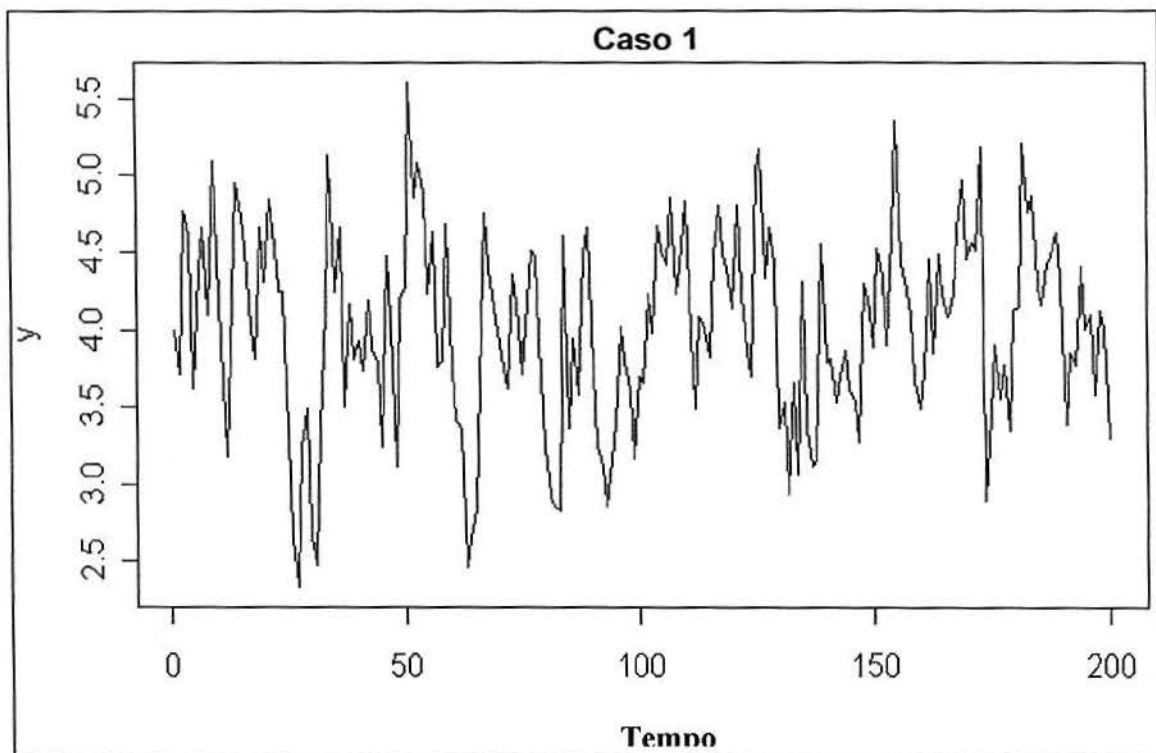


Figura 5.1: Gráfico da série temporal gerada no Caso 1.

Pela função de autocorrelação parcial apresentada na figura abaixo, é sugerido o ajuste de um modelo AR(1), pois apenas quando lag = 1, a função é significativamente diferente de zero.



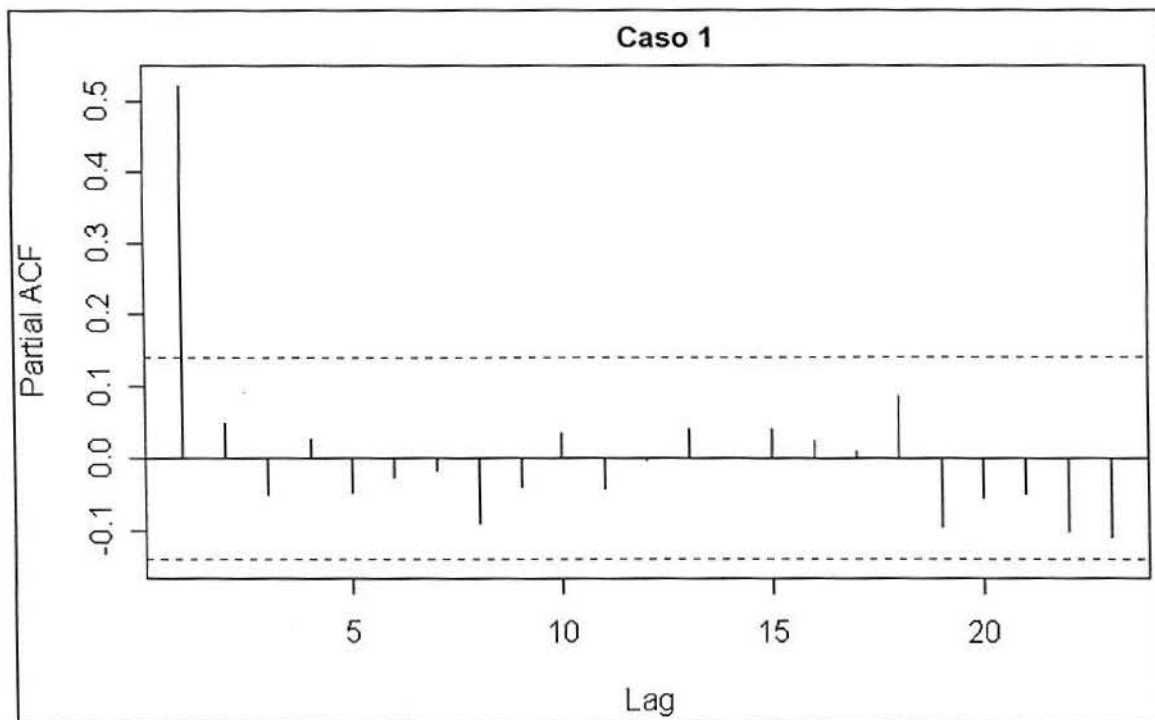


Figura 5.2: PACF para o Caso 1.

Nossa intenção ao gerar a série acima foi gerar uma série que apresentasse um comportamento autorregressivo seguindo um modelo AR(1). Porém, analisaremos esta série supondo que ela apresente um comportamento tipo MS(2)-AR(1).

Para inicializar o processo de estimação dos parâmetros do modelo, na rotina criada no R, entramos com as seguintes informações iniciais:

- número de regimes,  $N = 2$ ;
- tamanho da amostra,  $T = 200$ ;
- os termos constantes para cada regime, onde, para o regime 1 entramos com um valor abaixo da média amostral, e para o regime 2, com um valor acima da média amostral;
- os coeficientes regressores,  $\phi_1 = 0,6$  e  $\phi_2 = 0,4$ ;
- a variância do erro, onde utilizou-se a variância amostral da série;
- as probabilidades de transição em uma etapa,  $p_{ij} = 0,55$  e  $p_{ii} = 0,45$ ,  $i$  e  $j = 1, 2$ ;
- O critério de convergência adotado foi que o algoritmo só parasse o processo de estimação dos parâmetros quando a diferença entre a estimativa atual e a estimativa imediatamente anterior fosse menor que 0,00001;

Foram obtidas as seguintes estimativas para o vetor dos parâmetros  $\theta$ :

$\hat{c}_1 = 3,07$	$\hat{c}_2 = 1,29$	$\hat{\phi}_1 = 0,34$	$\hat{\phi}_2 = 0,61$	$\hat{\sigma}^2 = 0,17$
$\hat{p}_{11} = 0,38$	$\hat{p}_{12} = 0,62$	$\hat{p}_{21} = 0,39$	$\hat{p}_{22} = 0,61$	

Quadro 5.2: Estimativas para o vetor de parâmetros  $\theta$ .

Portanto, o modelo final ajustado - MS(2)-AR(1), forneceu as seguintes equações autorregressivas para cada regime da série:

Regime 1	$y_t = 3,07 + 0,34y_{t-1} + \varepsilon_t$
Regime 2	$y_t = 1,29 + 0,61y_{t-1} + \varepsilon_t$

Quadro 5.3: Equações autorregressivas para cada regime

Obtivemos as estimativas das probabilidades suavizadas, representadas graficamente abaixo:

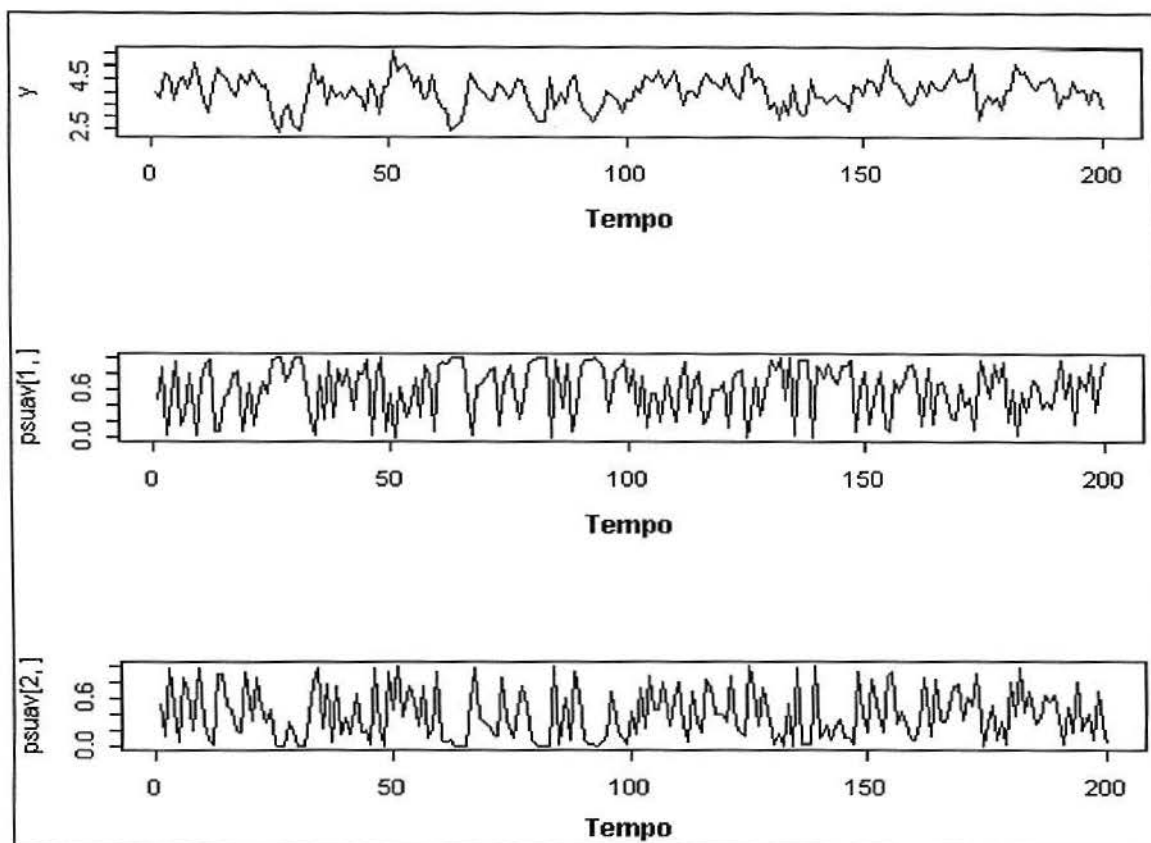


Figura 5.3: Gráfico da série com as probabilidades estimadas suavizadas para os respectivos regimes 1 e 2

Analisando a Figura 5.3, verificamos que o processo subjacente apresenta altas probabilidades suavizadas estimadas de estar no regime 1 quando a série apresenta valores baixos. Nestes casos, as probabilidades suavizadas estimadas de que o processo esteja no regime 2 são baixas.

Com o rigoroso critério de convergência adotado, as estimativas finais para  $\theta$  foram obtidas após 200 ciclos de estimação. Com isto concluímos que o processo de estimação dos parâmetros não convergiu rapidamente, talvez porque as informações iniciais não fossem muito precisas.

Faremos agora uma comparação entre os modelos ajustados com mudança de regime (MR) e sem MR, com a finalidade de verificar qual dos dois modelos se ajusta melhor. Para que fosse possível essa comparação, foram geradas aleatoriamente 50 séries temporais com 200 pontos cada, nas mesmas condições em que foi gerada a série original do Caso 1. Foram calculadas as estatísticas AIC e BIC para os modelos ajustados com e sem mudança de regime para cada uma das 50 séries. As estatísticas de AIC encontradas para as 50 séries geradas estão representadas na figura abaixo:

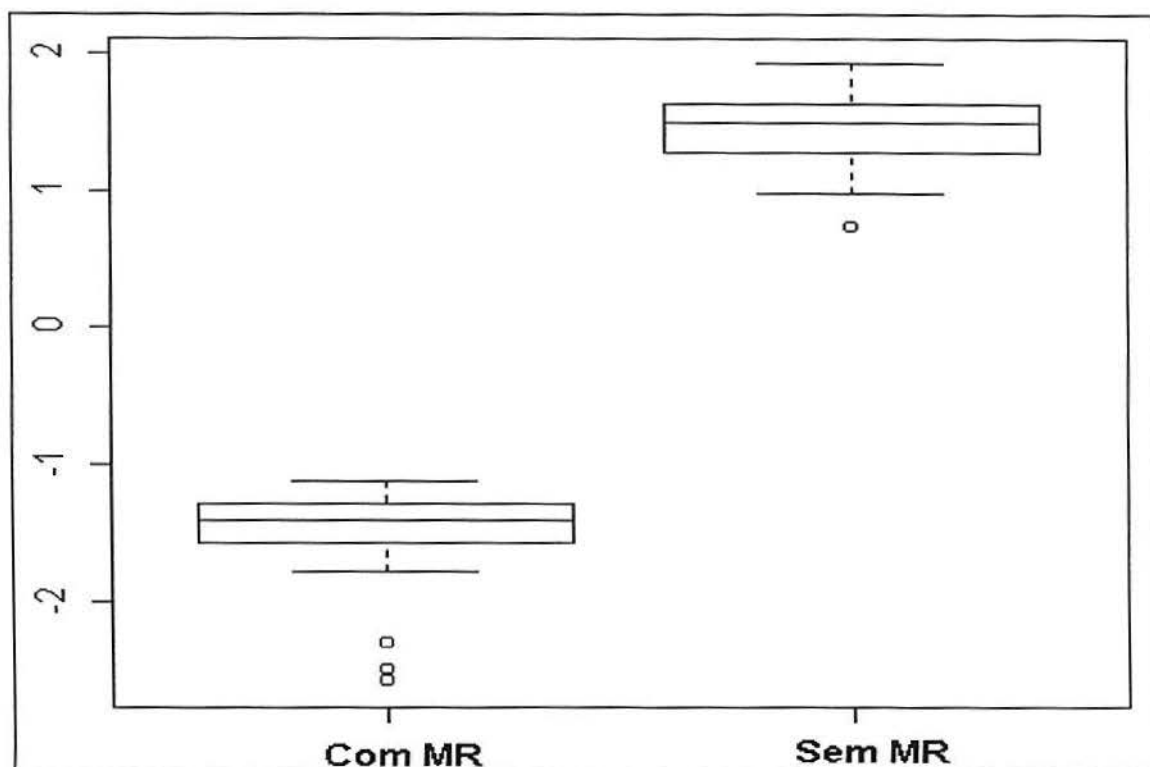


Figura 5.4: Boxplot comparando os AIC's obtidos para os modelos com e sem MR das 50 séries geradas

Verificando o boxplot apresentado na Figura 5.4, percebemos que as séries em que foram ajustados modelos com mudança de regime produziram ajustes superiores em

relação aqueles ajustes de modelos sem mudança de regime, uma vez que produziu medidas de AIC menores.

A seguir estão ilustradas as medidas de BIC para as 50 séries geradas.

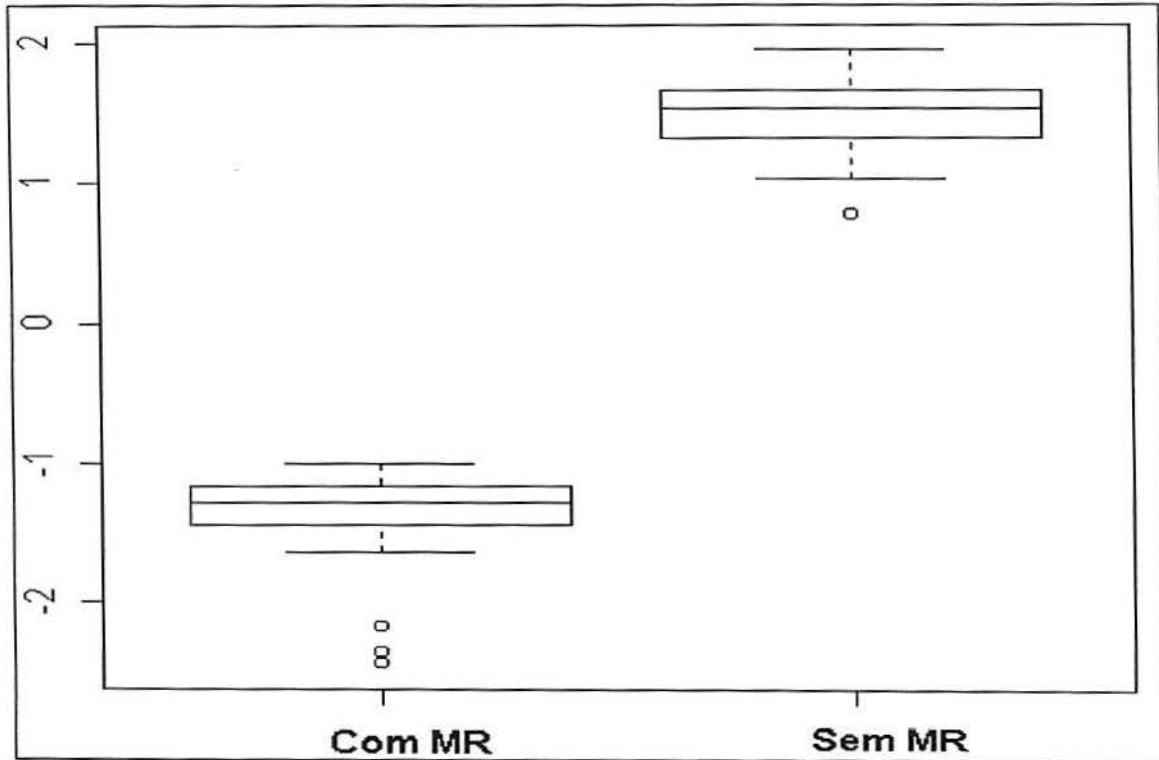


Figura 5.5: Boxplot comparando os BIC's obtidos para os modelos com e sem MR das 50 séries geradas

Percebemos pela Figura 5.5 que o critério de BIC apresentou resultados idênticos ao critério AIC. Podemos dizer então que ambos os critérios indicam o modelo MS(2)-AR(1) como sendo melhor explicativo das 50 séries geradas. Apesar de o modelo incorreto estar sendo escolhido, isso sugere uma certa robustez do modelo MS-AR.

### 5.1.2. CASO 2

Com base nas probabilidades de transição de uma Cadeia de Markov com dois regimes, foi simulada uma série de tamanho 200. Os parâmetros autorregressivos de cada regime foram:

$$c_1 = 2,$$

$$c_2 = 5,$$

$$\phi_1 = \phi_2 = 0,$$

$$\sigma_\varepsilon^2 = 1$$

Quadro 5.4: Valores iniciais para o processo de estimação

Assim, obteremos o seguinte gráfico para a série:

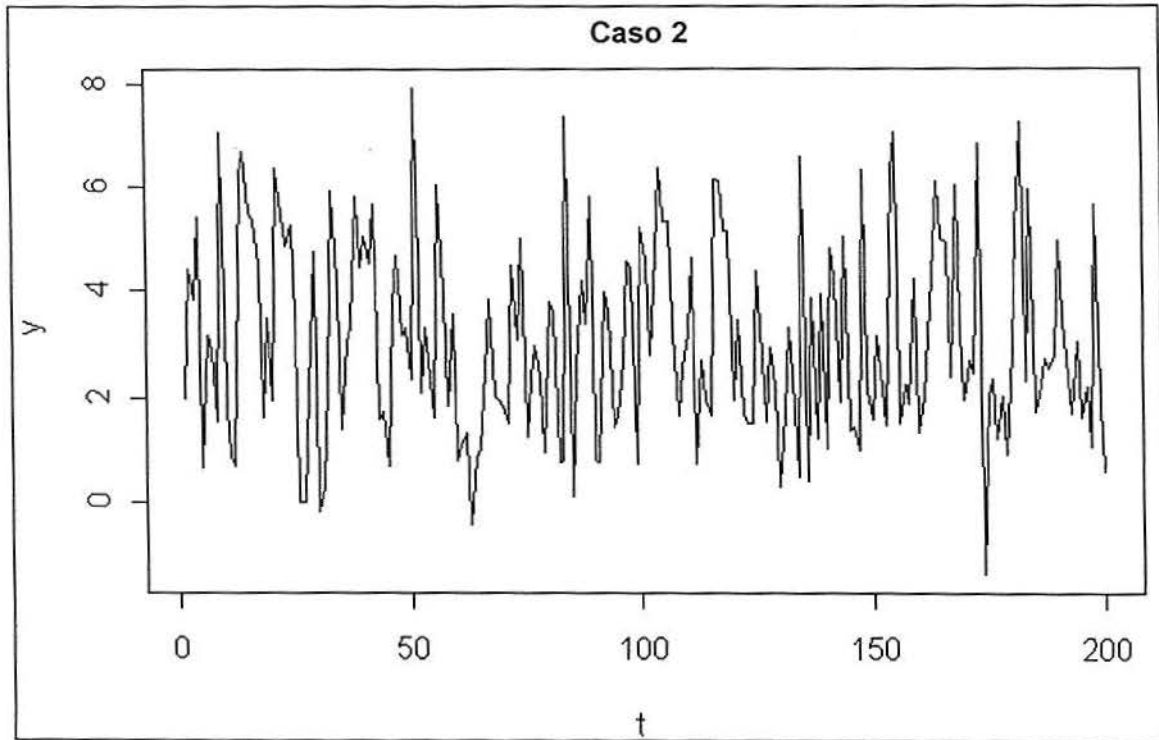


Figura 5.6: Gráfico da série para o Caso 2

Os gráficos de ACF e PACF para o Caso 2 são apresentados que segue.

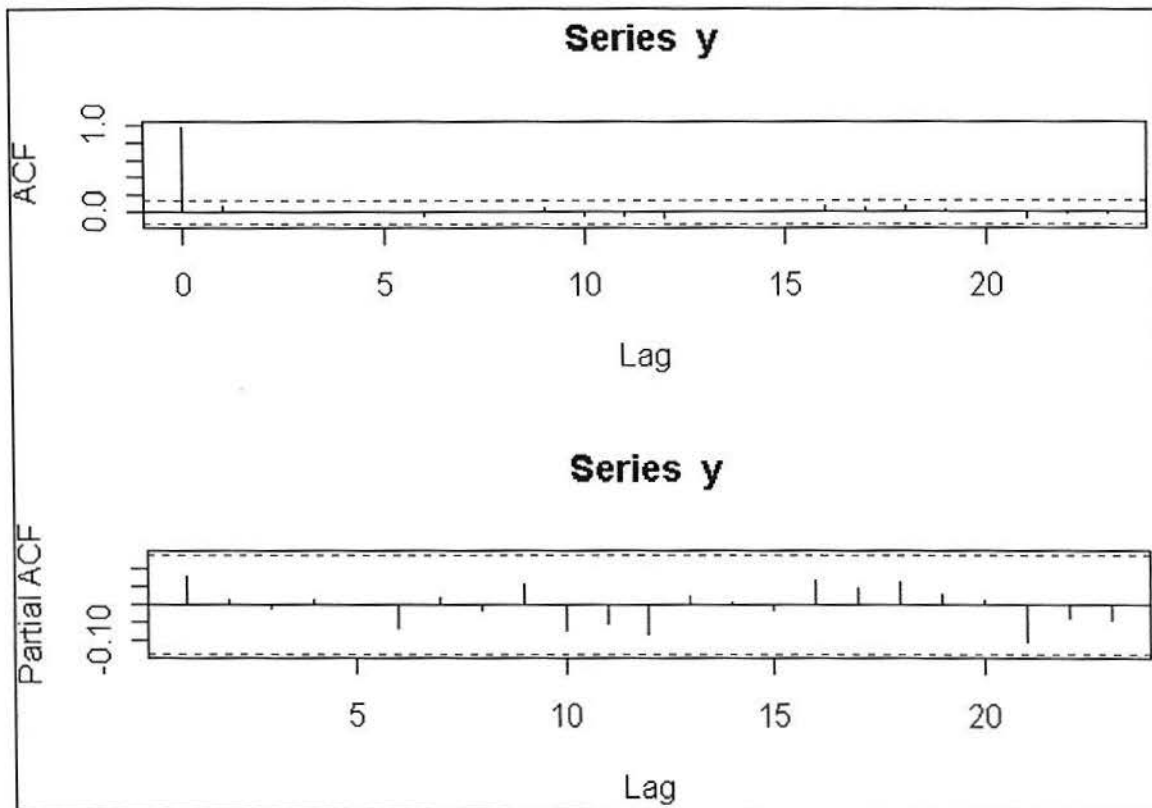


Figura 5.7: ACF e PACF para o Caso 2.

Na figura 5.7 verificamos que a série é um ruído, uma vez que as funções de ACF e PACF não são significativas para as defasagens (lags) apresentadas. A série é estacionária, pois há um rápido declínio da ACF, donde concluímos que o módulo dos regressores ( $\phi$ ) do modelo são menores que 1.

Suponhamos que, com base na Figura 5.6, pareça aceitável o ajuste de um modelo com dois regimes. Pela Figura 5.7 verificamos que não faz sentido adicionarmos os parâmetros de regressão  $\phi_1$  e  $\phi_2$  (ou considerá-los iguais a zero) ao modelo, pois constatamos que a série pode ser tratada como um ruído em torno da média.

Para inicializar o processo de estimação dos parâmetros do modelo, na rotina criada no *R*, entramos com as seguintes informações iniciais:

- número de regimes,  $N = 2$ ;
- tamanho da amostra,  $T = 200$ ;
- as constantes para cada regime, onde, por se tratar de um ruído, sabemos que a média para cada regime será igual ao termo constante do modelo do respectivo regime, como não sabemos qual a média de cada regime,

informamos que o valor inicial de  $c_1$  é igual a média do processo menos um desvio padrão, e  $c_2$  igual a média mais um desvio padrão;

- os coeficientes regressores,  $\phi_1 = \phi_2 = 0$ , por se tratar de um ruído;
- a variância do erro, onde utilizou-se a variância amostral da série;
- as probabilidades de transição em uma etapa,  $p_{ij} = 0,5$ , para  $i$  e  $j = 1, 2$ ; e
- O critério de convergência adotado foi que o algoritmo só parasse o processo de estimação dos parâmetros quando a diferença entre a estimativa atual e a estimativa imediatamente anterior fosse menor que 0,00001;

Foram obtidas as seguintes estimativas para o vetor dos parâmetros  $\theta$ :

$\hat{c}_1 = 2,62;$	$\hat{c}_2 = 5,65$	$\hat{\phi}_1 = -0,07$	$\hat{\phi}_2 = -0,11$	$\hat{\sigma}^2 = 1,2$
$\hat{p}_{11} = 0,75$	$\hat{p}_{12} = 0,25$	$\hat{p}_{21} = 0,52$	$\hat{p}_{22} = 0,48$	

Quadro 5.5: Estimativas para o vetor de parâmetros  $\theta$ .

Portanto, o modelo final ajustado foi um MS(2)-AR(1), que forneceu as seguintes equações autorregressivas para cada regime da série:

Regime 1	$y_t = 2,62 - 0,07y_{t-1} + \varepsilon_t$
Regime 2	$y_t = 5,65 - 0,11y_{t-1} + \varepsilon_t$

Quadro 5.6: Equações autorregressivas para cada regime.

Também foram obtidas as estimativas das probabilidades suavizadas, apresentadas na figura abaixo.

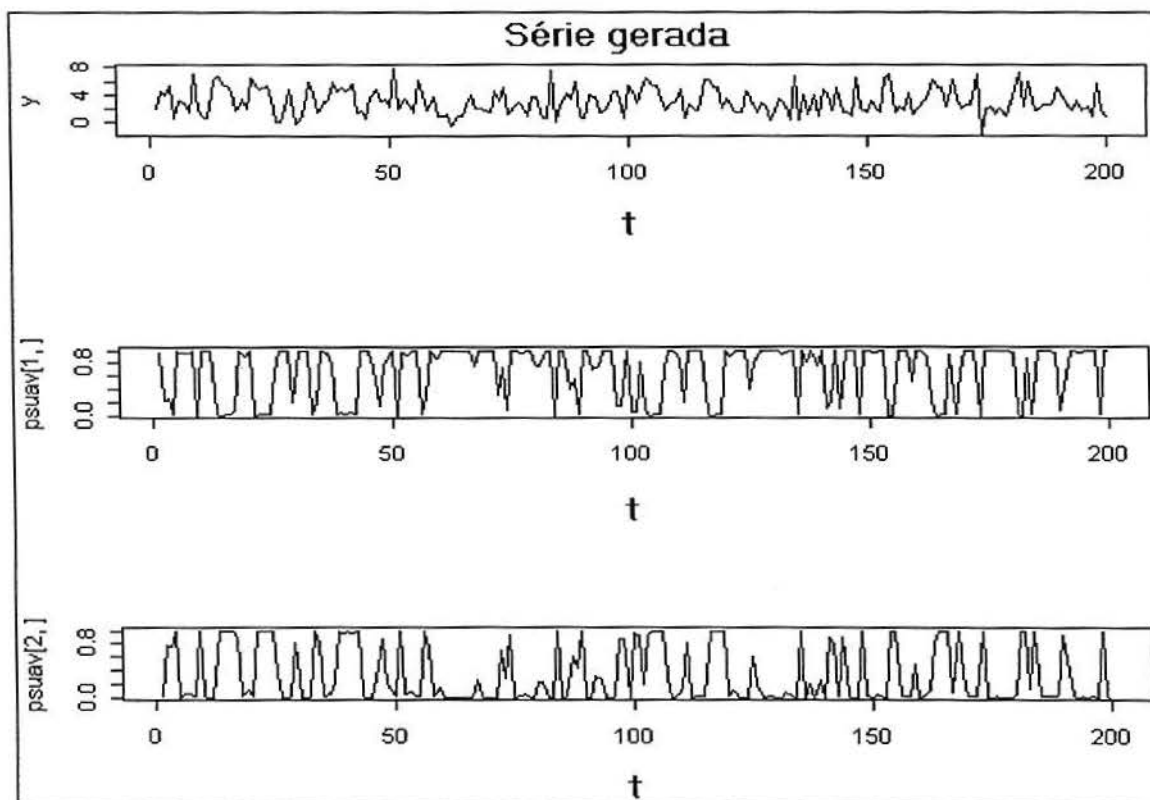


Figura 5.8: Gráfico da série com as probabilidades estimadas suavizadas para os respectivos regimes 1 e 2

Pela Figura 5.8, a exemplo do caso anterior, verificamos que o processo subjacente apresenta altas probabilidades suavizadas estimadas de estar no regime 1 quando a série apresenta baixos valores. Nestes casos, as probabilidades suavizadas estimadas de que o processo esteja no regime 2 são baixas.

Houve convergência do processo de estimação dos parâmetros após ter sido realizado 20 ciclos iterativos.

Para finalizar este caso, faremos uma comparação entre o ajuste de um modelo com mudança de regime e outro sem mudança. A finalidade desta comparação é verificar qual dos ajustes produz o menor erro de estimativa, ou seja, qual dos modelos produz o menor resíduo.

Para que fosse possível essa comparação, foram geradas aleatoriamente 50 séries temporais com 200 pontos cada, nas mesmas condições em que foi gerada a série para as análises anteriores contidas neste caso. Obteve-se as medidas de AIC e BIC para cada série gerada. Estas medidas são apresentadas nas Figuras 5.9 e 5.10.

Nas Figuras 5.9 e 5.10 fica evidenciado que, para o Caso 2, os modelos com mudança de regime se ajustam melhor do que aqueles que não consideram as mudanças de regime, uma vez que estes últimos apresentam erro de previsões maiores. Assim,



para o Caso 2, verificamos que os modelos MS(2)-AR(1) são mais adequados do que os modelos AR(1), uma vez que produziram medidas de AIC e BIC menores.

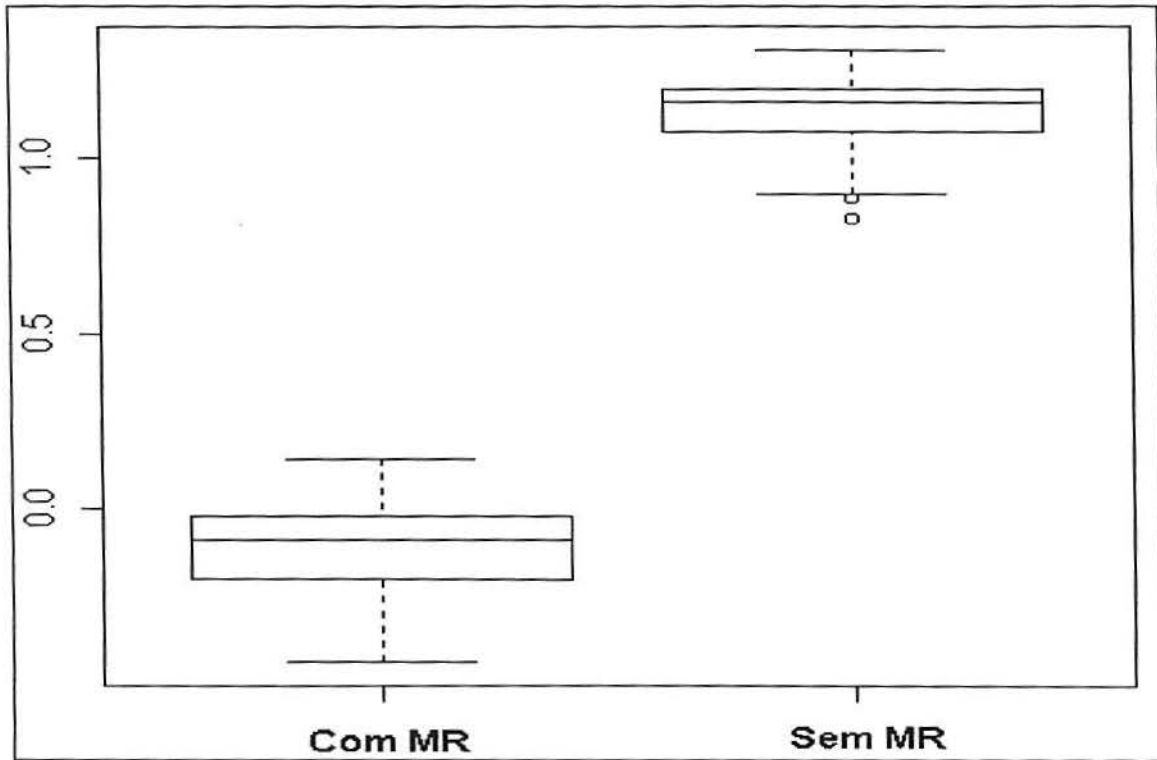


Figura 5.9: Boxplot comparando os AIC's obtidos para os modelos com e sem MR das 50 séries geradas

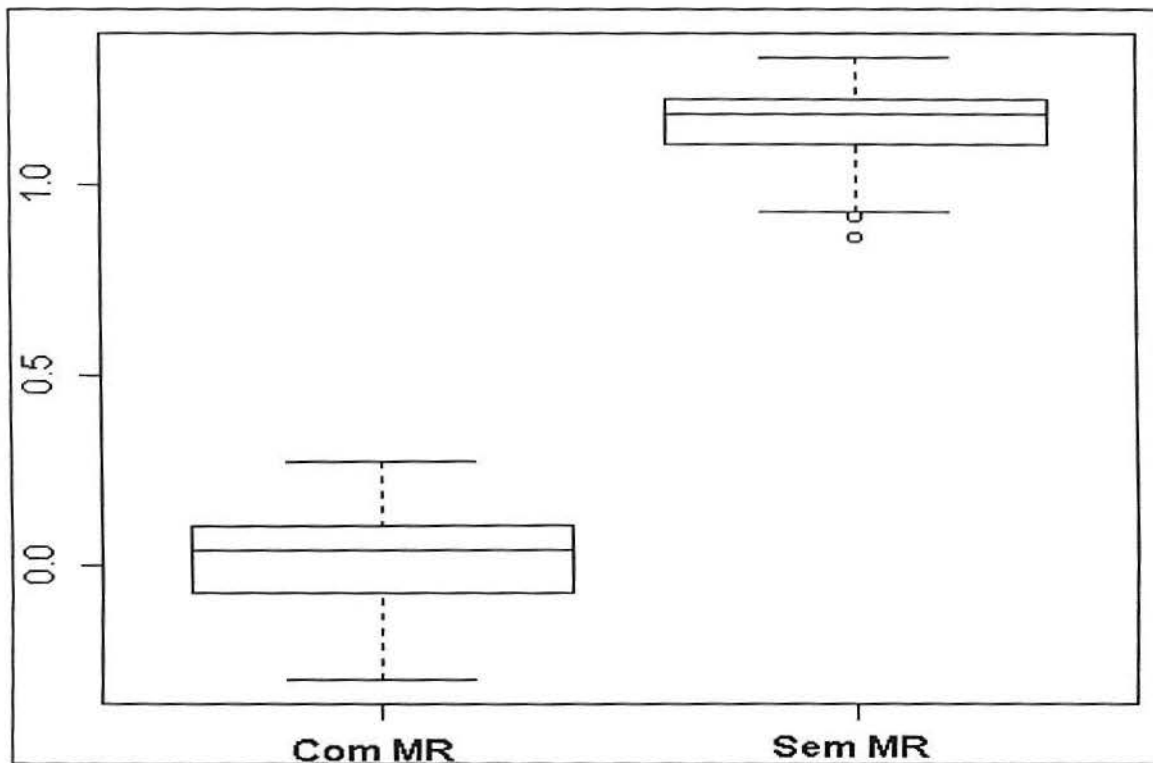


Figura 5.10: Boxplot comparando os BIC's obtidos para os modelos com e sem MR das 50 séries geradas

### 5.1.3. CASO 3

Identicamente ao Caso 2 geramos uma série temporal com base nas probabilidades de transição de uma cadeia de Markov para representar os dois regimes.

A diferença entre o Caso 2 para este serão os parâmetros autorregressivos simulados para cada regime, que aqui serão:

---

$c_1 = 2$	$c_2 = 1$	$\phi_1 = 0,8$	$\phi_2 = 0,2$	$\sigma_\varepsilon^2 = 1$
-----------	-----------	----------------	----------------	----------------------------

---

Quadro 5.7: Valores iniciais para o processo de estimação

Desta maneira obtivemos o seguinte gráfico para a série:

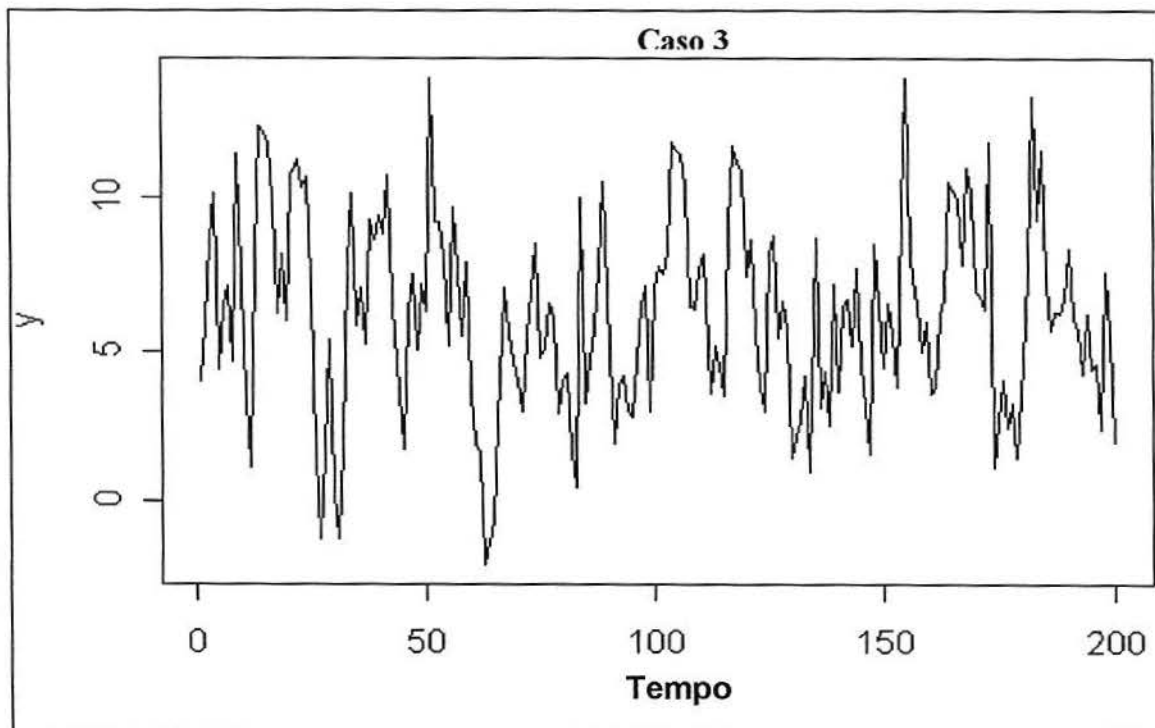


Figura 5.11: Gráfico da série para o Caso 3.

Abaixo apresentamos a ACF e PACF para o Caso 3:

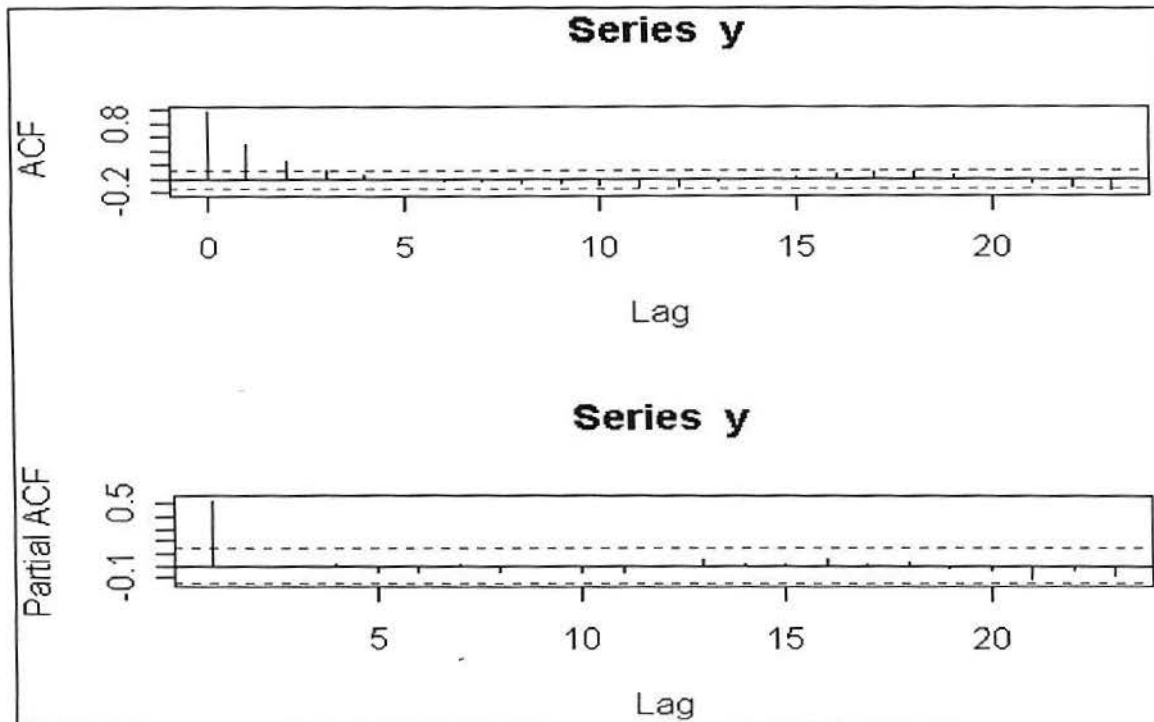


Figura 5.12: ACF e PACF para o Caso 3.

Pela Figura 5.11, verificamos que faz sentido olharmos para a série considerando que ela apresente dois regimes. Na Figura 5.12 vemos que a série é estacionária e um modelo possível seria um AR(1), pois a ACF cai rapidamente a medida que aumenta o *lag* e a função PACF apresenta autocorrelação significativamente positiva apenas para *lag* 1.

Portanto, verificaremos o ajuste de um modelo MS(2)-AR(1). Foram fornecidas as seguintes informações iniciais:

- número de regimes,  $N=2$ ;
- tamanho da amostra,  $T = 200$ ;
- $c_1$  e  $c_2$  serão respectivamente o valor da média amostral menos um desvio padrão e a média amostral mais um desvio padrão, seguindo a idéia apresentada no caso anterior;
- $\phi_1 = 0,4$  e  $\phi_2 = 0,6$ ;
- a variância do erro, onde utilizou-se a variância amostral da série;
- as probabilidades de transição em uma etapa,  $p_{ij} = 0,5$ , para  $i$  e  $j = 1, 2$ ; e
- O critério de convergência adotado foi que o algoritmo só parasse o processo de estimação dos parâmetros quando a diferença entre a estimativa atual e a estimativa imediatamente anterior fosse menor que 0,00001;

Dadas as estimativas iniciais acima, após 39 ciclos iterativos, obtivemos as seguintes estimativas para os parâmetros do vetor  $\theta$ :

$\hat{c}_1 = 1,72$	$\hat{c}_2 = 6,91$	$\hat{\phi}_1 = 0,58$	$\hat{\phi}_2 = 0,35$	$\hat{\sigma}^2 = 4,65$
$\hat{p}_{11} = 0,77$	$\hat{p}_{12} = 0,23$	$\hat{p}_{21} = 0,73$	$\hat{p}_{22} = 0,27$	

Quadro 5.8: Estimativas para o vetor de parâmetros  $\theta$ .

Portanto, o modelo final ajustado foi um MS(2)-AR(1), que forneceu as seguintes equações autorregressivas para cada regime da série:

Regime 1	$y_t = 1,72 + 0,58y_{t-1} + \varepsilon_t$
Regime 2	$y_t = 6,91 + 0,35y_{t-1} + \varepsilon_t$

Quadro 5.9: Equações autorregressivas para cada regime.

As estimativas das probabilidades suavizadas estão apresentadas abaixo:

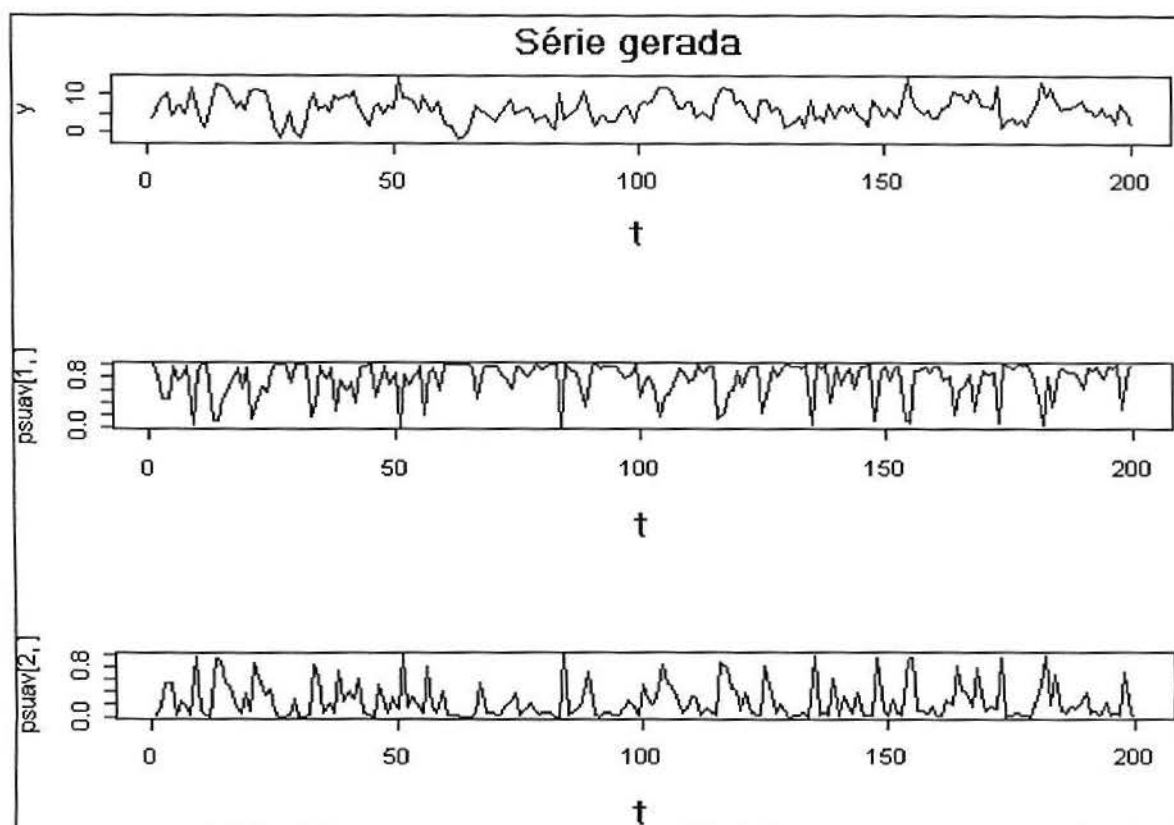


Figura 5.13: Gráfico da série com as probabilidades estimadas suavizadas para os respectivos regimes 1 e

Neste caso novamente verificamos que o processo subjacente apresenta altas probabilidades suavizadas estimadas de estar no regime 1 quando a série apresenta valores baixos. Assim, as probabilidades suavizadas estimadas de que o processo esteja no regime 2 são baixas.

Pelas Figuras 5.14 e 5.15 apresentadas a seguir podemos perceber que para o presente caso os modelos com mudança de regime forneceram os melhores ajustes, de acordo com os critérios de comparação AIC e BIC. Para o presente caso, o modelo MS(2)-AR(1) se ajusta melhor a série em relação ao modelo AR(1).

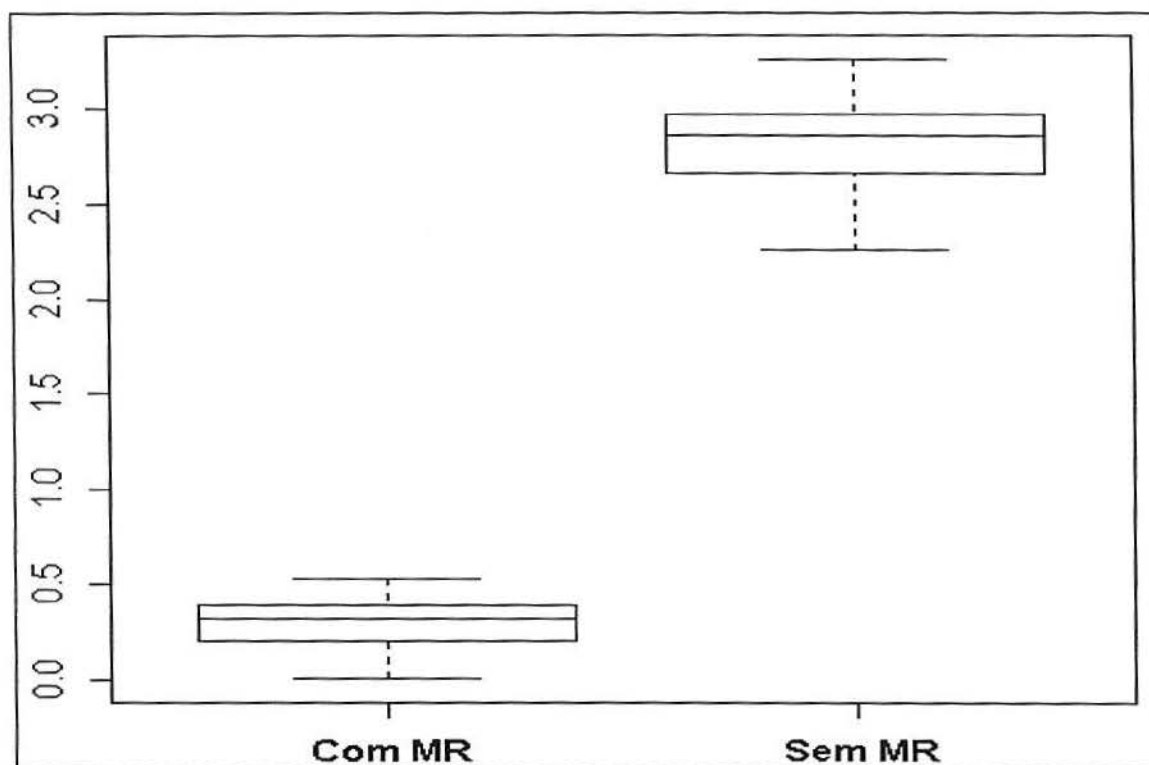


Figura 5.14: Boxplot comparando os AIC's obtidos para os modelos com e sem MR das 50 séries geradas

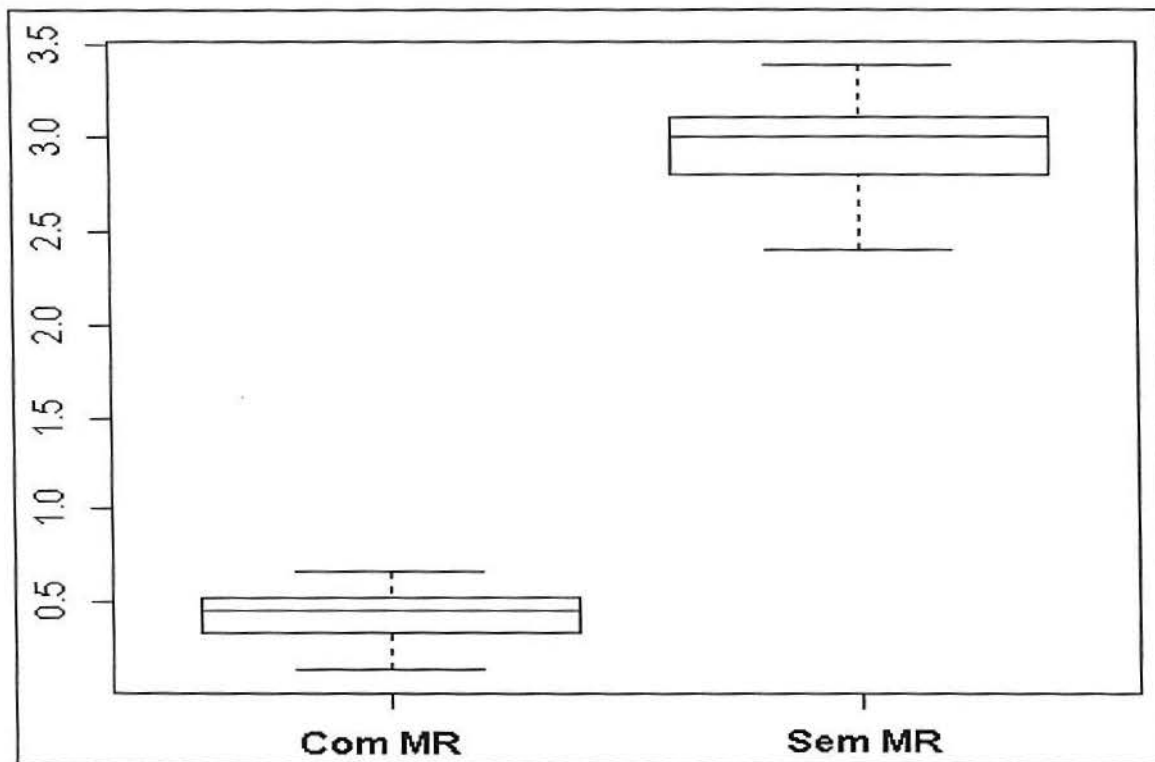


Figura 5.15: Boxplot comparando os BIC's obtidos para os modelos com e sem MR das 50 séries geradas

## 6. APLICAÇÃO DA METODOLOGIA MS-AR A UM CASO REAL

Neste capítulo, apresentaremos a aplicação de um modelo MS-AR a um conjunto de dados reais, observados trimestralmente, que representa o Produto Interno Bruto dos Estados Unidos (GNP), no período compreendido entre o primeiro trimestre de 1947 e o primeiro trimestre de 1989, totalizando 169 observações. Os dados representam o GNP em bilhões de dólares. Este mesmo conjunto de dados já foi analisado por Hamilton (1989), porém agora vamos sugerir um modelo distinto daquele proposto por Hamilton. Os dados podem ser obtidos na “homepage” de Hamilton, na internet (<http://econ.ucsd.edu/~jhamilto/>).

Na figura abaixo apresentamos a série original dos dados “GNP”:

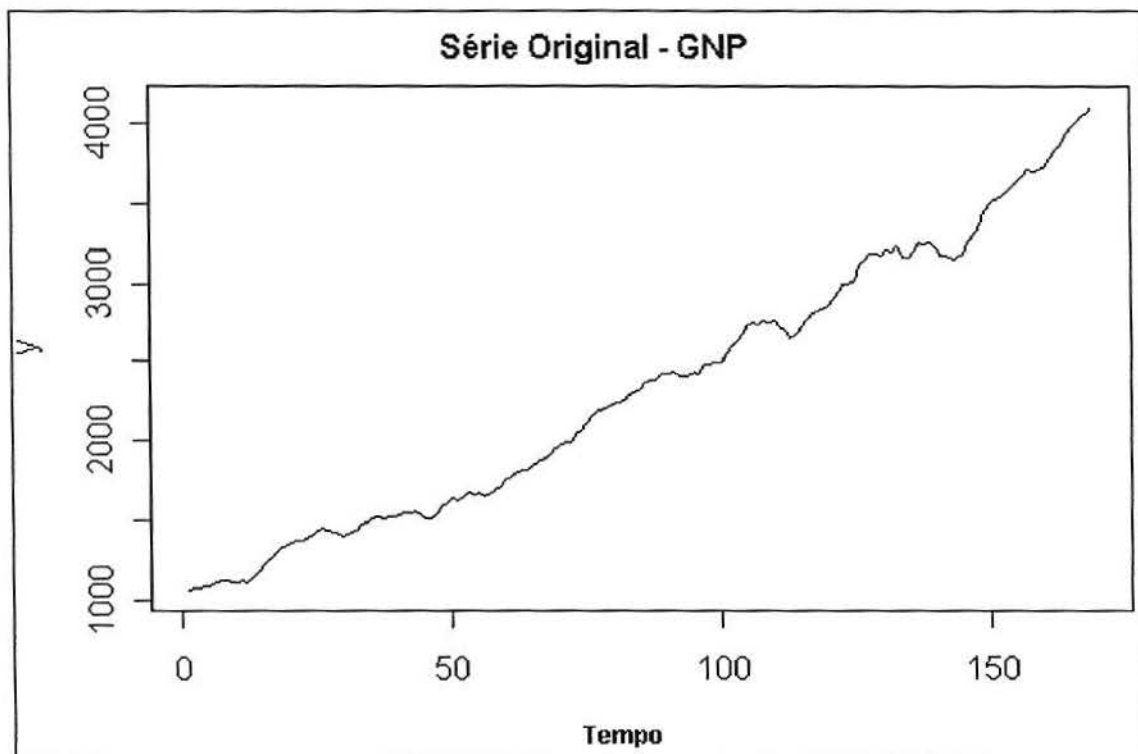


Figura 6.1: Série GNP.

A seguir são apresentados os gráficos das funções ACF e PACF.

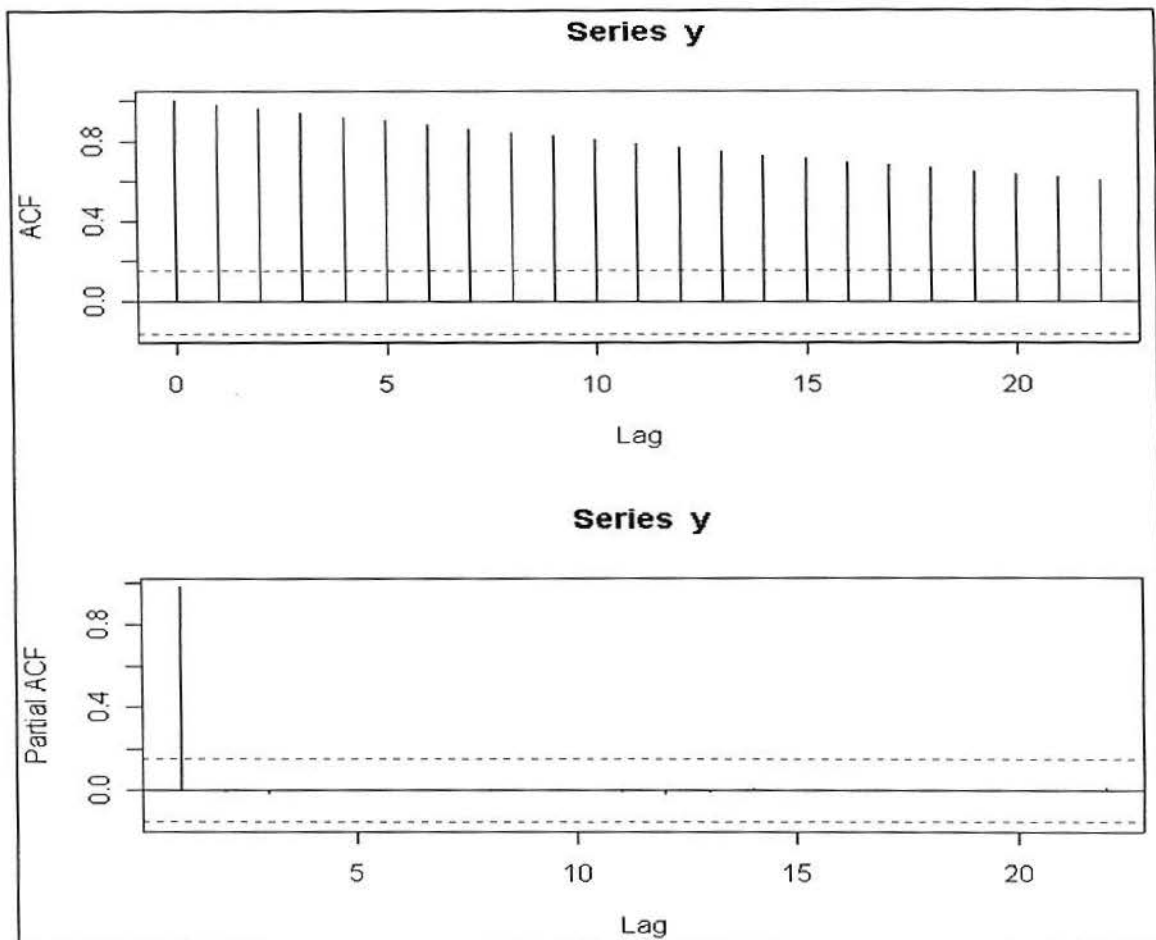


Figura 6.2: ACF e PACF da série GNP.

Pelo gráfico apresentado na Figura 6.1, vemos que a série não é estacionária, uma vez que há uma tendência crescente ao longo do tempo. A não-estacionariedade da série fica evidente olhando para o gráfico da ACF mostrado na Figura 6.2. Sendo assim, precisamos remover esta não-estacionariedade dos dados tomando-se quantas diferenças forem necessárias. Tomando-se a primeira diferença, obtivemos uma série estacionária, como pode ser verificado na figura abaixo:



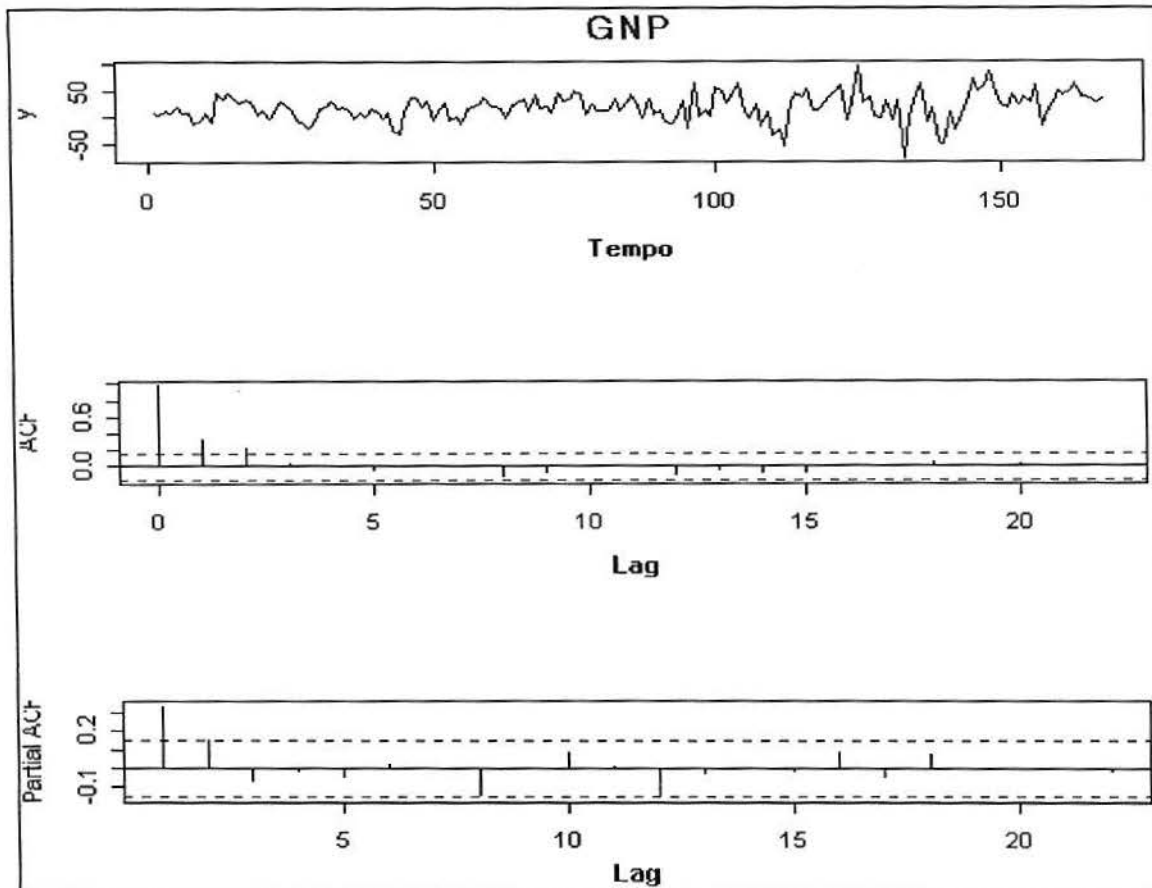


Figura 6.3: Primeira diferença da série GNP, ACF e PACF.

Verificamos que a primeira diferença da série GNP é estacionária uma vez que a ACF passou rapidamente a apresentar *lags* não significativos. Pela função PACF, vemos que parece plausível ajustar um modelo AR(1).

Vamos supor que a série possa ser governada por dois regimes, onde o regime 1 será de recessão e o regime 2 de ascensão econômica. Portanto, sugerimos o ajuste de um modelo MS(2)-AR(1) após ter sido tirada a primeira diferença da série original, ou seja, um modelo MS(2)-ARIMA(1,1,0).

Se, por outro lado, ajustarmos um modelo ARIMA(1,1,0) à série original, obteremos as seguintes estimativas de máxima verossimilhança:

$$\hat{c} = 42,00, \quad \hat{\phi} = 0,5676, \quad \hat{\sigma}^2 = 656,4.$$

Quadro 6.1: Estimativas dos parâmetros do modelo ARIMA(1,1,0)

Porém, como supomos haver mudança de regime na série, ajustaremos um modelo com mudança de regime. As estimativas iniciais foram aproximadas baseando-se no modelo ARIMA (1,1,0) proposto acima, conforme o quadro abaixo:

$\hat{\phi}_1 = 0,5042$	$\hat{\phi}_2 = 0,631$	$\hat{c}_1 = 36,63$	$\hat{c}_2 = 49,21$	$\hat{\sigma}^2 = 656,4$
$\hat{p}_{11} = 0,5$	$\hat{p}_{12} = 0,5$	$\hat{p}_{21} = 0,5$	$\hat{p}_{22} = 0,5$	

Quadro 6.2: Estimativas iniciais para o vetor de parâmetros  $\theta$ .

A rotina de estimação forneceu os resultados apresentados na tabela a seguir após o método iterativo de estimação ter sido realizado em 29 etapas. Note que o critério de convergência é de  $10^{-5}$  (0,00001) para cada parâmetro do modelo.

$\hat{\phi}_1 = 0,14$	$\hat{\phi}_2 = 0,9$	$\hat{c}_1 = 14,06$	$\hat{c}_2 = 5,29$	$\hat{\sigma}^2 = 484,54$
$\hat{p}_{11} = 0,67$	$\hat{p}_{12} = 0,33$	$\hat{p}_{21} = 0,68$	$\hat{p}_{22} = 0,32$	

Quadro 6.3: Estimativas finais para o vetor de parâmetros  $\theta$ .

Analisando-se as estimativas para  $\phi_1$ ,  $\phi_2$ ,  $c_1$  e  $c_2$ , percebe-se que o regime 2 representa períodos de expansão econômica, enquanto o regime 1 representa períodos de recessão  $\left( \frac{\hat{c}_2}{1-\hat{\phi}_2} > \frac{\hat{c}_1}{1-\hat{\phi}_1} \right)$ . Além disso, momentos de recessão são menos correlacionados com o passado do que os períodos de expansão ( $\hat{\phi}_1 = 0,14$  enquanto  $\hat{\phi}_2 = 0,9$ ).

No que diz respeito às estimativas das probabilidades de transição, elas nos indicam que os períodos de recessão são mais “persistentes” em relação aos períodos de expansão. Além disso, a longo prazo, as probabilidades ergódicas recorrentes indicam que há mais chance de em um determinado momento estarmos em recessão do que em expansão.

Analisaremos agora as probabilidades suavizadas para cada regime, conforme a figura abaixo:

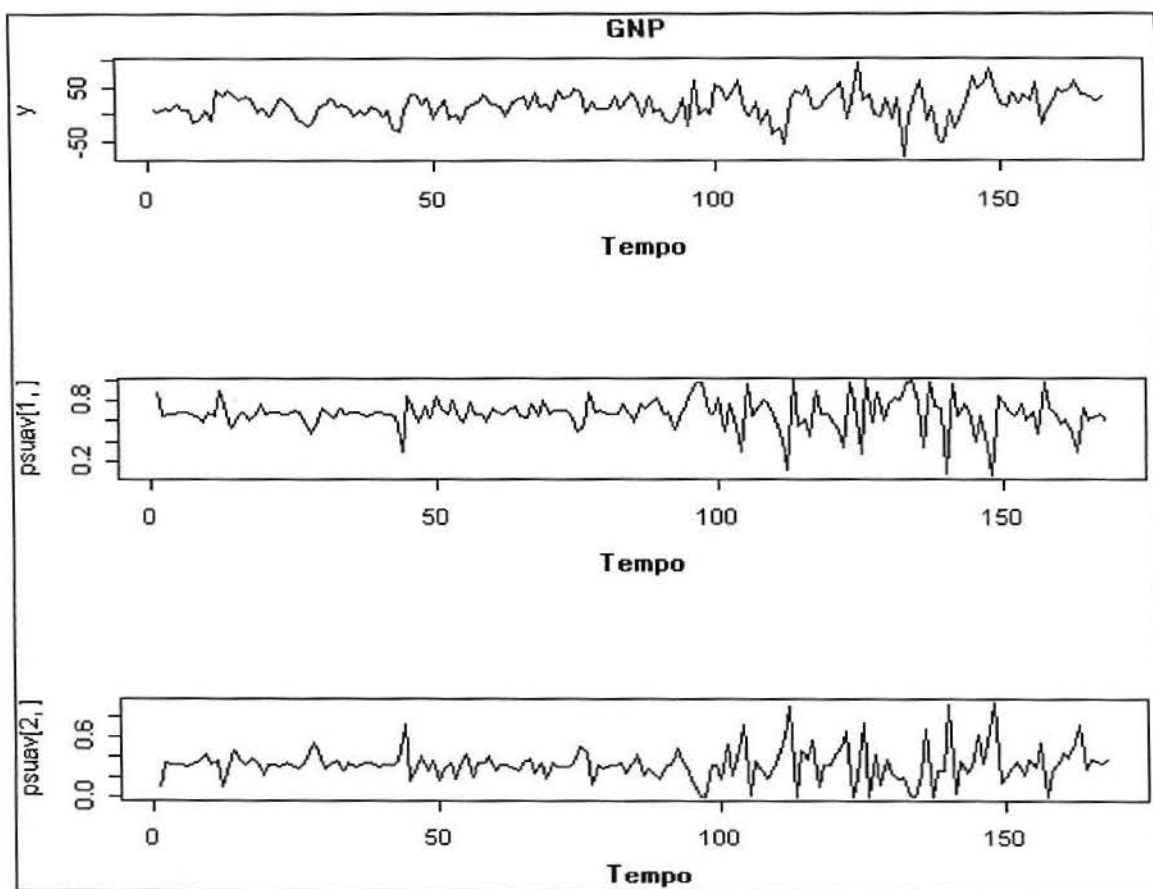


Figura 6.4: Gráfico da série GNP diferenciada e probabilidades suavizadas estimadas dos regimes 1 e 2

Através dos gráficos apresentados na Figura 6.4, vemos as probabilidades das observações estarem num regime de recessão ou estagnação econômica (regime 1) e de estarem num regime de ascensão econômica (regime 2). Assim, percebemos que a maior parte das observações apresenta maior probabilidade de “ingressar” num regime de recessão econômica e a medida que o processo avança surgem as maiores probabilidades de ascensão econômica.

Para verificar qual dos modelos se ajusta melhor aos dados, fizemos comparações com base no critério de Akaike (AIC) e no critério de Schwarz (BIC). O modelo com mudança de regime apresentou um grande número de parâmetros em relação ao modelo ARIMA (8 contra 2, pois foram incluídas as probabilidades de transição), o que inflacionou o valor de AIC e o de BIC.

Para a série analisada, verificamos que o modelo ARIMA(1,1,0) apresentou AIC igual a 7,49 e BIC igual a 7,53, enquanto que o modelo com mudança de regime MS(2)-ARIMA(1,1,0) apresentou AIC igual a 6,19 e BIC igual a 6,34. Ou seja, o modelo com mudança de regime proposto se ajusta melhor aos dados referentes ao Índice Geral de Produção Nacional dos E.U.A. no período analisado.

## 7. CONCLUSÃO

A partir da realização deste trabalho, percebemos que existem outros interessantes modelos para análise de séries temporais a serem abordados, que não os comumente vistos em um curso de graduação. Entre eles, podemos citar os modelos com mudança de regime, que foram o foco principal na realização deste estudo.

Como a maioria dos modelos de séries temporais, modelos com mudança de regime podem ser usados nas mais diversas áreas de conhecimento, embora até o momento não tenhamos visto muitos exemplos de aplicação que não sejam em áreas financeiras e econômicas.

Ao longo deste trabalho procuramos aprimorar o conhecimento a respeito das Cadeias de Markov, que, como pôde ser visto, são de fundamental relevância no estudo de modelos MS-AR. Nesses modelos, assumimos que estas cadeias governam o comportamento da série temporal. Além disso, foi preciso revisar os conhecimentos a respeito da metodologia de Box e Jenkins, utilizando-a para o ajuste de modelos do tipo SARIMA(p,d,q)(P,D,Q)<sub>s</sub>. Entretanto o enfoque principal foi dado aos modelos AR(p), seguindo o mesmo rumo dado em Hamilton (1994, Capítulo 22).

Verificamos que, para alguns modelos geradores que apresentavam alterações de comportamento ao longo do tempo, o uso de modelos com mudança de regime foi extremamente relevante. Além disso, inclusive para as particulares séries geradas para o caso sem mudança de regime, o ajuste do modelo com mudança de regime foi superior em termos do critério AIC e BIC.

Um elemento importante para fins de interpretação dos modelos ajustados com mudança de regime é a obtenção das probabilidades suavizadas estimadas, encontradas através de um processo iterativo envolvendo probabilidades preditas e probabilidades filtradas estimadas. Assim, podemos inferir sobre qual dos  $N$  regimes é responsável por uma determinada observação no tempo  $t$ .

Também não podem deixar de ser citadas as estimativas das probabilidades de transição que governam as Cadeias de Markov, pois, com base nestas probabilidades, poderemos ver, por exemplo, o risco de um determinado país passar a apresentar um regime de recessão econômica, dado que este se encontra num regime de expansão ou crescimento econômico. Além disso, probabilidades ergódicas estimadas podem ser obtidas.

O fato de o processo de obtenção das estimativas dos parâmetros do modelo ser feito através de várias etapas de um processo iterativo possibilitou-nos acompanhar a evolução das estimativas até que estas convergissem, onde percebemos que, quanto melhores (ou mais próximas) forem as estimativas iniciais dos parâmetros, menor o tempo que o processo de estimação levará para ser concluído. Na busca pelas melhores estimativas possíveis, foi adotado um rigoroso critério de convergência, o que fez com que praticamente não houvesse diferença entre as estimativas finais, dado que as estimativas iniciais poderiam ser levemente equivocadas.

Porém, a grande vantagem da realização deste trabalho foi a experiência acumulada principalmente em análise estatística de séries temporais, tanto pelas horas de estudo dedicadas para a realização do trabalho, quanto pelo desenvolvimento de rotinas de programação e algoritmos de previsão e simulação em linguagem *S* (utilizada pelo software *R*), que até então era praticamente desconhecida de minha parte.

Sendo assim, concluímos que o trabalho realizado foi de grande valia, uma vez que gerou conhecimento e experiência em diversas técnicas de análise estatística de série temporais, sempre levando-se em conta o rigor estatístico.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, N.M.C.G. e PEREIRA, P.L.V., (2002). Mudança de Regime para Séries Financeiras: Um Estudo Empírico Destes Modelos para a Realização de Regras de Mercado. USP e IBMEC Business Scholl, São Paulo, SP.

ANDERSON, T.W., (1971). *The Statistical Analysis of Times Series*. New York, John Wiley, 478.

ANDREWS, D.W.K. & PLOBERGER, W., (1994) Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative. *Econometrica* 62: 1383-1414.

BOLLERSLEV, T., ENGLE, R.F., NELSON, D.B., (1994). *ARCH Models: In Handbook of Econometrics, Amsterdam* (eds R. F. Engle and D. L. McFadden), Volume IV.

BOWERMAN, B.L. & O'CONNELL, R.T., (1993). *Forecasting and Time Series: An Applied Approach*. Third Edition. Miami University, Duxburg Press. Belmont, California.

BOX, G.E.P. & JENKINS, G.M., (1976). *Times Series Analysis: Forecasting and Control*. Holden-Day Inc., San Francisco, California.

DAVIES, R.B., (1977). Hypothesis testing When a Nuisance Parameter is Present Only under the Alternative. *Biometrika* 64:247-54.

DICKEY, D. & FULLER, W.A., (1979). Distribution of the Estimates for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74: 427-31. June.

ENDERS, W., (1995). *Applied Econometric Time Series*. Iowa State University. John Wiley & Sons, Inc.

HAMILTON, J.D., (1989). *A New Approach to the Economic Analysis of Nonstationary Times Series and the Business Cycle*. Princeton University Press. Princeton, New Jersey.

HAMILTON, J.D., (1994). *Time Series Analysis*. Princeton University Press. Princeton, New Jersey.

HAMILTON, J.D., (1996). *Specification testing in Markov-Switching Time Series Models*. University of California, San Diego.

HAMILTON, J.D., (2002). *Advances in Markov-Switching Models*. Princeton University Press. Princeton, New Jersey.

HANSEN, B.E., (1992). The Likelihood ratio Test under Non-standard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics* 7: S61-82.

KIM, C. J., (1993). Dynamic Linear Models with Markov Switching. *Journal of Econometrics* 60:1-22.

LAY, D., (1999). *Álgebra Linear e suas Aplicações*. LTC – Livros Técnicos e Científicos. Rio de Janeiro, RJ.

MORETTIN, P.A. & TOLOI, C.M.C., (1981). *Modelos para Previsão de Séries Temporais*. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ.

REISEN, V. A., ABRAHAN, B., LOPES, S.R.C., (2000). *Estimation of the Parameters in ARFIMA Processes: A Simulation Study*. Communications In Statistics Simulation and Computing.

SCHUMMWAY, R.H., (1988) *Applied Statistical Time Series Analysis*. Prentice Hall.

STINCHCOMBE, M. & WHITE, H., (1993) An approach to Consistent Specification Testing Using Duality and Banach Limit Theory. University of California, San Diego, Mimeo.

ZIEGELMAN, F.A., (2002) *Estimation of Volatility Functions: Nonparametrics and Semi-parametric Methods*. Ph.D. Dissertation. University of Kent at Canterbury, UK.



## 9. ANEXOS: PROGRAMAS EM R

**Obs.:** não serão aqui apresentadas todas as rotinas de programação utilizadas neste trabalho. Apresentaremos apenas as rotinas utilizadas que não apresentam grandes semelhanças de caso para caso. Motivo: a estrutura das rotinas é a mesma, mudando apenas os parâmetros da série simulada e as estimativas iniciais dos parâmetros.

### SIMULAÇÃO DA SÉRIE UTILILIZADA NO CASO 1

```
set.seed(17)
g <- c(.7,.6,1,1)
ga <- matrix(g,2,2,) # matriz acumulada
T <- 200 # tamanho da amostra
est <- rep(0,T)
s <- c(1,2) # espaço dos estados
p0 <- c(2/3,1/3) # probabilidades iniciais de cada estado
f <- sample(s,size=1,prob=p0) # sorteio do estado inicial
est[1] <- f
# a seguir é feita a trajetória aleatória
for(i in 2:T) {
  u <- runif(1,0,1)
  j <- 1
  while(u > ga[est[i-1],j])
    {j <- j+1}
  est[i] <- j }
# Simulando um MS(2)-AR(1)
y <- rep(0,T)
c1 <- 2
c2 <- 2
fi1 <- .5
fi2 <- .5
erro=rnorm(T,0,.5)
y[1] <- 4
for (i in 2:T) {
  if (est[i]<2)
    {y[i] <- c1+(fi1*y[i-1])+erro[i]}
  else
    {y[i] <- c2+(fi2*y[i-1])+erro[i]} }
par(mfrow=c(3,1))
ts.plot(y) # gráfico da série
acf(y) # gráfico de autocorrelação amostral
pacf(y) # gráfico de autocorrelação parcial amostral
```

## FUNÇÃO UTILIZADA PARA ESTIMAR OS PARÂMETROS DO CASO 1, CASO 3 E SÉRIE GNP

```

vetpar<-function(y, N, T, def, c1, c2, fi1, fi2, var, p11a, p12a, p21a, p22a,
etain=matrix(c(.5), (.5)), nrow=N), x=(c(p11a, p12a, p21a, p22a)), P=matrix(x, nrow=N)) {
#Estimativas das densidades condicionais e das probabilidades filtradas e preditas
pfilta <- rep(0,N*T)
ppreda <- rep(0,N*T)
psuava <- rep(0,N*T)
pfiltf <- matrix(pfilta, nrow=N, ncol=T)
ppredf <- matrix(ppreda, nrow=N, ncol=T)
psuavf <- matrix(psuava, nrow=N, ncol=T)
ppredf[,1] <- etain
y01 <- c1/(1-fi1)
y02 <- c2/(1-fi2)
eta1 <- (2*pi*var)^(-.5)*exp(-.5*((y[1]-(c1+fi1*y01))^2/var))
eta2 <- (2*pi*var)^(-.5)*exp(-.5*((y[1]-(c2+fi2*y02))^2/var))
eta=matrix(c(eta1,eta2),nrow=N, ncol=1)
A1=ppredf[,1]*eta
B1=((ppredf[1,1])*(eta[1,1]))+((ppredf[2,1])*(eta[2,1]))
pfiltf[,1] <- A1/B1
for (i in 2:T) {
ppredf[,i] <- P%*%pfiltf[,i-1]
eta1 <- (2*pi*var)^(-.5)*exp(-.5*((y[i]-(c1+fi1*y[i-1]))^2/var))
eta2 <- (2*pi*var)^(-.5)*exp(-.5*((y[i]-(c2+fi2*y[i-1]))^2/var))
eta=matrix(c(eta1,eta2),nrow=2, ncol=1)
A=ppredf[,i]*eta
B=((ppredf[1,i])*(eta[1,1]))+((ppredf[2,i])*(eta[2,1]))
pfiltf[,i]=A/B }
#Estimativas das probabilidades suavizadas
psuavf[,T]=pfiltf[,T]
for (j in 1:(T-1)) { psuavf[,T-j]=pfiltf[,T-j]*(P%*%(psuavf[,T-j+1]/ppredf[,T-j+1])) }
#Estimativas dos parâmetros autorregressivos
ybarra <- 0
for (i in 1:T){
termo <- y[i]
ybarra <- ybarra+termo }
ybarra <- ybarra/T
ztin <- rep(0,N*T)
ztf <- matrix(ztin, nrow=2, ncol=T)
z1 <- matrix(c(1, ybarra), ncol=1)
ztf[,1] <- z1
for (i in 2:T){
ztf[,i] <- matrix(c(1, y[i-1]), nrow=2) }
zbar <- rep(0,def*T)
zbar1 <- matrix(zbar, nrow=def, ncol=T)
zbar2 <- matrix(zbar, nrow=def, ncol=T)
ybar <- rep(0,def*T)
ybar <- matrix(ybar, nrow=def, ncol=T)
beta1 <- rep(0,def)

```

```

beta2 <- rep(0,def)
for (i in 1:T) {
ybar[1,i]=y[i]*sqrt(psuavf[1,i])
ybar[2,i]=y[i]*sqrt(psuavf[2,i]) }
zbar1[,1]=z1*sqrt(psuavf[1,1])
zbar2[,1]=z1*sqrt(psuavf[2,1])
  for (i in 2:T) {
zbar1[,i]=ztf[,i]*sqrt(psuavf[1,i])
zbar2[,i]=ztf[,i]*sqrt(psuavf[2,i]) }
parte11 <- 0
  for (j in 1:T){
termo1 <- zbar1[,j]*%*%t(zbar1[,j])
parte11 <- parte11+termo1}
parte12 <- 0
  for (j in 1:T){
termo2 <- zbar2[,j]*%*%t(zbar2[,j])
parte12 <- parte12+termo2 }
parte21 <- 0
  for (j in 1:T){
termo3 <- zbar1[,j]*ybar[1,j]
parte21 <- parte21+termo3 }
parte22 <- 0
  for (j in 1:T){
termo4 <- zbar2[,j]*ybar[2,j]
parte22 <- parte22+termo4 }
beta1 <- solve(parte11)%*%parte21
beta2 <- solve(parte12)%*%parte22
beta <- matrix(c(beta1, beta2), ncol=N)
# Variância estimada
d <- 0
  for (i in 1:T) { for (j in 1:N) {
  termo5 <- ((y[i] - t(ztf[,i])%*%beta[,j])^2)*psuavf[j,i]
  d <- d+termo5 } }
varest <- d/T
#Probabilidades de transição
p1ja <- rep(0,N*N)
p1jf <- matrix(p1ja, nrow=N, ncol=N)
p11 <- 0
p11c <- 0
p11m <- 0
p12 <- 0
p12c <- 0
p12m <- 0
p21 <- 0
p21c <- 0
p21m <- 0
p22 <- 0
p22c <- 0
p22m <- 0
  for (t in 2:T){
  termoc11 <- psuavf[1,t-1]*psuavf[1,t]
  p11c <- p11c+termoc11
  termom11 <- psuavf[1,t-1]

```

```

p11m <- p11m+termom11
p11 <- p11c/p11m
termoc12 <- psuavf[1,t-1]*psuavf[2,t]
p12c <- p12c+termoc12
termom12 <- psuavf[1,t-1]
p12m <- p12m+termom12
p12 <- p12c/p12m
termoc21 <- psuavf[2,t-1]*psuavf[1,t]
p21c <- p21c+termoc21
termom21 <- psuavf[2,t-1]
p21m <- p21m+termom21
p21 <- p21c/p21m
termoc22 <- psuavf[2,t-1]*psuavf[2,t]
p22c <- p22c+termoc22
termom22 <- psuavf[2,t-1]
p22m <- p22m+termom22
p22 <- p22c/p22m }
pijf[1,1] <- p11
pijf[1,2] <- p21
pijf[2,1] <- p12
pijf[2,2] <- p22
etai <- psuavf[,1]
c1f <- beta[1,1]
dif1 <- (c1f-c1)^2
c2f <- beta[1,2]
dif2 <- (c2f-c2)^2
fi1f <- beta[2,1]
dif3 <- (fi1f-fi1)^2
fi2f <- beta[2,2]
dif4 <- (fi2f-fi2)^2
varf <- varest
dif5 <- (varf-var)^2
p11f <- pijf[1,1]
dif6 <- (p11f-p11a)^2
p12f <- pijf[2,1]
dif7 <- (p12f-p12a)^2
p21f <- pijf[1,2]
dif8 <- (p21f-p21a)^2
p22f <- pijf[2,2]
dif9 <- (p22f-p22a)^2
diffinal <- dif1+dif2+dif3+dif4+dif5+dif6+dif7+dif8+dif9
# Previsões
h <- rep(0, N*T)
h <- matrix(h, nrow=N)
h[1,1] <- c1 + fi1*y01
h[2,1] <- c2 + fi2*y02
for (i in 2:T){
h[1,i] <- c1 + fi1*y[i-1]
h[2,i] <- c2 + fi2*y[i-1] }
esp <- rep(0, T)
esp <- matrix(esp, ncol=T)
for (i in 1:T){ esp[,i] <- (t(h[,i]))%*%ppredf[,i] }
return(P, ppredf, pi1f, psuavf, ybar, zbar1, zbar2, esp, beta, varest, pijf, diffinal) }

```

## PROCESSO DE ESTIMAÇÃO REALIZADO NO CASO 1

```
# Estimativas iniciais
N <- 2
T <- 200
def <- 2
c1 <- mean(y)-.25*sd(y)
c2 <- mean(y)+.25*sd(y)
fi1 <- .6
fi2 <- .4
var <- var(y)
p11a <- .45
p12a <- .55
p21a <- .55
p22a <- .45
dif <- 99
cont <- 0
# Critério de convergência
while(dif>.00001) {
# Obtenção das estimativas
cont <- cont+1
a <- vetpar(y, N, T, def, c1, c2, fi1, fi2, var, p11a, p12a, p21a, p22a)
c1 <- a$beta[1,1]
c2 <- a$beta[1,2]
fi1 <- a$beta[2,1]
fi2 <- a$beta[2,2]
var <- a$varest
p11a <- a$pijf[1,1]
p12a <- a$pijf[2,1]
p21a <- a$pijf[1,2]
p22a <- a$pijf[2,2]
dif <- a$diffinal }
par(mfrow=c(3,1))
ts.plot(y) # gráfico da série
psuav <- a$psuavf
ts.plot(psuav[1,]) # gráfico das probabilidades suavizadas estimadas para o regime 1
ts.plot(psuav[2,]) # gráfico das probabilidades suavizadas estimadas para o regime 2
```

## COMPARAÇÃO ATRAVÉS DOS CRITÉRIOS AIC E BIC PARA O CASO 1

```
ns <- 50 # número de séries geradas
g <- c(.7,.6,1,1)
ga <- matrix(g,2,2,) # matriz acumulada
T <- 200 # tamanho da amostra
est <- rep(0,T)
s <- c(1,2) # espaço dos estados
p0 <- c(2/3,1/3) # probabilidades iniciais de cada estado
varSMR <- rep(0, ns) # variância dos resíduos do modelo sem mudança de regime
varMR <- rep(0, ns) # variância dos resíduos do modelo com mudança de regime
AICSMR <- rep(0, ns) # Estatística de Akaike para um modelo sem mudança de regime
AICMR <- rep(0, ns) # Estatística de Akaike para um modelo com mudança de regime
BICSMR <- rep(0, ns) # Estatística de Schwarz para um modelo sem mudança de regime
BICMR <- rep(0, ns) # Estatística de Schwarz para um modelo com mudança de regime
for (k in 1:ns) {
  f <- sample(s,size=1,prob=p0) # sorteio do estado inicial
  est[1] <- f
  # a seguir é feita a trajetória aleatória
  for (i in 2:T) {
    u <- runif(1,0,1)
    j <- 1
    while(u > ga[est[i-1],j])
      {j <- j+1}
    est[i] <- j
    # Simulando um MS(2)-AR(1)
    y <- rep(0,T)
    c1 <- 2
    c2 <- 2
    fi1 <- .5
    fi2 <- .5
    erro=rnorm(T,0,.5)
    y[1] <- 4
    for (i in 2:T) {
      if (est[i]<2)
        {y[i] <- c1+(fi1*y[i-1])+erro[i]}
      else
        {y[i] <- c2+(fi2*y[i-1])+erro[i]}
    }
    # Estimativas iniciais
    N <- 2
    def <- 2
    c1 <- 3.75
    c2 <- 4.25
    fi1 <- .6
    fi2 <- .4
    var <- var(y)
    p11a <- .45
    p12a <- .55
    p21a <- .55
    p22a <- .45
    dif <- 99
  }
}
```

```

cont <- 0
# Critério de convergência
while(dif>.00001) {
# Obtenção das estimativas
cont <- cont+1
a <- vetpar(y, N, T, def, c1, c2, fi1, fi2, var, p11a, p12a, p21a, p22a)
c1 <- a$beta[1,1]
c2 <- a$beta[1,2]
fi1 <- a$beta[2,1]
fi2 <- a$beta[2,2]
var <- a$varest
p11a <- a$pijf[1,1]
p12a <- a$pijf[2,1]
p21a <- a$pijf[1,2]
p22a <- a$pijf[2,2]
p11a <- a$pijf[1,1]
dif <- a$diffinal }
psuav <- a$psuavf
# Obtenção dos resíduos, AIC e BIC
ary <- arima(y, order=c(1,0,0))
c <- ary$coef[2]
fi <- ary$coef[1]
vetsem <- rep(0,T-1)
vetest <- rep(0,T-1)
ress <- rep(0, T-1)
resc <- rep(0, T-1)
for (l in 2:T) {
vetsem[l-1] <- c+fi*y[l-1]
ress[l-1] <- (y[l]-vetsem[l-1])^2
if ( psuav[1,l]>.5)
    vetest[l-1] <- c1+fi1*y[l-1]
    else
    vetest[l-1] <- c2+fi2*y[l-1]
resc[l-1] <- (y[l]-vetest[l-1])^2 }
varSMR[k] <- mean(ress)
varMR[k] <- mean(resc)
AICSMR[k] <- log(varSMR[k])+(2*2)/T
AICMR[k] <- log(varMR[k])+(2*8)/T
BICSMR[k] <- log(varSMR[k])+(2*log(T))/T
BICMR[k] <- log(varMR[k])+(8*log(T))/T }

```

## FUNÇÃO UTILIZADA PARA ESTIMAR OS PARÂMETROS DO CASO 2

Neste caso, a função apresentada difere das demais por não conter os parâmetros autorregressivos  $fi1$  e  $fi2$ .

```
vetpar<-function(y, N, T, def, c1, c2, fi1, fi2, var, p11a, p12a, p21a, p22a,
etain=matrix(c(.5), (.5)), nrow=N), x=(c(p11a, p12a, p21a, p22a)), P=matrix(x, nrow=N)) {
#Estimativas das densidades condicionais e das probabilidades filtradas e preditas
pfilta <- rep(0,N*T)
ppreda <- rep(0,N*T)
psuava <- rep(0,N*T)
pfiltf <- matrix(pfilta, nrow=N, ncol=T)
ppredf <- matrix(ppreda, nrow=N, ncol=T)
psuavf <- matrix(psuava, nrow=N, ncol=T)
ppredf[,1] <- etain
y01 <- c1/(1-fi1)
y02 <- c2/(1-fi2)
eta1 <- (2*pi*var)^(-.5)*exp(-.5*((y[1]-(c1+fi1*y01))^2/var))
eta2 <- (2*pi*var)^(-.5)*exp(-.5*((y[1]-(c2+fi2*y02))^2/var))
eta=matrix(c(eta1,eta2),nrow=N, ncol=1)
A1=ppredf[,1]*eta
B1=((ppredf[1,1])*(eta[1,1]))+((ppredf[2,1])*(eta[2,1]))
pfiltf[,1] <- A1/B1
for (i in 2:T) {
ppredf[,i] <- P%*%pfiltf[,i-1]
eta1 <- (2*pi*var)^(-.5)*exp(-.5*((y[i]-(c1+fi1*y[i-1]))^2/var))
eta2 <- (2*pi*var)^(-.5)*exp(-.5*((y[i]-(c2+fi2*y[i-1]))^2/var))
eta=matrix(c(eta1,eta2),nrow=2, ncol=1)
A=ppredf[,i]*eta
B=((ppredf[1,i])*(eta[1,1]))+((ppredf[2,i])*(eta[2,1]))
pfiltf[,i]=A/B }
#Estimativas das probabilidades suavizadas
psuavf[,T]=pfiltf[,T]
for (j in 1:(T-1)) { psuavf[,T-j]=pfiltf[,T-j]*(t(P)%*(psuavf[,T-j+1]/ppredf[,T-j+1])) }
#Estimativas dos parâmetros autorregressivos
ybarra <- 0
for (i in 1:T){
termo <- y[i]
ybarra <- ybarra+termo }
ybarra <- ybarra/T
ztin <- rep(0,N*T)
ztf <- matrix(ztin, nrow=2, ncol=T)
z1 <- matrix(c(1, ybarra), ncol=1)
ztf[,1] <- z1
for (i in 2:T){
ztf[,i] <- matrix(c(1, y[i-1]), nrow=2) }
zbar <- rep(0,def*T)
zbar1 <- matrix(zbar, nrow=def, ncol=T)
zbar2 <- matrix(zbar, nrow=def, ncol=T)
ybar <- rep(0,def*T)
ybar <- matrix(ybar, nrow=def, ncol=T)
beta1 <- rep(0,def)
```



```

beta2 <- rep(0,def)
for (i in 1:T) {
ybar[1,i]=y[i]*sqrt(psuavf[1,i])
ybar[2,i]=y[i]*sqrt(psuavf[2,i]) }
zbar1[,1]=z1*sqrt(psuavf[1,1])
zbar2[,1]=z1*sqrt(psuavf[2,1])
  for (i in 2:T) {
zbar1[,i]=ztf[,i]*sqrt(psuavf[1,i])
zbar2[,i]=ztf[,i]*sqrt(psuavf[2,i]) }
parte11 <- 0
  for (j in 1:T){
termo1 <- zbar1[,j]*zbar1[,j]
parte11 <- parte11+termo1 }
parte12 <- 0
  for (j in 1:T){
termo2 <- zbar2[,j]*zbar2[,j]
parte12 <- parte12+termo2 }
parte21 <- 0
  for (j in 1:T){
termo3 <- zbar1[,j]*ybar[1,j]
parte21 <- parte21+termo3 }
parte22 <- 0
  for (j in 1:T){
termo4 <- zbar2[,j]*ybar[2,j]
parte22 <- parte22+termo4 }
beta1 <- solve(parte11)%*%parte21
beta2 <- solve(parte12)%*%parte22
beta <- matrix(c(beta1, beta2), ncol=N)
# Variância estimada
d <- 0
  for (i in 1:T) {
    for (j in 1:N) {
      termo5 <- ((y[i] - t(ztf[,i])%*%beta[,j])^2)*psuavf[j,i]
      d <- d+termo5 } }
varest <- d/T
#Probabilidades de transição:
pija <- rep(0,N*N)
pijf <- matrix(pija, nrow=N, ncol=N)
p11 <- 0
p11c <- 0
p11m <- 0
p12 <- 0
p12c <- 0
p12m <- 0
p21 <- 0
p21c <- 0
p21m <- 0
p22 <- 0
p22c <- 0
p22m <- 0
  for (t in 2:T){
    termoc11 <- psuavf[1,t-1]*psuavf[1,t]
    p11c <- p11c+termoc11

```

```

termom11 <- psuavf[1,t-1]
p11m <- p11m+termom11
p11 <- p11c/p11m
termoc12 <- psuavf[1,t-1]*psuavf[2,t]
p12c <- p12c+termoc12
termom12 <- psuavf[1,t-1]
p12m <- p12m+termom12
p12 <- p12c/p12m
termoc21 <- psuavf[2,t-1]*psuavf[1,t]
p21c <- p21c+termoc21
termom21 <- psuavf[2,t-1]
p21m <- p21m+termom21
p21 <- p21c/p21m
termoc22 <- psuavf[2,t-1]*psuavf[2,t]
p22c <- p22c+termoc22
termom22 <- psuavf[2,t-1]
p22m <- p22m+termom22
p22 <- p22c/p22m }
pijf[1,1] <- p11
pijf[1,2] <- p21
pijf[2,1] <- p12
pijf[2,2] <- p22
eta1 <- psuavf[,1]
c1f <- beta[1,1]
dif1 <- (c1f-c1)^2
c2f <- beta[1,2]
dif2 <- (c2f-c2)^2
varf <- varest
dif5 <- (varf-var)^2
p11f <- pijf[1,1]
dif6 <- (p11f-p11a)^2
p12f <- pijf[2,1]
dif7 <- (p12f-p12a)^2
p21f <- pijf[1,2]
dif8 <- (p21f-p21a)^2
p22f <- pijf[2,2]
dif9 <- (p22f-p22a)^2
diffinal <- dif1+dif2+dif5+dif6+dif7+dif8+dif9
# Previsões
h <- rep(0, N*T)
h <- matrix(h, nrow=N)
h[1,1] <- c1 + fi1*y01
h[2,1] <- c2 + fi2*y02
for (i in 2:T){
h[1,i] <- c1 + fi1*y[i-1]
h[2,i] <- c2 + fi2*y[i-1] }
esp <- rep(0, T)
esp <- matrix(esp, ncol=T)
for (i in 1:T){
esp[,i] <- (t(h[,i]))%*%ppredf[,i] }
return(P, ppredf, pfillf, psuavf, ybar, zbar1, zbar2, esp, beta, varest, pijf, diffinal) }

```

## PROCESSO DE ESTIMAÇÃO REALIZADO NA APLICAÇÃO A SÉRIE GNP

```
# Estimativas iniciais
N <- 2
T <- 169
lagy <- 2
c1 <- 18.16
c2 <- 49.21
fi1 <- .5042
fi2 <- .631
var <- var(y)
p11a <- .5
p12a <- .5
p21a <- .5
p22a <- .5
dif <- 99
cont <- 0
# Critério de convergência
while(dif>.00001) {
# Obtenção das estimativas
cont <- cont+1
a <- vetpar(y, N, T, lagy, c1, c2, fi1, fi2, var, p11a, p12a, p21a, p22a)
c1 <- a$beta[1,1]
c2 <- a$beta[1,2]
fi1 <- a$beta[2,1]
fi2 <- a$beta[2,2]
var <- a$varest
p11a <- a$pijf[1,1]
p12a <- a$pijf[2,1]
p21a <- a$pijf[1,2]
p22a <- a$pijf[2,2]
dif <- a$diffinal }
par(mfrow=c(3,1))
ts.plot(y) # gráfico da série
psuav <- a$psuavf
ts.plot(psuav[1,]) # gráfico das probabilidades suavizadas estimadas para o regime 1
ts.plot(psuav[2,]) # gráfico das probabilidades suavizadas estimadas para o regime 2
```

SÉRIE GNP:

#### OBTENÇÃO DE AIC E BIC PARA O MODELO COM MUDANÇA DE REGIME

```
psuav <- a$psuavf
resMR <- 0
K <- 8 # Número de parâmetros do modelo
vetc <- rep(0,T-1)
resc <- rep(0,T-1)
# Obtenção dos resíduos gerados pelo modelo
for (l in 2:T) {
  if ( psuav[l,]>.5) vetc[l-1]<-c1+fi1*dy[l-1]
  else
    vetc[l-1] <- c2+fi2*dy[l-1]
  resc[l-1] <- (dy[l]-vetc[l-1])^2
  resMR[l] <- resMR[l-1]+resc[l-1] }
varMR <- mean(resc)
AIC <- log(varMR)+(2*K)/T
BIC <- log(varMR)+(K*log(T))/T
```

#### OBTENÇÃO DE AIC E BIC PARA O MODELO SEM MUDANÇA DE REGIME

```
psuav <- a$psuavf
resSMR <- 0
K <- 2 # Número de parâmetros do modelo
vetc <- rep(0,T-1)
resc <- rep(0,T-1)
dy <- diff(y)
# Obtenção dos resíduos gerados pelo modelo
for (l in 2:T) {
  vetc[l-1]<-c1+fi1*dy[l-1]
  resc[l-1] <- (dy[l]-vetc[l-1])^2
  resSMR[l] <- resSMR[l-1]+resc[l-1] }
varSMR <- mean(resc)
AIC <- log(varSMR)+(2*K)/T
BIC <- log(varSMR)+(K*log(T))/T
```