MODELOS DE MUDANÇAS MARKOVIÁNAS DE REGIMES APLICADOS A SÉRIES TEMPORAIS FINANCEIRAS

Luiz Alberto Rabi Junior

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU

DE

MESTRE EM ESTATÍSTICA

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Pedro Luiz Valls Pereira**

-São Paulo, Junho de 1998-

Modelos de Mudanças Markovianas de Regimes Aplicados à Séries Temporais Financeiras

Este exemplar corresponde à
redação final da dissertação
devidamente corrigida e defendida
por Luiz Alberto Rabi Junior e
aprovada pela comissão julgadora.

São Paulo, 08 de junho de 1998

Banca Examinadora:

- Prof. Dr. Pedro Luiz Valls Pereira (Orientador) - IME-USP
- Prof. Dr. Luiz Koodi Hotta - IMECC-UNICAMP
- Prof. Dr. Marcos Eugênio da Silva - FEA-USP

Agradecimentos

Agradeço aos meus pais pelo constante apoio e incentivo.

Agradeço à minha esposa pela paciência e compreensão.

Agradeço ao Prof. Dr. Pedro L. Valls Pereira pela ótima orientação.

Agradeço ao CNPq pela bolsa concedida.

ÍNDICE GERAL

APRESENTAÇÃO 3

CAPÍTULO 1 - "MISTURA DE DISTRIBUIÇÕES I.I.D." - PRIMEIRA ABORDAGEM AOS MODELOS DE MUDANÇAS DE REGIME 6

I - INTRODUÇÃO 6
II - O MODELO UTILIZADO 6
III - DERIVAÇÃO DOS ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA 7
IV - RELAÇÃO ENTRE A INFERÊNCIA A POSTERIORI E OS ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA: O ALGORITMO EM 9
V - APLICAÇÃO DO MODELO: O CASO DO IBOVESPA 10
VI - CONSIDERAÇÕES FINAIS 14

CAPÍTULO 2 - MODELOS DE MUDANÇAS MARKOVIANAS DE REGIMES EM SÉRIES TEMPORAIS AUTORREGRESSIVAS 15

I - INTRODUÇÃO 15
II - O MODELO UTILIZADO 16
III - PROBABILIDADES FILTRADAS E PROBABILIDADES SUAVIZADAS 17
III.1 - Probabilidades Filtradas 18
III.1.1 - Notações e Definições Básicas 18
III.1.2 - Derivação do Algoritmo 19
III.2 - Probabilidades Suavizadas 20
III.2.1 - Derivação do Algoritmo 21
IV- ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA 23
V- ESTIMATIVAS DE MÁXIMA VEROSSIMILHANÇA OBTIDAS NUMERICAMENTE 29
VI- EXEMPLO 1: RETORNO DIÁRIO DOS TÍTULOS DE 30 ANOS DO TESOURO NORTE-AMERICANO 31
VII- EXEMPLO 2: SPREAD DIÁRIO ENTRE DOIS TÍTULOS DA DÍVIDA EXTERNA BRASILEIRA 41

CAPÍTULO 3 - MUDANÇAS DE REGIME EM VARIÂNCIAS CONDICIONAIS 52

I - INTRODUÇÃO 52
II - O MODELO GARCH (P,Q) 53
II.1 - Definições Básicas e Principais Resultados 53
II.2 - Persistência 54
III - OS MODELOS SWARCH 56
III.1 - O Modelo de Cai 56
III.2 - O Modelo de Hamilton 57
IV - OS MODELOS SWGARCH 58
V - MODELOS UTILIZADOS NESTE CAPÍTULO 60
VI - AJUSTE DOS MODELOS SWARCH E SWGARCH SOBRE TELEBRÁS PN 62

NOTAS CONCLUSIVAS 74

BIBLIOGRAFIA 76

ÍNDICE DE TABELAS

TABELA 1.1 - ESTIMATIVAS DOS MODELOS PARA O IBOVESPA 11
TABELA 2.1 - MODELOS ALTERNATIVOS PARA A SÉRIE DOS T-BONDS 30 ANOS 35
TABELA 2.2 - ERROS QUADRÁTICOS DE PROJEÇÃO DOS MODELOS PARA O T-BOND 30 ANOS 41
TABELA 2.3 - MODELOS ALTERNATIVOS PARA O SPREAD ENTRE OS TÍTULOS DA DÍVIDA EXTERNA BRASILEIRA 45
TABELA 2.4 - AJUSTE DOS MODELOS DE 2 REGIMES PARA O SPREAD ENTRE OS TÍTULOS DA DÍVIDA EXTERNA BRASILEIRA 47
TABELA 2.5 - CRITÉRIOS DE SELEÇÃO DE MODELOS PARA O SPREAD ENTRE OS TÍTULOS DA DÍVIDA EXTERNA BRASILEIRA 47
TABELA 2.6 - ERROS DE PROJEÇÃO A UM PASSO DOS MODELOS PARA "SPREAD" 50
TABELA 3.1 - MODELOS DE ALTERNÂNCIA DE REGIMES PROPOSTOS PARA OS RETORNOS DIÁRIOS DE TELEBRÁS. 63
TABELA 3.2 - RESTRIÇÕES AOS PARÂMETROS DOS MODELOS DE MUDANÇAS DE REGIMES PROPOSTOS 64
TABELA 3.3 - CRITÉRIOS DE SELEÇÃO PARA OS MODELOS ALTERNATIVOS 66
TABELA 3.4 - ESTIMATIVAS DOS PARÂMETROS PARA OS MODELOS SWARCH (1) ALTERNATIVOS 67
TABELA 3.5 - ESTIMATIVAS DOS PARÂMETROS PARA OS MODELOS SWARCH (2) ALTERNATIVOS 68
TABELA 3.6 - ESTIMATIVAS DOS PARÂMETROS PARA OS MODELOS SWGARCH (1,1) ALTERNATIVOS 69
TABELA 3.7 - TESTE LR PARA OS MODELOS SWGARCH (1,1) ALTERNATIVOS 69
TABELA 3.8 - COMPARATIVO ENTRE OS ERROS DE PROJEÇÃO DOS MODELOS PARA TELEBRÁS PN 73
TABELA 3.9 - COMPARATIVO ENTRE A PERSISTÊNCIA (δ) DOS MODELOS 73

ÍNDICE DE FIGURAS

FIG. 1.1 - DESVIO-PADRÃO CENTRADO - JANELA DE 21 DIAS ÚTEIS 10
FIG. 1.2 - IBOVESPA - RETORNOS DIÁRIOS 12
FIG. 1.3 - PROBABILIDADE DE OCORRÊNCIA DO REGIME 2 - ALTA VOLATILIDADE 12
FIG. 1.4 - DISTRIBUIÇÕES DE PROBABILIDADES 13
FIG. 1.5 - DISTRIBUIÇÕES DE PROBABILIDADES 13
FIG. 2.1 - SÉRIE DOS PREÇOS DE FECHAMENTO DIÁRIOS DOS T-BONDS 30 ANOS 32
FIG. 2.2. - RETORNOS DIÁRIOS - T-BONDS 30 ANOS 32
FIG. 2.3 - PROBABILIDADE SUAVIZADA DE OCORRÊNCIA DO REGIME DE ALTA 37
FIG. 2.4 - SÉRIE DE "SPREAD" DE FECHAMENTO DIÁRIO ENTRE O C-BOND E O PAR-BOND 42
FIG. 2.5 - SPREAD E PROBABILIDADE SUAVIZADA DE OCORRÊNCIA DO REGIME 1 48
FIG. 3.1 - RETORNOS DIÁRIOS DE TELEBRÁS PN 70
FIG. 3.2 - PROBABILIDADES SUAVIZADAS DOS DIFERENTES REGIMES 71

APRESENTAÇÃO

Durante muito tempo, permaneceu um consenso de que os processos estocásticos associados às séries financeiras (ações, taxa de câmbio, juros, etc.) poderiam ser modelados através de processos lineares e, neste sentido, quase sempre mediante algum modelo de passeio aleatório. Este enfoque permaneceu inalterado até o início da década de 70, com o advento da metodologia desenvolvida por Box & Jenkins (1970), pela qual ampliou-se o espectro dos modelos lineares. Porém, no caso dos retornos de séries financeiras, as funções de autocorrelação e autocorrelação parcial amostrais, salvo em raríssimos casos, não apresentavam evidências suficientes para se rejeitar a hipótese nula de passeio aleatório.

Entretanto, certas transformações aplicadas sobre os retornos (quadrado dos retornos, valor absoluto dos retornos ou o logaritmo do quadrado dos retornos) geravam funções de autocorrelação e autocorrelação parcial amostrais incompatíveis com a hipótese de independência, uma vez que apresentavam defasagens de baixa ordem significativas, sugerindo a inadequação dos modelos lineares gaussianos na descrição das séries financeiras.

A partir daí, por volta do início da década de 80, foi ganhando cada vez mais força a idéia de que modelos não lineares seriam os mais indicados para se explicar o comportamento de séries financeiras. Campbell, Lo & MacKinlay (1997) apresentam a classe dos modelos não lineares da seguinte forma:

De acordo com o Teorema da Decomposição de Wold, toda série temporal linear pode ser representada por uma média móvel linear, de ordem infinita, dos choques aleatórios. Em geral, nestes modelos, a seqüência dos choques é assumida como sendo não correlacionada, porém não necessariamente I.I.D.. Nos modelos não lineares, os choques geralmente são tidos como I.I.D., entretanto a relação funcional entre a série temporal e os choques aleatórios deixa de ser linear, ou seja, $y_t = f(u_t, u_{t-1}, u_{t-2}, \ldots)$ sendo que E $(u_t) = 0$ e Var $(u_t) = 1$ para todo t e $f(\cdot)$ é uma função desconhecida.

Apesar da generalização apresentada acima, a maioria dos modelos não lineares utilizados na prática recai numa subclasse que pode ser representada por

$$y_t = g(u_{t-1}, u_{t-2}, \ldots) + u_t \cdot h(u_{t-1}, u_{t-2}, \ldots).$$

Nesta caso, a função $g(\cdot)$ representa a esperança de y_t condicionada à informação passada. Ao mesmo tempo, $h(\cdot)^2$ representa a variância de y_t condicionada à informação passada. Modelos em que $g(\cdot)$ é uma função não linear são ditos modelos não lineares na média e modelos em que $h(\cdot)$ é uma função não linear são ditos modelos não lineares na variância. Assim, por exemplo, o modelo $y_t = u_t + \psi \cdot u_{t-1}^2$ é não linear somente na média uma vez que $g(\cdot) = \psi \cdot u_{t-1}^2$ e $h(\cdot) = 1$. Já o modelo ARCH (1), introduzido por Engle (1982), o qual

pode ser representado por $y_t = u_t\sqrt{\psi \cdot u_{t-1}^2}$, é não linear somente na variância pois $g(\cdot) = 0$ ao passo que $h(\cdot) = \sqrt{\psi \cdot u_{t-1}^2}$.

Dentre os modelos não lineares, algumas propostas interessantes têm surgido como, por exemplo, a abordagem hamiltoniana de mudanças de regime a qual supõe que a série temporal realiza mudanças de regime, sendo que estas mudanças são caracterizadas por uma alteração na distribuição de probabilidades e, ao mesmo tempo, governadas por alguma lei probabilística. Desta forma, baseando-se nas observações disponíveis, realizam-se inferências sobre a probabilidade a posteriori de se estar em cada um dos respectivos regimes, incorporando estas informações na equação de predição.

Um exemplo típico dos modelos de mudança markoviana de regimes é dado por:

$$y_t = \begin{cases} \varsigma_1 + \phi_1 \cdot y_{t-1} + u_{1t} & \text{se} \quad s_t = 1 \\ \varsigma_2 + \phi_2 \cdot y_{t-1} + u_{2t} & \text{se} \quad s_t = 2 \end{cases}$$

Neste modelo, em ambos os regimes y_t segue um AR(1), mas os parâmetros (inclusive a variância dos termos aleatórios) diferem de acordo com os regimes. Estes são determinados por uma variável aleatória não observada s_t, que segue uma Cadeia de Markov de dois estados, com matriz de transição $\mathbb{P}$.

Outro modelo que também incorpora a idéia de regimes alternantes é o chamado Limiar Autorregressivo (TAR), que pode ser caracterizado como:

$$y_t = \begin{cases} \varsigma_1 + \phi_1 \cdot y_{t-1} + u_t & \text{se} \quad y_{t-1} < \kappa \\ \varsigma_2 + \phi_2 \cdot y_{t-1} + u_t & \text{se} \quad y_{t-1} \geq \kappa \end{cases}$$

Neste caso, y_t também segue um AR(1) com parâmetros que se modificam de acordo com a relação entre y_{t-1} e o valor κ chamado de limiar. Este modelo pode ser estendido incorporando dinâmica autorregressiva de ordens superiores além de múltiplos limiares. A este respeito veja por exemplo Correia (1997).

A grande diferença entre estas duas abordagem é que, no modelo de mudança markoviana, as alterações entre os regimes não são governadas pelo nível do processo mas por uma variável aleatória não observada que segue uma Cadeia de Markov. Ou seja, as mudanças de regimes são causadas por outros fatores, externos à série temporal em si.

Além disto, os modelos de mudança markoviana de regimes oferecem uma importante vantagem sobre os modelos do tipo TAR, bem como sobre os demais modelos de quebras estruturais. As alterações de regimes são identificadas pela interação entre a série temporal e

a Cadeia de Markov, e não através de uma inspeção a priori sobre a série. Isto evita que os modelos de mudança de regimes incorram em viéses semelhantes ao que se verifica quando se utiliza os modelos de quebras estruturais na descrição de séries econômicas - raramente podemos dizer com absoluta certeza em que determinado instante uma série temporal econômica apresenta uma quebra estrutural (se é que esta última realmente existiu). Logo, a utilização deste modelos, assumindo existirem quebras estruturais onde, de fato, não existem, podem distorcer dramaticamente os resultados obtidos.

Por estas razões, nosso trabalho focaliza exclusivamente algumas das diferentes formas de se aplicar os modelos de mudanças markovianas de regimes sobre séries financeiras. Para tanto, este está divido em três capítulos. O capítulo 1 promove uma primeira abordagem sobre a aplicação dos modelos de mudança de regimes em séries financeiras, apresentaremos neste capítulo a sua versão mais simplificada, conhecida como "Mistura de Distribuições I. I. D." e que pode ser encontrada em Hamilton (1994). Além disto, será apresentada uma breve aplicação do modelo para a série do Índice Bovespa.

O capítulo 2 relaxa algumas hipóteses restritivas impostas no capítulo anterior, as quais impediam que o modelo de "Mistura de Distribuições I. I. D.", devido à sua extrema simplicidade, captasse algumas características adicionais peculiares a séries financeiras, tais como assimetria e movimentações em tendências. Neste segundo capítulo é introduzida a hipótese de que os regimes sejam governados por uma Cadeia de Markov, além de se permitir dinâmicas autorregressivas na série temporal. São tratadas também a questão da inferência sobre os regimes - probabilidades filtradas e suavizadas - além dos diferentes métodos de obtenção dos estimadores de máxima verossimilhança dos parâmetros. Para ilustrar a forma de se trabalhar com este modelo mais complexo, utilizaremos as séries dos títulos de 30 anos do tesouro dos EUA e o "spread" entre dois dos principais títulos da dívida externa brasileira, o C-bond e o Par-bond.

No capítulo 3, mostraremos como o modelo de mudanças markovianas de regimes pode ser também utilizado em modelos condicionalmente heterocedásticos. Estaremos comparando os resultados obtidos por estes modelos com os aqueles gerados pelo modelo GARCH (1,1), introduzido por Bollerslev (1986), quando aplicados sobre a série dos retornos diários de Telebrás PN.

Ao final do trabalho, breve notas conclusivas serão apresentadas.

CAPÍTULO 1 - "MISTURA DE DISTRIBUIÇÕES I.I.D." - PRIMEIRA ABORDAGEM AOS MODELOS DE MUDANÇAS DE REGIME

I - Introdução

Neste capítulo, apresentaremos a versão mais simplificada dos modelos de alternância de regimes, conhecida como "Mistura de Distribuições I. I. D." e que pode ser encontrada em Hamilton (1994).

Nas seções seguintes, descreveremos o modelo, derivaremos os estimadores de máxima verossimilhança de seus parâmetros e traçaremos suas relações com a inferência *a posteriori* a cada instante sobre os diferentes regimes. Além disto, será apresentada uma breve aplicação do modelo para a série do Índice Bovespa.

II - O Modelo Utilizado

Seja $\{X_t\}$ uma série temporal financeira qualquer (índice de ações, cotações no mercado futuro de algum ativo financeiro, taxa de câmbio, etc.).

Definimos por $y_t = ln\, x_t - ln\, x_{t-1}$ o retorno em regime de capitalização contínua no instante t.

Seja $S_t \in \{1,2,3,\ldots,k\}$ a v.a. que indica qual regime está vigorando no instante t

São pressupostos básicos do modelo de "Mistura de Distribuições I. I. D." a independência entre os retornos e a distribuição destes condicionada aos regimes depende apenas daquele que está atualmente em vigor, ou seja:

$$f\left(y_t \middle| y_{t-1}, y_{t-2}, \ldots, y_1, s_t, s_{t-1}, s_{t-2}, \ldots, s_1\right) = f\left(y_t \middle| s_t\right)$$

Assumindo então normalidade na distribuição de probabilidade dos retornos temos que:

$$f\left(y_t \middle| s_t = j\right) = \frac{1}{\sqrt{2\pi}\sigma_j} exp\left\{-\frac{\left(y_t - \mu_j\right)^2}{2\sigma_j^2}\right\} \quad (1.II.1)$$

Além disto a distribuição de probabilidades não condicional de nos encontrarmos nos regimes 1, 2, 3,...,k é dada por:

$$P\{s_t = j\} = \pi_j, j \in \{1,2,3,\ldots,k\}, \sum_{j=1}^{k} \pi_j = 1$$

Ou seja, o processo estocástico $\{S_t\}$ forma uma seqüência de variáveis aleatórias i. i. d..

Utilizando-se do Teorema da Partição do Espaço Amostral temos que a distribuição não condicional de y_t possui a seguinte densidade:

$$f(y_t) = \sum_{j=1}^{k} \pi_j \times f(y_t | s_t = j) \quad (1.II.2)$$

Finalmente, utilizando o Teorema de Bayes, chegamos à expressão da probabilidade à posteriori de S_t:

$$P\{s_t = j | y_t\} = \frac{P\{s_t = j; y_t\}}{f(y_t)} = \frac{f(y_t | s_t = j) \times \pi_j}{f(y_t)} \quad (1.II.3)$$

Vemos então que, para calcularmos as probabilidades *a posteriori* em cada instante t, necessitamos de estimativas para os parâmetros μ_j, σ_j^2 e π_j.

III - Derivação dos Estimadores de Máxima Verossimilhança

Interessa-nos agora obter os estimadores de máxima verossimilhança de μ_j, σ_j^2 e π_j. Seja θ o vetor que reúne os parâmetros do modelo. Pelo que foi exposto na seção anterior, a expressão da log-verossimilhança do modelo em questão é dada por:

$\ell(\theta) = \sum_{t=1}^{T} lnf(y_t;\theta)$, onde $f(y_t;\theta)$ é dado pela expressão (1.II.2)

Temos então que, para encontrarmos as expressões dos estimadores de máxima verossimilhança de θ, devemos maximizar $\ell(\theta)$ sujeito à restrição $\sum_{j=1}^{k} \pi_j = 1$. Assim, construímos o seguinte lagrangeano:

$$J(\theta) = \ell(\theta) + \lambda \times \left(1 - \sum_{j=1}^{k} \pi_j\right)$$

Note ainda que :

$$\frac{\partial lnf(y_t;\theta)}{\partial\theta} = \frac{1}{f(y_t;\theta)} \times \frac{\partial f(y_t;\theta)}{\partial\theta}$$

Desta forma, derivando-se $\ell(\theta)$ em relação a θ, obtemos:

$$\frac{\partial\ell(\theta)}{\partial\theta} = \sum_{t=1}^{T} \frac{1}{f(y_t;\theta)} \times \frac{\partial f(y_t;\theta)}{\partial\theta}$$

Como $\theta = \{\mu_1,\ldots,\mu_k,\sigma_1,\ldots,\sigma_k,\pi_1,\ldots,\pi_k\}$, e assumindo normalidade nas distribuições condicionais de y_t, a expressão acima se desdobra nas seguintes derivadas parciais:

$$\frac{\partial\ell(\theta)}{\partial\pi_j} = \sum_{t=1}^{T} \frac{1}{f(y_t;\theta)} \times f(y_t|s_t = j;\theta) = \pi_j^{-1}\sum_{t=1}^{T} P\{s_t = j|y_t;\theta\}$$

$$\frac{\partial\ell(\theta)}{\partial\mu_j} = \sum_{t=1}^{T} \frac{1}{f(y_t;\theta)} \times \frac{y_t-\mu_j}{\sigma_j^2} f(y_t,s_t = j;\theta) = \sum_{t=1}^{T} \frac{y_t-\mu_j}{\sigma_j^2} P\{s_t = j|y_t;\theta\}$$

$$\frac{\partial\ell(\theta)}{\partial\sigma_j^2} = \sum_{t=1}^{T} \frac{1}{f(y_t;\theta)}\left\{-\frac{1}{2\sigma_j^2} + \frac{(y_t-\mu_j)^2}{2\sigma_j^4}\right\} f(y_t,s_t = j;\theta) = \sum_{t=1}^{T}\left\{-\frac{1}{2\sigma_j^2} + \frac{(y_t-\mu_j)^2}{2\sigma_j^4}\right\} P\{s_t = j|y_t;\theta\}$$

Assim, obtemos:

$$\frac{\partial J(\theta)}{\partial\pi_j} = \pi_j^{-1}\sum_{t=1}^{T} P\{s_t = j|y_t;\theta\} - \lambda$$

$$\frac{\partial J(\theta)}{\partial\mu_j} = \sum_{t=1}^{T} \frac{y_t-\mu_j}{\sigma_j^2} P\{s_t = j|y_t;\theta\}$$

$$\frac{\partial J(\theta)}{\partial\sigma_j^2} = \sum_{t=1}^{T}\left\{-\frac{1}{2\sigma_j^2} + \frac{(y_t-\mu_j)^2}{2\sigma_j^4}\right\} P\{s_t = j|y_t;\theta\}$$

Igualando as derivadas parciais a zero, obtemos:

$$\hat{\mu}_j^{(\ell+1)} = \frac{\sum_{t=1}^{T} y_t \times P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\}}$$

$$\hat{\sigma}_j^{2;(\ell+1)} = \frac{\sum_{t=1}^{T} \left(y_t - \hat{\mu}_j^{(\ell+1)}\right)^2 \times P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\}}$$

$$\hat{\pi}_j^{(\ell+1)} = \frac{\sum_{t=1}^{T} P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\}}{T}$$

Onde T é o número total de observações e $\hat{\theta}^{(\ell)}$ é vetor das estimativas dos parâmetros obtido proveniente da ℓ - ésima iteração.

A última expressão é obtida verificando-se que:

$$\frac{\partial J(\theta)}{\partial \pi_j} = 0 \Rightarrow \sum_{t=1}^{T} P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\} = \hat{\lambda}\hat{\pi}_j^{(\ell+1)}$$

e, somando-se os dois membros da expressão acima em relação a j, obtemos:

$$\sum_{j=1}^{k}\sum_{t=1}^{T} P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\} = \hat{\lambda}\sum_{j=1}^{k}\hat{\pi}_j^{(\ell+1)} \Rightarrow \sum_{t=1}^{T}\sum_{j=1}^{k} P\left\{s_t = j \middle| y_t; \hat{\theta}^{(\ell)}\right\} = \hat{\lambda} \Rightarrow \hat{\lambda} = T$$

IV - Relação entre a Inferência a posteriori e os Estimadores de Máxima Verossimilhança: O Algoritmo EM

Como podemos perceber, as expressões dos estimadores de máxima verossimilhança dos parâmetros não são funções lineares das observações, logo, não podem ser resolvidos analiticamente. Neste caso utilizamos um algoritmo chamado de EM, o qual consiste em:

1) Atribuir estimativas iniciais para $\hat{\mu}_j, \hat{\sigma}_j^2$ e $\hat{\pi}_j$

2) Com base nestes valores iniciais, calcular através das expressões (1.II.1) e (1.II.2) as probabilidades *a posteriori* dadas por (1.II.3)

3) Tendo os valores das probabilidades a posteriori, recalcular através dos estimadores de máxima verossimilhança. novos valores para $\hat{\mu}_j, \hat{\sigma}_j^2$ e $\hat{\pi}_j$

4) Repetir o processo até que os valores dos estimadores não se alterem significativamente segundo algum critério de convergência.

V - Aplicação do Modelo: O Caso do Ibovespa

A fim de exemplificarmos a utilização do modelo acima descrito, escolhemos a série histórica das cotações de fechamento diário do Ibovespa à vista, correspondente ao período de 04/01/1993 a 23/07/96 (T = 873 observações).

Como foi apresentado, devemos obter as estimativas iniciais dos parâmetros a fim de se iniciar o algoritmo de estimação. Calculamos, então, através da série dos retornos diários, uma nova série com os desvios padrões centrados utilizando-se 21 observações - janela de 21 dias úteis (Fig 1.1). Podemos perceber, de antemão, períodos de altas e baixas volatilidades. Consideramos a presença de dois regimes (regime 1 - baixa volatilidade e regime 2 - alta volatilidade) e obtivemos nossas estimativas iniciais para os parâmetros que caracterizam a distribuição de probabilidades de ambos os regimes, calculando-se a média e o desvio padrão dos retornos conforme o desvio padrão centrado do instante t esteja acima ou abaixo do desvio padrão da série global dos retornos, representado pela linha vermelha na Fig. 1.1.

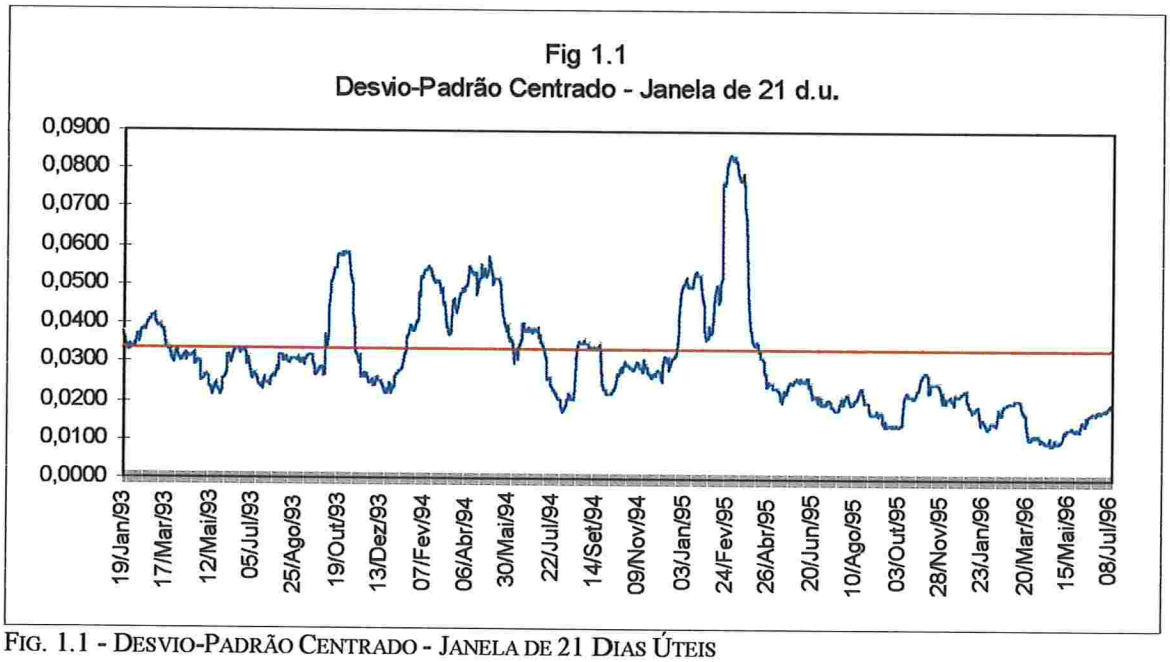

FIG. 1.1 - DESVIO-PADRÃO CENTRADO - JANELA DE 21 DIAS ÚTEIS

Seguem abaixo, as estimativas assim obtidos :

$\hat{\mu}_1 = 0.00250 \quad \hat{\sigma}_1^2 = 0.00071 \quad \hat{\pi}_1 = 0.51348 \quad \hat{\mu}_2 = 0.00067 \quad \hat{\sigma}_2^2 = 0.00159$

Note que, mesmo se tratando de estimativas iniciais, a variância do regime 2 é maior que o dobro da verificada no regime 1.

A tabela 1.1. resume o resultados obtidos aplicando-se o algoritmo EM. No caso dos modelos de mudanças de regimes consideramos também a hipótese $\mu_1 = \mu_2 = \mu$.

Tabela 1.1 - Estimativas dos Modelos para o Ibovespa

PARÂMETRO	HAMILTON 1 ($\mu_1 \neq \mu_2$)	HAMILTON 2 ($\mu_1 = \mu_2 = \mu$)	RANDOM WALK
μ	-	0.00177 (0.00104)	0.00162 (0.00114)
μ_1	0.00202 (0.00130)	-	-
μ_2	0.00024 (0.00519)	-	-
σ_1^2	0.00063 (0.00007)	0.00062 (0.00007)	-
σ_2^2	0.00288 (0.00040)	0.00285 (0.00037)	-
σ^2	-	-	0.00114 (0.00005)
π_1	0.77537 (0.06301)	0.77066 (0.06093)	-
Log Verossimilhança	1752.81	1752.76	1721.45
Teste LR [valor-p]	62.7 [0.000]	62.5 [0.000]	-
AIC (*)	-3495.6	-3497.4	-3438.9
SC (**)	-3471.8	-3478.3	-3429.4

(*) $AIC = -2 \cdot \ell(\hat{\theta}) + 2 \cdot r$, onde r é o número de parâmetros a serem estimados.

(**) $SC = -2 \cdot \ell(\hat{\theta}) + r \cdot \ln T$

Considerando-se os critérios de Akaike (AIC) bem como o de Schwarz (SC) , constatamos a superioridade dos modelos de mudanças de regime sobre aquele que seria obtido utilizando-se as técnicas tradicionais de séries temporais (vale mencionar que o modelo de passeio aleatório seria o mais adequado uma vez que as funções amostrais de autocorrelação e autocorrelação parcial apontam nesta direção). Outra confirmação provém do fato de que o teste da razão de verossimilhança rejeita a hipótese nula de passeio aleatório em favor dos modelos de mudanças de regimes. Além disto, parece haver evidências que a caracterização básica entre os dois regimes é mesmo a mudança na volatilidade uma vez que o melhor

ajuste foi obtido quando consideramos a hipótese de igualdade entre as médias dos retornos diários de ambos os regimes.

Para completar este pequeno exemplo, apresentamos a seguir as Fig. 1.2 e Fig. 1.3 referentes aos retornos diários do Ibovespa em regime de capitalização contínua (Fig. 1.2) e às probabilidades condicionais de nos encontrarmos no regime 2 - alta volatilidade (Fig. 1.3).

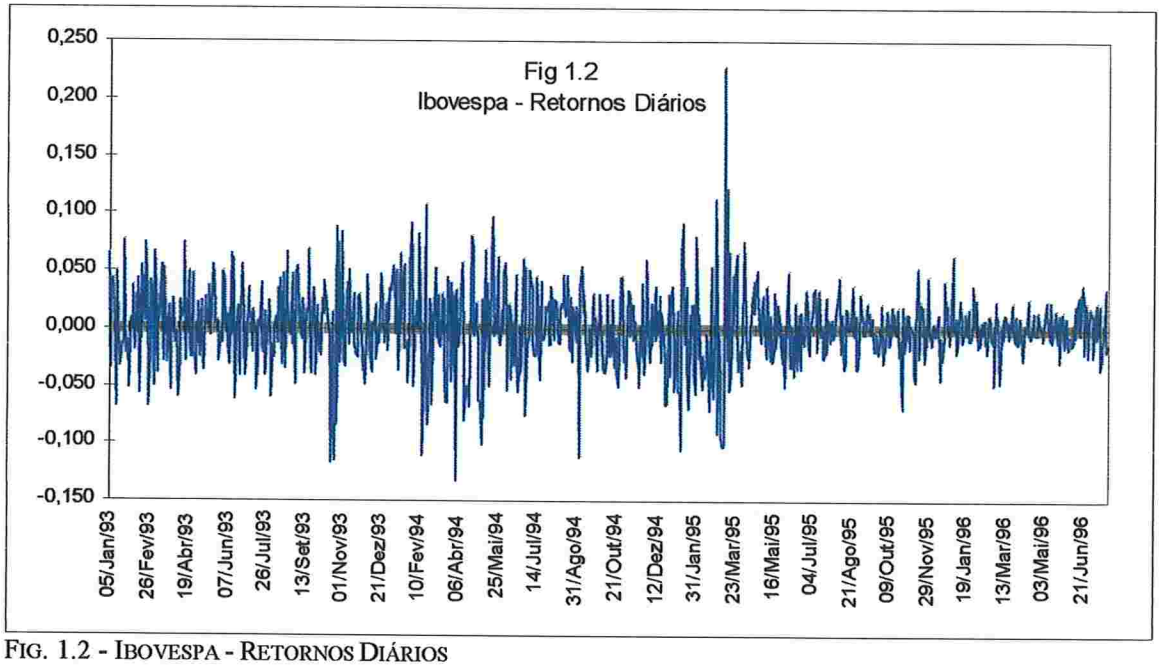

FIG. 1.2 - IBOVESPA - RETORNOS DIÁRIOS

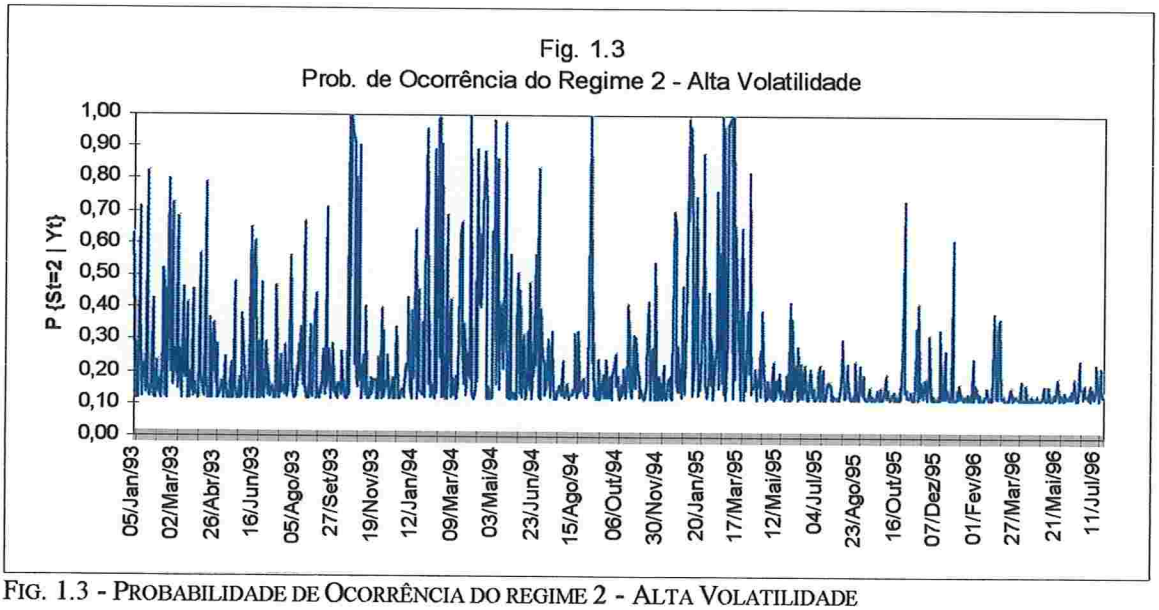

FIG. 1.3 - PROBABILIDADE DE OCORRÊNCIA DO REGIME 2 - ALTA VOLATILIDADE

De acordo com as figuras 1.2 e 1.3, observamos que em várias ocasiões, o modelo ajustado identifica períodos de volatilidade elevada (probabilidade de nos encontrarmos no regime 2

maior que 50%) sendo que em algumas delas, esta probabilidade chegou a concentrar valores superiores a 90%: outubro de 93, fevereiro a maio de 94 (implantação da URV) e dezembro a março de 95 (crise mexicana e mudança na política cambial brasileira).

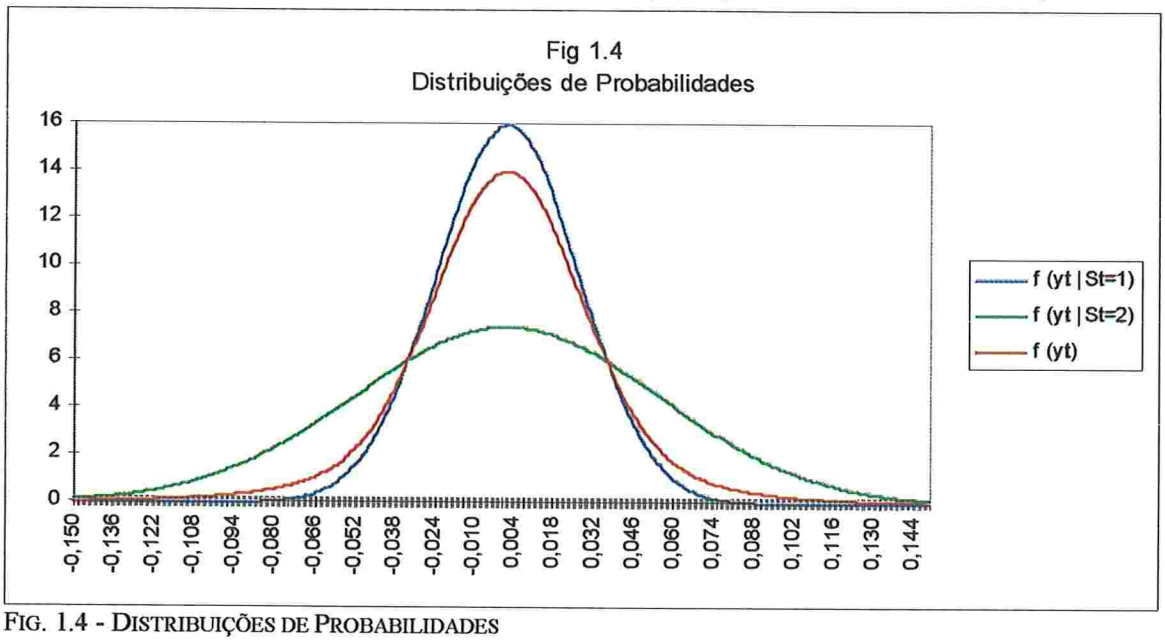

FIG. 1.4 - DISTRIBUIÇÕES DE PROBABILIDADES

Por último, apresentamos também a Fig. 1.4 contendo as estimativas das distribuições de probabilidades dos retornos diários do Ibovespa condicionados aos diferentes regimes bem como a distribuição não-condicional dos mesmos (linha vermelha) obtida através das estimativas do modelo adotado. Note que, conforme exibe a Fig. 1.5, esta última possui caudas mais pesadas do que a distribuição normal que seria obtida caso o modelo ajustado fosse o de passeio aleatório, sugerindo maior adequabilidade de modelos não-lineares.

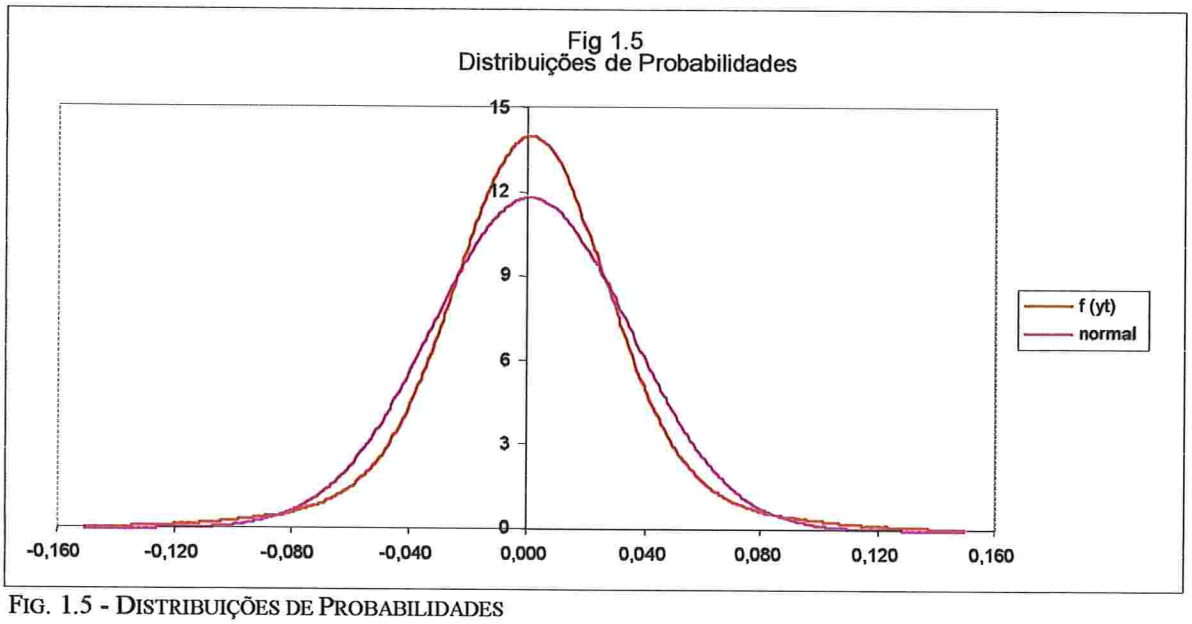

FIG. 1.5 - DISTRIBUIÇÕES DE PROBABILIDADES

VI - Considerações Adicionais

O modelo "Mistura de Distribuições I. I. D." acima descrito é extremamente simples, porém têm o mérito de introduzir algumas questões importantes sobre os modelos de mudanças de regimes tais como:

1) A sua não-linearidade;

2) A (in)dependência temporal entre os regimes;

3) As relações entre as probabilidades *a posteriori* de nos encontrarmos em cada instante num determinado regime e as expressões dos estimadores de máxima verossimilhança, que são a base do algoritmo EM de estimação dos parâmetros do modelo.

Estas características estarão presentes nos modelos mais sofisticados que passaremos a apresentar a partir do próximo capítulo, os quais são obtidos relaxando-se as hipóteses de independência entre as observações e/ou atribuindo-se uma estrutura markoviana para caracterizar o processo estocástico determinante dos diferentes regimes.

CAPÍTULO 2 - MODELOS DE MUDANÇAS MARKOVIANAS DE REGIMES EM SÉRIES TEMPORAIS AUTORREGRESSIVAS

I - Introdução

No capítulo anterior apresentamos a versão mais simplificada dos modelos de mudanças de regimes conhecida como "Mistura de Distribuições I. I. D." . Entretanto, a sua extrema simplicidade não permite captar algumas características adicionais das séries temporais financeiras, porém não menos importantes, tais como:

1) Assimetria: muitas variáveis econômico-financeiras se comportam de maneira assimétrica perante os ciclos econômicos. Algumas delas, como as ações por exemplo, quando sobem costumam fazê-lo de maneira longa e gradual e, quando caem, geralmente realizam movimentos abruptos e curtos. Neste caso, poderíamos esperar que tanto a média como a variância que caracterizariam o regime 1 (regime de alta) fossem, em valor absoluto, menores do que as do regime 2 (regime de baixa). Além disto, podemos ter também tempos médios de permanência diferentes em cada um dos regimes.

2) Movimentos em tendências: Há entre os operadores de Wall Street um ditado muito popular que diz: "The trend is your friend". No fundo, o que isto quer dizer é que as variáveis financeiras se movimentam em tendências ao longo do tempo as quais tendem a se perpetuar até que condições econômicas, políticas, sociais, etc. produzam fatores que acabam, de maneira quase que abrupta, promovendo reversões nestas tendências.

Assim, ao contrário do que é pressuposto pelo modelo apresentado no capítulo 1, a hipótese de independência entre os regimes parece não ser adequada. Ou seja, quando nos encontramos dentro de um particular regime no instante t, há maiores chances de nos encontrarmos neste mesmo regime no instante $t+1$, do que saltarmos para um regime diferente. Além disto, é comum perceber que o tempo de permanência dentro dos regimes geralmente não é o mesmo.

Diante destes fatos, é mais sensato supor que os regimes sejam governados por uma Cadeia de Markov, uma vez que esta estrutura markoviana permite o relaxamento da hipótese de independência e, ao mesmo tempo, consegue abrir espaço para tempos de permanência diferentes dentro dos regimes, já que as probabilidades de transição podem assumir valores dos mais variados.

Por outro lado, é comum no mercado financeiro a elaboração de análises de "spreads" (diferença entre preços de dois ativos) e, geralmente, séries temporais desta natureza

possuem algum tipo de estrutura de reversão à média, que pode ser caracterizada por um modelo autorregressivo de ordem m. Em geral, consideramos $m = 1$.

Surge então, a necessidade de se desenvolver um modelo mais sofisticado de mudanças de regimes, quer dizer, que possua estrutura markoviana para os diferentes regimes em séries temporais autorregressivas.

É claro que teremos agora um modelo mais completo, porém a tarefa de se encontrar expressões para os estimadores de máxima verossimilhança dos parâmetros torna-se bastante ingrata. Além disto, a inferência probabilística sobre os regimes não depende mais apenas da observação do instante t.

Iremos, assim, neste capítulo tratar passo-a-passo cada uma destas questões, apresentando o modelo de mudanças markovianas de regimes em séries autorregressivas, a forma de se obter os estimadores de máxima verossimilhança dos parâmetros, a questão da inferência sobre os regimes, a equação de predição e, por fim, apresentaremos duas aplicações deste modelo em questão, sendo uma delas relacionadas à análise de "spread".

II - O Modelo Utilizado

Seja $\{Y_t\}$ um processo estocástico assumindo valores no conjunto $\Re$.

Supomos que ocorrem ocasionalmente mudanças discretas no nível, na variância e na dinâmica autorregressiva do processo estocástico, sendo que $S_t \in \{1,2,3,\ldots,k\}$ é a v.a. que indica qual regime está vigorando no instante t. Ou seja, existem k possíveis diferentes regimes pelos quais uma particular observação de $\{Y_t\}$ é governada.

Assumimos também que existe uma dinâmica autorregressiva, no máximo de primeira ordem, para o processo estocástico $\{Y_t\}$ tal que y_t depende apenas dos valores imediatamente mais recentes de s_t, y_t e de um vetor de parâmetros θ:

$$f\left(y_t \middle| y_{t-1}, y_{t-2}, \ldots, y_1, s_t, s_{t-1}, s_{t-2}, \ldots, s_1; \theta\right) = f\left(y_t \middle| s_t, s_{t-1}, y_{t-1}; \theta\right)$$

Além disto, assumimos que a transição entre os regimes 1, 2, 3,...,k é dada por uma Cadeia de Markov de primeira ordem:

$$P\left\{s_t = j \middle| s_{t-1} = i\right\} = p_{ij}; i, j \in \{1,2,3,\ldots,k\}, \sum_{j=1}^{k} p_{ij} = 1$$

Chamaremos também de **P** a Matriz de Transição (k $\times$ k) desta Cadeia de Markov, cuja soma dos elementos de uma determinada coluna j, por definição, somam 1.

Por hipótese, assumimos também que, condicionado a s_{t-1}, s_t é independente com relação a valores passados de s_t.

Do que foi exposto acima, podemos exprimir o nosso modelo através da seguinte equação:

$y_t = \mu_{s_t} + \phi_{s_t} \times y_{t-1} + u_{t,s_t}$; sendo que u_{t,s_t} tem distribuição normal com média 0 e variância $\sigma^2_{s_t}$.

Nestas condições, nosso vetor de parâmetros a ser estimado é dado por:

$$\theta = \{\mu_1, \mu_2, \ldots, \mu_k, \phi_1, \phi_2, \ldots, \phi_k, \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2, p_{11}, p_{12}, \ldots p_{1k}, \ldots, p_{k1}, p_{k2}, \ldots, p_{kk}\}$$

Note que não fazem parte do vetor de parâmetros a distribuição de probabilidades inicial da Cadeia de Markov. Por questão de simplicidade, assumimos como distribuição inicial a medida invariante da Cadeia de Markov.

Antes de tratarmos da estimação dos parâmetros do modelo acima descrito, convém abordarmos a questão da inferência a posteriori sobre os diferentes regimes, uma vez que esta se torna mais complexa, ao mesmo tempo em que é fundamental para a estimação dos parâmetros, conforme veremos a seguir.

III - Probabilidades Filtradas e Probabilidades Suavizadas

Como vimos no capítulo anterior, a inferência a posteriori sobre os regimes, isto é, a probabilidade de nos encontrarmos no instante t em um determinado regime dado que já conhecíamos a observação relativa ao instante t da série temporal dependia somente desta única observação. Isto era fruto da hipótese de independência entre os regimes juntamente com a ausência de uma estrutura autorregressiva no processo estocástico.

Entretanto, o relaxamento destas hipóteses faz com que esta inferência a posteriori não dependa apenas da observação no instante t. Outrossim, surge agora um desdobramento na forma de se calcular a probabilidade da ocorrência dos diferentes regimes dado o conjunto de observações da nossa série temporal. Uma delas, denominada "Probabilidade Filtrada" refere-se a inferência a posteriori sobre o regime no instante t levando-se em consideração toda a informação disponível sobre a série temporal do instante 1 até o instante t. A outra, conhecida como "Probabilidade Suavizada" utilizamos toda a série temporal no cálculo das probabilidades a posteriori sobre o regime no instante t.

Em resumo, temos:

$\mathbf{Y}_t = \{y_t, y_{t-1}, \ldots, y_1\}$ é o conjunto das observações sobre a série temporal até o instante t.

$\mathbf{Y}_T = \{y_T, y_{T-1}, \ldots, y_1\}$ é o conjunto de todas as observações disponíveis da série temporal, ou seja, até o instante T.

$P\{s_t = j|\mathbf{Y}_t\}$, $j \in \{1,2,3,\ldots,k\}$ é a "Probabilidade Filtrada" de nos encontrarmos no instante t em um determinado regime j, inferida utilizando-se o conjunto de informação até o instante t.

$P\{s_t = j|\mathbf{Y}_T\}$, $j \in \{1,2,3,\ldots,k\}$ é a "Probabilidade Suavizada" de nos encontrarmos no instante t em um determinado regime j, inferida utilizando-se o todo conjunto de informação.

Note que apenas quando t = T as duas probabilidades acima coincidirão, quer dizer, serão as mesmas quando realizarmos a inferência no instante em que a informação mais recente estiver disponível.

A seguir, apresentaremos os algoritmos utilizados no cálculo destas probabilidades, uma vez que elas não são mais obtidas através de uma forma direta conforme foi apresentado no capítulo 1, quando o modelo de "Mistura de Distribuições I. I. D." foi abordado.

III.1 - Probabilidades Filtradas

III.1.1 - Notações e Definições Básicas

Inicialmente, vale ressaltar que o desenvolvimento dos algoritmos necessários ao cálculo das probabilidades filtradas e suavizadas pressupõe o conhecimento de θ. Entretanto, mesmo conhecendo θ, nós nunca saberemos com certeza em que regime o processo esteve em cada um dos instantes ao longo da amostra. Alternativamente, o melhor que podemos fazer é realizar uma afirmação probabilística sobre os regimes a cada instante de tempo e, a primeira forma delas é a que já denominamos de "Probabilidades Filtradas".

Seja $P\{s_t = j|\mathbf{Y}_t;\theta\}$ a afirmação probabilística a cerca do valor de s_t baseada nas observações até o instante t e no conhecimento do vetor paramétrico populacional θ. Esta afirmação toma a forma de uma probabilidade condicional que é atribuída à possibilidade que a t-ésima observação tenha sido governada pelo regime j. Agrupamos estas probabilidades, $P\{s_t = j|\mathbf{Y}_t;\theta\}$ $j \in \{1,2,3,\ldots,k\}$ num vetor (k $\times$ 1) denotado por $\zeta_{t|t}$.

Também definimos outro vetor (k $\times$ 1) η_t, o qual agrupa as densidades condicionais de y_t dados o regime que está vigorando em t, o conjunto de informações passadas $\mathbf{Y}_{t-1}$ e o vetor

de parâmetros θ, ou seja cada uma destas densidades condicionais é dada por $f(y_t|s_t = j, \mathbf{Y}_{t-1};\theta)$.

Vale lembrar que, por hipótese, a densidade condicional acima depende apenas do regime no instante t, mas não dos regimes anteriores:

$$f(y_t|s_t = j, \mathbf{Y}_{t-1};\theta) = f(y_t|s_t = j, s_{t-1} = i, \ldots, \mathbf{Y}_{t-1};\theta).$$

Por outro lado, podemos construir um outro vetor $(k \times 1)$ denotado por $\zeta_{t+1|t}$, o qual representa a previsão do regime j no instante $t+1$ dada a informação disponível até o instante t. Assim, cada elemento deste vetor é $P\{s_{t+1} = j|\mathbf{Y}_t;\theta\}$.

III.1.2 - Derivação do Algoritmo

Como vimos acima, o j-ésimo elemento de $\zeta_{t|t-1}$ é dado por $P\{s_t = j|\mathbf{Y}_{t-1};\theta\}$. Por outro lado, o j-ésimo elemento de η_t é $f(y_t|s_t = j, \mathbf{Y}_{t-1};\theta)$. Assim, o j-ésimo elemento do vetor $(\zeta_{t|t-1} \odot \eta_t)$ é o produto destas duas magnitudes acima descritas, o qual pode ser interpretado como a densidade conjunta de y_t e s_t:

$$f(y_t, s_t = j|\mathbf{Y}_{t-1};\theta) = P\{s_t = j|\mathbf{Y}_{t-1};\theta\} \times f(y_t|s_t = j, \mathbf{Y}_{t-1};\theta)$$

Logo, a densidade de y_t condicionada apenas ao conjunto das observações passadas $\mathbf{Y}_{t-1}$ é a soma para j=1, 2,...,k, das k quantidades acima descritas:

$f(y_t|\mathbf{Y}_{t-1};\theta) = \sum_{j=1}^{k} f(y_t, s_t = j|\mathbf{Y}_{t-1};\theta)$, e que pode ser escrito em notação matricial como:

$$f(y_t|\mathbf{Y}_{t-1};\theta) = \mathbf{1}'(\zeta_{t|t-1} \odot \eta_t)$$

Se dividirmos a densidade conjunta de y_t e s_t pela densidade de y_t, obviamente obteremos a densidade condicionada de s_t a $\mathbf{Y}_t$:

$$P\{s_t = j|\mathbf{Y}_t;\theta\} = P\{s_t = j|y_t; \mathbf{Y}_{t-1};\theta\} = \frac{f(y_t, s_t = j|\mathbf{Y}_{t-1};\theta)}{f(y_t|\mathbf{Y}_{t-1};\theta)}$$

Usando a notação apresentada nesta seção e, utilizando-se a notação matricial, conseguimos reescrever a expressão acima:

$$\zeta_{t|t} = \frac{(\zeta_{t|t-1} \odot \eta_t)}{\mathbf{1}'(\zeta_{t|t-1} \odot \eta_t)} \quad (2.\text{III}.1)$$

Adicionalmente, utilizando-se teorias elementares de Cadeias de Markov, é fácil notar que:

$$\zeta_{t+1|t} = \mathbf{P} \cdot \zeta_{t|t} \qquad (2.\text{III}.2)$$

As expressões (2.III.1) e (2.III.2) proporcionam o algoritmo necessário ao cálculo das "probabilidades filtradas" para cada instante t da amostra: dadas uma estimativa para o vetor de parâmetros $\hat{\theta}$ (o qual inclui as probabilidades de transição), a distribuição de probabilidades inicial $\hat{\zeta}_{1|0}$ e, uma vez realizada a observação no instante 1, através de (2.III.1) obtemos $\hat{\zeta}_{1|1}$. A seguir, utilizando-se (2.III.2) obtemos $\hat{\zeta}_{2|1}$. Desta forma, o processo é repetido até atingirmos o instante T (última observação da amostra).

É importante salientar que sempre utilizaremos em nossos modelos, como distribuição inicial das probabilidades $\hat{\zeta}_{1|0}$ a distribuição invariante da Cadeia de Markov caracterizada pela Matriz de Transição **P**.

III.2 - Probabilidades Suavizadas

A grande diferença entre as probabilidades filtradas e as probabilidades suavizadas é, como já vimos, o fato de que as últimas são calculadas considerando-se toda a informação disponível desde o instante t = 1. Como veremos a seguir, outra diferença básica está associada ao sentido temporal do algoritmo de cálculo. Enquanto que as probabilidades filtradas são obtidas partindo-se do instante t = 1 até o final da série, as probabilidades suavizadas são obtidas de maneira inversa, ou seja, obtemos inicialmente a probabilidade suavizada do instante t = T e, a partir desta, regride-se no tempo até obtermos a probabilidade suavizada do instante t = 1.

É importante observar que, como conseqüência automática das definições apresentadas na seção III, a probabilidade suavizada do instante t = T é a própria probabilidade filtrada deste instante. Com efeito:

Probabilidade Suavizada do instante T = $P\{s_T = j|\mathbf{Y}_T\}$, por definição.

Probabilidade Filtrada do instante t = $P\{s_t = j|\mathbf{Y}_t\}$. Fazendo t = T, obtemos a magnitude acima.

Daí o porquê do algoritmo de cálculo das probabilidades suavizadas iniciar-se a partir do final da amostra.

III.2.1 - Derivação do Algoritmo

Inicialmente, vale lembrar que, sob as hipóteses do modelo, o regime s_t depende do conjunto das observações passadas $\mathbf{Y}_{t-1}$ apenas através de s_{t-1}. Da mesma forma, s_t depende das observações futuras apenas através de s_{t+1}:

$$P\{s_t = j|s_{t+1} = i, \mathbf{Y}_{t+1};\theta\} = P\{s_t = j|s_{t+1} = i, y_{t+1}, \mathbf{Y}_t;\theta\} = \frac{f(y_{t+1}, s_t = j|s_{t+1} = i, \mathbf{Y}_t;\theta)}{f(y_{t+1}|s_{t+1} = i, \mathbf{Y}_t;\theta)} =$$

$$= \frac{f(y_{t+1}|s_t = j, s_{t+1} = i, \mathbf{Y}_t;\theta) \times P\{s_t = j|s_{t+1} = i, \mathbf{Y}_t;\theta\}}{f(y_{t+1}|s_{t+1} = i, \mathbf{Y}_t;\theta)}$$

porém como por hipótese $f(y_{t+1}|s_t = j, s_{t+1} = i, \mathbf{Y}_t;\theta) = f(y_{t+1}|s_{t+1} = i, \mathbf{Y}_t;\theta)$, temos que:

$$P\{s_t = j|s_{t+1} = i, \mathbf{Y}_{t+1};\theta\} = P\{s_t = j|s_{t+1} = i, \mathbf{Y}_t;\theta\}.$$

Prosseguindo por indução, pode-se demonstrar que:

$$P\{s_t = j|s_{t+1} = i, \mathbf{Y}_{t+m};\theta\} = P\{s_t = j|s_{t+1} = i, \mathbf{Y}_t;\theta\}, \text{ para } m = 1,2,\ldots,T\text{-}t.$$

De fato, assumindo por hipótese a igualdade acima e tendo em vista que para m =1 já demonstramos, temos que:

$$P\{s_t = j|s_{t+1} = i, \mathbf{Y}_{t+m+1};\theta\} = P\{s_t = j|s_{t+1} = i, y_{t+m+1}, \mathbf{Y}_{t+m};\theta\}$$

$$= \frac{f(y_{t+m+1}, s_t = j|s_{t+1} = i, \mathbf{Y}_{t+m};\theta)}{f(y_{t+m+1}|s_{t+1} = i, \mathbf{Y}_{t+m};\theta)} =$$

$$= \frac{f(y_{t+m+1}|s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m};\theta) \times P\{s_t = j|s_{t+1} = i, \mathbf{Y}_{t+m};\theta\}}{f(y_{t+m+1}|s_{t+1} = i, \mathbf{Y}_{t+m};\theta)}$$

Note agora que:

$$f(y_{t+m+1}|s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m};\theta) = \sum_{l=1}^{k} f(y_{t+m+1}, s_{t+m+1} = l|s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m};\theta) =$$

$$= \sum_{l=1}^{k} f\left(y_{t+m+1} \middle| s_{t+m+1} = l, s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right) \times P\left\{s_{t+m+1} = l \middle| s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right\} =$$

$$= \sum_{l=1}^{k} f\left(y_{t+m+1} \middle| s_{t+m+1} = l, s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right) \times P\left\{s_{t+m+1} = l \middle| s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right\},$$ onde esta última igualdade segue do fato de que, por hipótese, y_{t+m+1} depende de $\{S_t\}$ somente através de s_{t+m+1} e também pela propriedade das Cadeias de Markov de 1ª ordem, o que faz com que $P\left\{s_{t+m+1} = l \middle| s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right\} = P\left\{s_{t+m+1} = l \middle| s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right\}$.

Assim, a igualdade entre $f\left(y_{t+m+1} \middle| s_t = j, s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right)$ e $f\left(y_{t+m+1} \middle| s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right)$ se verifica, fazendo com que $P\left\{s_t = j \middle| s_{t+1} = i, \mathbf{Y}_{t+m+1}; \theta\right\} = P\left\{s_t = j \middle| s_{t+1} = i, \mathbf{Y}_{t+m}; \theta\right\}$ (c.q.d)

Retornando à derivação do algoritmo, note agora que:

$$P\left\{s_t = j \middle| s_{t+1} = i, \mathbf{Y}_t; \theta\right\} = \frac{P\left\{s_t = j, s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}}{P\left\{s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}} = \frac{P\left\{s_t = j \middle| \mathbf{Y}_t; \theta\right\} \times P\left\{s_{t+1} = i \middle| s_t = j; \theta\right\}}{P\left\{s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}} =$$

$$= \frac{p_{ji} \cdot P\left\{s_t = j \middle| \mathbf{Y}_t; \theta\right\}}{P\left\{s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}}$$

Por outro lado, podemos também escrever:

$$P\left\{s_t = j, s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\} = P\left\{s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\} \cdot P\left\{s_t = j \middle| s_{t+1} = i, \mathbf{Y}_T; \theta\right\} =$$

$$= P\left\{s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\} \cdot P\left\{s_t = j \middle| s_{t+1} = i, \mathbf{Y}_t; \theta\right\} = P\left\{s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\} \cdot \frac{p_{ji} \cdot P\left\{s_t = j \middle| \mathbf{Y}_t; \theta\right\}}{P\left\{s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}}$$

Finalmente, a probabilidade suavizada do instante t é obtida somando-se para i = 1,2,...,k a expressão $P\left\{s_t = j, s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\}$, ou seja:

$$P\left\{s_t = j \middle| \mathbf{Y}_T; \theta\right\} = \sum_{i=1}^{k} P\left\{s_t = j, s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\} = \sum_{i=1}^{k} P\left\{s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\} \cdot \frac{p_{ji} \cdot P\left\{s_t = j \middle| \mathbf{Y}_t; \theta\right\}}{P\left\{s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}} =$$

$$= P\left\{s_t = j \middle| \mathbf{Y}_t; \theta\right\} \sum_{i=1}^{k} \frac{p_{ji} \cdot P\left\{s_{t+1} = i \middle| \mathbf{Y}_T; \theta\right\}}{P\left\{s_{t+1} = i \middle| \mathbf{Y}_t; \theta\right\}}$$

Utilizando a notação apresentada em III.1.1 e fazendo $p'_j = [p_{j1}\ p_{j2} \ldots\ p_{jk}]$, obtemos:

$$P\{s_t = j|Y_T;\theta\} = P\{s_t = j|Y_t;\theta\} \cdot p'_j \cdot (\zeta_{t+1|T}(\div)\zeta_{t+1|t})$$

Na expressão anterior, o vetor $p'_j = [p_{j1}\ p_{j2} \ldots\ p_{jk}]$ é equivalente à j-ésima linha de **P'** e o sinal $(\div)$ significa divisão elemento por elemento.

Para j = 1,2,...,k agrupamos as probabilidades $P\{s_t = j|Y_T;\theta\}$ num único vetor (k × 1):

$$\zeta_{t|T} = \zeta_{t|t} \odot (\mathbf{P'} \cdot (\zeta_{t+1|T}(\div)\zeta_{t+1|t})) \quad (2.III.3)$$

A expressão (2.III.3) assim obtida representa o algoritmo de cálculo das probabilidades suavizadas, as quais obtêm-se calculando a quantidade acima para t = T-1, T-2,...,1 , respectivamente.

IV- Estimadores de Máxima Verossimilhança

Hamilton (1990) demonstra que as expressões dos estimadores de máxima verossimilhança dos parâmetros relativos a um processo autorregressivo de ordem *m* que realiza mudanças markovianas de regimes, são obtidas resolvendo-se as seguintes equações normais:

$$\hat{p}_{ij}^{(\ell+1)} = \frac{\sum_{t=m-1}^{T} P\{s_t = j, s_{t-1} = i|Y_T;\hat{\theta}^{(\ell)}\}}{\sum_{t=m-1}^{T} P\{s_{t-1} = i|Y_T;\hat{\theta}^{(\ell)}\}}, \text{ para } i, j = 1,2,\ldots,k$$

$$\sum_{t=m+1}^{T} \sum_{st=1}^{k} \cdots \sum_{st-m=1}^{k} \frac{\partial \ln f(y_t|s_t, s_{t-1}, \ldots s_{t-m}, y_{t-1}, y_{t-2}, \ldots y_{t-m};\alpha)}{\partial\alpha}\Big|_{\alpha=\hat{\alpha}^{(\ell+1)}} \cdot P\{s_t, \ldots, s_{t-m}|Y_T;\hat{\theta}^{(\ell)}\} = 0,$$

onde α é o vetor θ a menos das probabilidades de transição.

Note que, já havíamos mencionado no final do capítulo anterior, um das implicações da hipótese de não dependência entre os regimes é o fato de que, na expressão dos estimadores de máxima verossimilhança dos parâmetros, as probabilidades suavizadas substituem as probabilidades filtradas. Daí a necessidade de se desenvolver um algoritmo para o cálculo das primeiras, tal como foi feito na seção III.2.1.

Analisaremos agora, alguns exemplos onde as expressões acima são aplicadas na prática.

Exemplo 1 - 2 regimes com $\mu_1 \neq \mu_2; \sigma_1^2 \neq \sigma_2^2$ e $m = 0$.

Neste exemplo, y_t é um processo no qual s_t segue uma Cadeia de Markov de primeira ordem composta de dois estados. A distribuição de y_t condicionada a $s_t = j$ é assumida normal com média μ_j e variância σ_j^2, para $j \in \{1,2\}$. Estas, como sabemos, variam de acordo com s_t e, por hipótese, não há estrutura autorregressiva em y_t. Assim, temos o seguinte vetor de parâmetros: $\theta = \{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p_{11}, p_{22}\}$. Diferenciando o logaritmo da densidade condicional de y_t com relação aos parâmetros, obtemos:

$$\frac{\partial lnf(y_t|s_t;\alpha)}{\partial \mu_j} = \frac{y_t - \mu_j}{\sigma_j^2}, \text{ se } s_t = j$$

$$= 0, \text{ caso contrário.}$$

$$\frac{\partial lnf(y_t|s_t;\alpha)}{\partial \sigma_j^2} = \frac{\sigma_j^2}{2} - \frac{(y_t - \mu_j)^2}{2}, \text{ se } s_t = j$$

$$= 0, \text{ caso contrário.}$$

Assim, com relação aos parâmetros μ_j e σ^2_j, as equações normais tomam a seguinte forma:

$$\sum_{t=1}^{T} \frac{y_t - \hat{\mu}_j^{(\ell+1)}}{\hat{\sigma}_j^{2;(\ell+1)}} \cdot P\left\{s_t = j \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\} = 0$$

$$\sum_{t=1}^{T} \left(\frac{\hat{\sigma}_j^{2;(\ell+1)}}{2} - \frac{\left(y_t - \hat{\mu}_j^{(\ell+1)}\right)^2}{2} \right) \cdot P\left\{s_t = j \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\} = 0$$

Resolvendo-se as equações acima e, considerando-se também as expressões relativas às probabilidades de transição, apresentadas no início desta seção, obtemos as expressões finais para os estimadores de máxima verossimilhança deste exemplo:

$$\hat{\mu}_j^{(\ell+1)} = \frac{\sum_{t=1}^{T} y_t \cdot P\left\{s_t = j \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_t = i \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}, \; j = 1,2$$

$$\hat{\sigma}_j^{2;(\ell+1)} = \frac{\sum_{t=1}^{T} \left(y_t - \hat{\mu}_j^{(\ell+1)}\right)^2 \cdot P\left\{s_t = j \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_t = i \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}, \; j = 1,2$$

$$\hat{p}_{11}^{(\ell+1)} = \frac{\sum_{t=1}^{T} P\left\{s_t = 1, s_{t-1} = 1 \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_{t-1} = 1 \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}$$

$$\hat{p}_{22}^{(\ell+1)} = \frac{\sum_{t=1}^{T} P\left\{s_t = 2, s_{t-1} = 2 \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_{t-1} = 2 \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}$$

Sobre este resultado é importante notar que:

1) As expressões obtidas para as médias e variâncias, μ_j e σ^2_j respectivamente, são semelhantes àquelas obtidas no capítulo anterior, quando expusemos o modelo de "Mistura de Distribuições I. I. D.". Entretanto, apresentam uma diferença básica: as observações são ponderadas agora pelas probabilidades suavizadas. Contudo, a estimação destes dois parâmetros dentro de cada regime ainda pode ser encarada como uma média ponderada das observações (y_t para μ_j e $\left(y_t - \mu_j\right)^2$ para σ^2_j) cujos pesos são proporcionais à probabilidade suavizada de cada observação correspondente ter sido gerada dentro do regime em questão.

2) Nas expressões dos estimadores para as probabilidades de transição encontramos, no numerador, as seguintes quantidades: $P\left\{s_t = 1, s_{t-1} = 1 \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}$ e $P\left\{s_t = 2, s_{t-1} = 2 \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}$. Vale lembrar que as mesmas são obtidas ao longo do algoritmo de cálculo das probabilidades suavizadas.

3) A obtenção dos valores numéricos finais relativos às estimativas de máxima verossimilhança segue o mesmo processo apresentado no capítulo anterior, o algoritmo EM. Assim, inicialmente introduzimos condições iniciais para $\theta = \{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p_{11}, p_{22}\}$ e, calcula-se daí as probabilidades suavizadas. Estas últimas são utilizadas nas expressões de máxima verossimilhança para se obter o novo valor de θ. O processo segue até que, de acordo com algum critério de convergência, os valores não sofram mais variações significativas a cada nova iteração.

Exemplo 2 - k regimes com $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_k; \sigma_1^2 = \sigma_2^2 = \ldots = \sigma^2$ e $m = 0$.

Neste caso y_t segue um processo que se alterna em k diferentes regimes porém sem apresentar estrutura dinâmica autorregressiva. No entanto, os regimes são diferenciados apenas pelas médias possuindo, em compensação, variâncias idênticas.

De acordo com estas suposições, teremos:

$$\frac{\partial lnf(y_t|s_t;\alpha)}{\partial \mu_j} = \frac{y_t - \mu_j}{\sigma_j^2}, \text{ se } s_t = j$$

$$= 0, \text{ caso contrário.}$$

$$\frac{\partial lnf(y_t|s_t;\alpha)}{\partial \sigma^2} = \frac{\sigma^2}{2} - \frac{(y_t - \mu_j)^2}{2}, \text{ se } s_t = j$$

O que produz para os estimadores de máxima verossimilhança para as médias e variância:

$$\hat{\mu}_j^{(\ell+1)} = \frac{\sum_{t=1}^{T} y_t \cdot P\{s_t = j | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}{\sum_{t=1}^{T} P\{s_t = i | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}, j = 1,2,\ldots,k$$

$$\hat{\sigma}^{2;(\ell+1)} = \frac{\sum_{j=1}^{k}\sum_{t=1}^{T}(y_t - \hat{\mu}_j^{(\ell+1)})^2 \cdot P\{s_t = j | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}{T}$$

Para as probabilidades de transição teremos:

$$\hat{p}_{ij}^{(\ell+1)} = \frac{\sum_{t=1}^{T} P\left\{s_t = j, s_{t-1} = i \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}{\sum_{t=1}^{T} P\left\{s_{t-1} = i \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}}, \text{ para } i = 1,2,...k; j = 1,2,...,k-1$$

$$\hat{p}_{ik} = 1 - \hat{p}_{i1} - ... - \hat{p}_{ik-1}$$

Da mesma forma do exemplo anterior, os valores finais para as estimativas de máxima verossimilhança são obtidos segundo o algoritmo EM.

Exemplo 3 - k regimes com $\mu_1 \neq \mu_2 \neq ... \neq \mu_k; \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$ e $m \neq 0$ com $\phi_1 \neq \phi_2 \neq ... \neq \phi_k$.

Neste caso permitimos uma dinâmica autorregressiva de ordem *m*, sendo que os parâmetros que a caracterizam (média e coeficientes autorregressivos) variam de acordo com a alternância entre os regimes. Somente a variância do termo aleatório u_t permanece constante ao longo do processo.

Do que foi exposto acima, podemos escrever:

$$y_t = \mu_{s_t} + \left(\phi_{1,s_t} \times y_{t-1}\right) + \left(\phi_{2,s_t} \times y_{t-2}\right) ... + \left(\phi_{m,s_t} \times y_{t-m}\right) + u_t,$$

onde u_t possui uma distribuição normal com média 0 e variância σ^2, constante ao longo do processo.

Utilizando-se da notação matricial, podemos reescrever a expressão acima como sendo:

$y_t = z_t' \beta_{s_t} + u_t$, onde:

$$z_t' = \left(1, y_{t-1}, y_{t-2}, ..., y_{t-m}\right) \text{ e } \beta_{s_t} = \left(\mu_{s_t}, \phi_{1,s_t}, \phi_{2,s_t}, ..., \phi_{m,s_t}\right).$$

Assim,

$$f\left(y_t \middle| s_t, s_{t-1}, ... s_{t-m}, y_{t-1}, y_{t-2}, ..., y_{t-m}; \alpha\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{\frac{-\left(y_t - z_t' \beta_{s_t}\right)^2}{2\sigma^2}\right\}$$

Derivando-se a expressão acima com relação aos parâmetros que compõem α obtemos:

$$\frac{\partial lnf\left(y_t \middle| s_t, s_{t-1}, \ldots s_{t-m}, y_{t-1}, y_{t-2}, \ldots, y_{t-m}; \alpha\right)}{\partial \beta_j} = \frac{\left(y_t - z_t'\beta_j\right) \cdot z_t}{\sigma^2}, \text{ se } s_t = j$$

$$= 0, \text{ caso contrário.}$$

$$\frac{\partial lnf\left(y_t \middle| s_t, s_{t-1}, \ldots s_{t-m}, y_{t-1}, y_{t-2}, \ldots, y_{t-m}; \alpha\right)}{\partial \sigma^2} = \frac{\sigma^2}{2} - \frac{\left(y_t - z_t'\beta_{s_t}\right)^2}{2}$$

Substituindo estes resultados nas expressões relativas às equações normais, obtemos:

$$\sum_{t=m+1}^{T} \frac{\left(y_t - z_t'\hat{\beta}_j^{(\ell+1)}\right) \cdot z_t}{\hat{\sigma}^{2;(\ell+1)}} \cdot P\left\{s_t = j \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\} = 0$$

$$\frac{\hat{\sigma}^{2;(\ell+1)}}{2} \cdot (T - m) = \sum_{t=m+1}^{T} \sum_{j=1}^{k} \frac{\left(y_t - z_t'\hat{\beta}_j^{(\ell+1)}\right)^2}{2} \cdot P\left\{s_t = j \middle| \mathbf{Y}_T; \hat{\theta}^{(\ell)}\right\}$$

Vale mencionar que as condições para a estimação das probabilidades de transição continuam as mesmas.

O algoritmo EM de estimação deve, então, ser utilizado da seguinte forma:

1) Atribui-se um valor inicial para θ e, calcula-se para cada j, as probabilidades suavizadas para todo t, $P\left\{s_t = j \middle| \mathbf{Y}_T; \theta\right\}$.

2) Novamente, para cada j, multiplicamos cada observação y_t e o vetor z_t pela raiz quadrada da probabilidade suavizada no instante t:

$$\tilde{y}_{t;j} = y_t \cdot \sqrt{P\left\{s_t = j \middle| \mathbf{Y}_T; \theta\right\}} \text{ e } \tilde{z}_{t;j} = z_t \cdot \sqrt{P\left\{s_t = j \middle| \mathbf{Y}_T; \theta\right\}}$$

3) A regressão de $\tilde{y}_{t;j}$ sobre $\tilde{z}_{t;j}$ produz um vetor β_j atualizado dos parâmetros, para cada j.

4) Efetuando-se esta regressão para todo j, obtemos ($\beta_1, \beta_2, \ldots, \beta_k$).

5) A estimativa atualizada para a variância σ^2 é obtida através da soma ao quadrado dos resíduos de cada uma das k regressões acima:

$$\hat{\sigma}^2 = \sum_{j=1}^{k} \sum_{t=m+1}^{T} \frac{\left(\tilde{y}_{t;j} - \tilde{z}_{t;j}' \ \beta_j\right)^2}{T-m}$$

6) Tendo em mãos o valor atualizado de θ, o processo é repetido até atingirmos o valor final para estimativa do vetor dos parâmetros, obedecendo algum critério de convergência.

V- Estimativas de Máxima Verossimilhança Obtidas Numericamente

A utilização do algoritmo EM na obtenção dos valores relativos aos estimadores de máxima verossimilhança requer, como vimos, a derivação analítica destes últimos. Uma vez obtidas as expressões dos estimadores, o processo iterativo pode ser aplicado mediante a introdução de conjuntos de valores iniciais para os parâmetros.

Entretanto, dependendo da estrutura do modelo a ser utilizado, a derivação analítica das expressões relativas aos estimadores de máxima verossimilhança torna-se uma tarefa de difícil solução, como por exemplo, quando introduzimos mudanças de regimes na especificação da variância condicional em modelos do tipo ARCH. Nestes casos, surge a necessidade de se encontrar uma forma alternativa de se obter as estimativas de máxima verossimilhança para os parâmetros. Assim, métodos numéricos são utilizados como forma alternativa de se conseguir encontrar estas estimativas.

De qualquer forma, o que queremos encontrar é o valor de θ que maximiza $f\left(y_T, y_{T-1}, \ldots, y_1; \theta\right)$.

Desenvolvendo-se a expressão acima, temos:

$$f\left(y_T, y_{T-1}, \ldots, y_1; \theta\right) = f\left(y_T \middle| \ y_{T-1}, \ldots, y_1; \theta\right) \cdot f\left(y_{T-1} \middle| \ y_{T-2}, \ldots, y_1; \theta\right) \cdot \ldots \cdot f\left(y_1; \theta\right)$$

E, aplicando-se o logaritmo natural em ambos os membros:

$$lnf\left(y_T, y_{T-1}, \ldots, y_1; \theta\right) = lnf\left(y_T \middle| \ y_{T-1}, \ldots, y_1; \theta\right) + lnf\left(y_{T-1} \middle| \ y_{T-2}, \ldots, y_1; \theta\right) + \ldots + lnf\left(y_1; \theta\right),$$ ou seja:

$$lnf\left(y_T, y_{T-1}, \ldots, y_1; \theta\right) = \sum_{t=1}^{T} lnf\left(y_t \mid y_{t-1}, \ldots, y_1; \theta\right)$$

Portanto, queremos encontrar θ que torna máximo o valor deste último somatório, cujas parcelas representam o logaritmo da densidade condicional de y_t dadas as observações passadas desde o instante 1 até o instante t-1.

Entretanto, este logaritmo da densidade condicional é um subproduto do algoritmo de cálculo das probabilidades filtradas, conforme foi exposto na seção III.1.2.. Assim, a tarefa que nos resta é somá-los e, utilizando-se algum método de otimização numérica (no caso estaremos utilizando o método BHHH), encontrar os valores dos parâmetros que maximizam esta soma.

Desta forma, o procedimento para o cálculo da log-verossimilhança pode ser descrito em 4 etapas, conforme descrito por Hamilton (1989):

Etapa 1 - Tomando-se inicialmente as k^{m+1} quantidades $p\left(s_t, s_{t-1}, \ldots, s_{t-m} \middle| \mathbf{Y}_t; \theta\right)$, calculamos as k^{m+2} expressões:

$$p\left(s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m} \middle| \mathbf{Y}_t; \theta\right) = p_{s_t, s_{t+1}} \times p\left(s_t, s_{t-1}, \ldots, s_{t-m} \middle| \mathbf{Y}_t; \theta\right)$$

Etapa 2 - Multiplicamos as k^{m+2} quantidades obtidas acima por $f\left(y_{t+1} \middle| y_t, y_{t-1}, \ldots, y_{t-m+1}, s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m+1}; \theta\right)$ para obtermos:

$$p\left(s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m}, y_{t+1} \middle| \mathbf{Y}_t; \theta\right) = f\left(y_{t+1} \middle| y_t, y_{t-1}, \ldots, y_{t-m+1}, s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m+1}; \theta\right) \times$$
$$\times\, p\left(s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m} \middle| \mathbf{Y}_t; \theta\right)$$

Etapa 3 - Obtemos a densidade condicional de y_{t+1} somando-se:

$$f\left(y_{t+1} \middle| \mathbf{Y}_t; \theta\right) = \sum_{s_{t+1}=1}^{k} \sum_{s_t=1}^{k} \ldots \sum_{s_{t-m}=1}^{k} p\left(s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m}, y_{t+1} \middle| \mathbf{Y}_t; \theta\right)$$

Etapa 4 - Somamos para os k possíveis valores de s_{t-m} as quantidades obtidas na etapa 2 e as dividimos pela densidade condicional encontrada na etapa 3 e obtemos as k^{m+1} quantidades:

$$p\left(s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m+1} \middle| \mathbf{Y}_{t+1}; \theta\right) = \frac{\sum_{s_{t-m}=1}^{k} p\left(s_{t+1}, s_t, s_{t-1}, \ldots, s_{t-m}, y_{t+1} \middle| \mathbf{Y}_t; \theta\right)}{f\left(y_{t+1} \middle| \mathbf{Y}_t; \theta\right)}$$

Ao final desta quarta etapa obtemos as quantidades que servirão de início para a iteração do próximo instante.

Desta forma, executando-se este procedimento para t = {1,2,...,T}, encontramos para um valor de θ fixado, a log-verossimilhança do modelo. A parti daí, utilizando-se métodos de otimização numérica, como por exemplo o método BHHH, podemos encontrar o valor de θ que maximiza a log-verossimilhança do modelo.

VI- Exemplo 1: Retorno Diário dos Títulos de 30 anos do Tesouro Norte-Americano

Depois de ter apresentado o arcabouço teórico dos modelos de mudanças de regimes ao longo das seções II a V, iremos agora ver como estes modelos funcionam na prática e, desta forma compararemos seu desempenho como modelos que seriam naturalmente ajustados à série temporal em questão caso tivéssemos tratado-a como um processo linear, ou seja, utilizando-se o enfoque tradicional de séries temporais.

Com o propósito de ilustrarmos a utilização da metodologia apresentada neste trabalho, abordaremos inicialmente o caso em que a série temporal em estudo não possui uma característica autorregressiva. Segundo a notação desenvolvida neste capítulo, estamos tratando do caso onde $m = 0$. Este caso é freqüentemente encontrado quando queremos modelar retornos diários de ativos financeiros, ao invés de seus respectivos preços, já que na maioria das vezes, as funções de autocorrelação e autocorrelação parcial amostrais não acusam estrutura autorregressiva.

Tomaremos então como um primeiro exemplo a série dos retornos diários, calculados em regime de capitalização contínua, do preço no mercado futuro dos títulos de 30 anos do Tesouro dos EUA, durante 03/01/1990 a 24/01/1997, totalizando 1778 observações. Assim, o retorno no instante t é dado por:

$$y_t = ln X_t - ln X_{t-1}$$

Onde X_t é o preço do referido título no instante t.

As Fig. 2.1 e Fig. 2.2 abaixo apresentam a evolução temporal dos preços do título (X_t) e dos respectivos retornos diários (y_t).

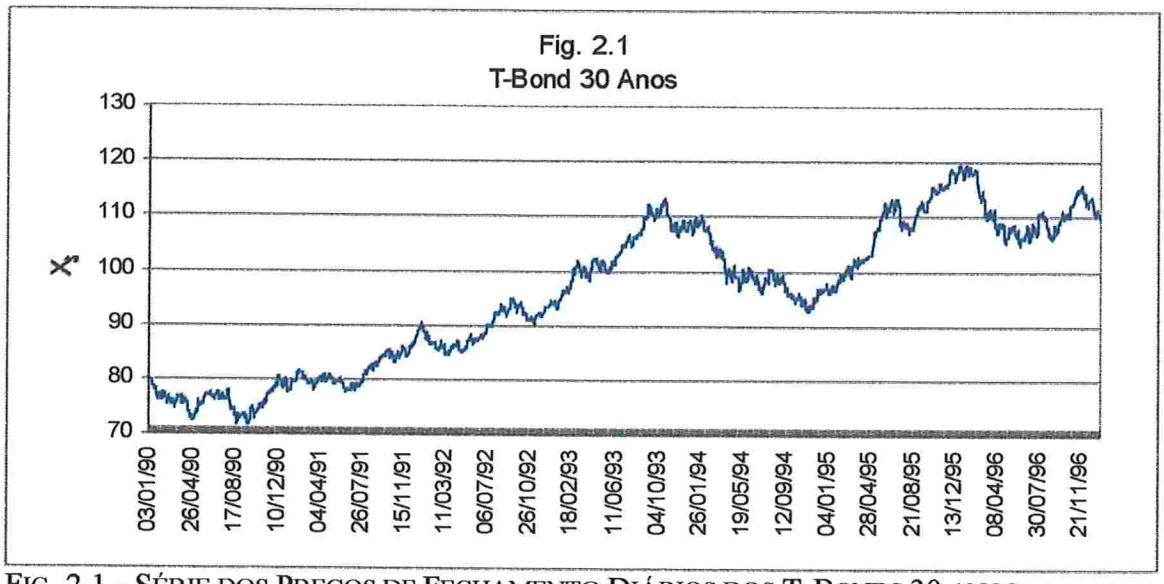

FIG. 2.1 - SÉRIE DOS PREÇOS DE FECHAMENTO DIÁRIOS DOS T-BONDS 30 ANOS

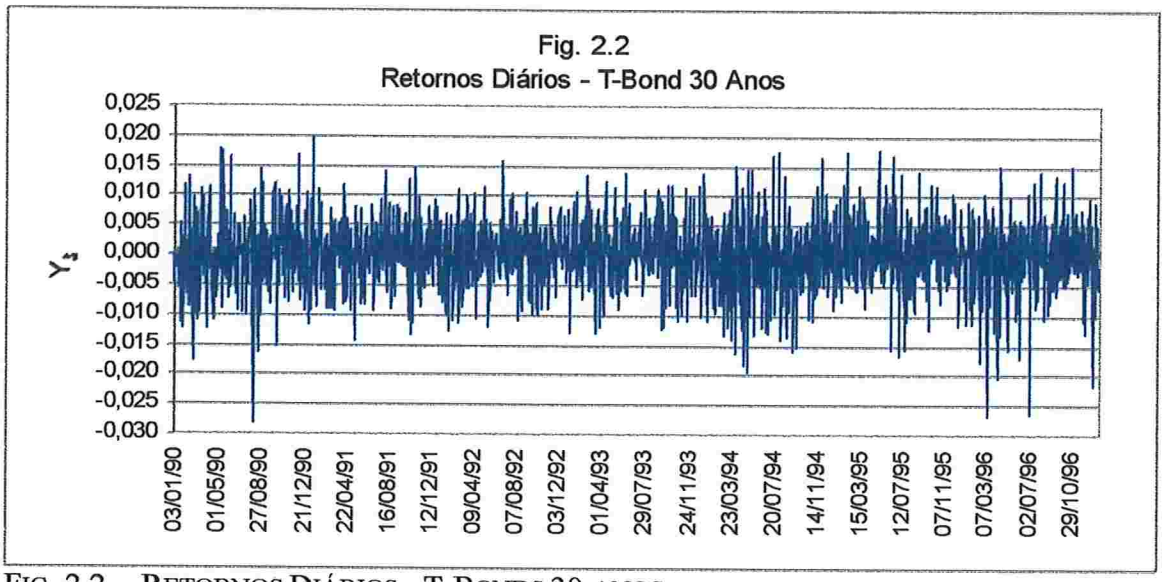

FIG. 2.2. - RETORNOS DIÁRIOS - T-BONDS 30 ANOS

Olhando para a evolução de X_t notamos que, inicialmente, os preços do título vinham evoluindo numa trajetória declinante, até setembro de 1990, quando passaram a apresentar uma tendência ascendente. Esta permaneceu até o início do segundo semestre de 1993, momento em que novamente observamos um recuo nos preços que prevaleceu até o final de 1994. Daí em diante, os preços tornaram novamente a subir, movimento que prevaleceu até o início de 1996, quando uma tendência de queda novamente se formou, porém de pouca duração na medida em que a partir do segundo semestre de 1996, os preços parecem ter reiniciado uma trajetória de elevação.

Em casos como este, onde nos deparamos com reversões de tendências, nosso interesse em aplicar os modelos de mudanças de regimes é o de justamente poder identificar estas reversões, diminuindo consideravelmente o erro de predição. Assim, podemos definir a

existência de dois regimes, um para o qual predomina a tendência ascendente (regime 1) e outro para o qual prevalece a tendência declinante dos preços (regime 2). Além disto, é de se supor que no regime 1 haja uma incidência maior dos retornos diários positivos enquanto que no regime 2, os retornos negativos deverão ocorrer com maior freqüência.

Outra característica que pode ser incorporada nos modelos de mudanças de regimes é o próprio processo de alternância entre eles. Como podemos perceber pelo gráfico de X_t (Fig. 2.1), aparentemente identificamos 6 reversões de tendências (mudanças de regimes) em pouco mais de 7 anos de observações. Ou seja, as mudanças ocorrem mas as tendências não são de pequena duração. Desta forma, ao modelar as mudanças de regimes por uma Cadeia de Markov, é de se esperar que as probabilidades de transição relativas à permanência dentro do regime (p_{11}, p_{22},...) sejam elevadas. Por outro lado, não precisam ser necessariamente idênticas: é possível que, embora haja a presença de dois ou mais regimes governando a evolução da série, um ou mais deles prevaleçam sobre os demais.

Assim, pelo sua maior flexibilidade e na medida em que permitem incorporar características fundamentais das séries financeiras em sua especificação, os modelos de mudanças de regimes, à primeira vista, devem fornecer projeções com erros menores que os modelos lineares tradicionais.

Feito este pequeno comentário, retornaremos ao nosso exemplo em questão. Pelo que foi exposto acima, parece-nos adequado tratar a série dos retornos como sendo governada por dois regimes: regime 1 para tendências de alta, caracterizadas pela ocorrência com maior freqüência de retornos diários positivos e regime 2 para tendências de baixa, caracterizadas pela ocorrência com maior freqüência de retornos diários negativos. Além disto, a transição entre os regimes é dada por uma Cadeia de Markov de primeira ordem. Assumindo, então, normalidade temos:

$$y_t | s_t = 1 \sim N(\mu_1, \sigma_1^2)$$

$$y_t | s_t = 2 \sim N(\mu_2, \sigma_2^2)$$

$$P\{s_t = 1 | s_{t-1} = 1\} = p_{11}$$

$$P\{s_t = 2 | s_{t-1} = 2\} = p_{22}$$

$$P\{s_t = 2 | s_{t-1} = 1\} = 1 - p_{11}$$

$$P\{s_t = 1 | s_{t-1} = 2\} = 1 - p_{22}$$

Utilizando-se a medida invariante da Cadeia de Markov como distribuição do estado inicial, temos:

$$P\{s_1 = 1\} = \frac{1 - p_{22}}{(1 - p_{11}) + (1 - p_{22})} = \pi_1$$

$$P\{s_1 = 2\} = \frac{1 - p_{11}}{(1 - p_{11}) + (1 - p_{22})} = 1 - \pi_1$$

Neste caso, o vetor de parâmetros a ser estimado é composto por:

$$\theta = \{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p_{11}, p_{22}\}$$

E, na ausência de dinâmica autorregressiva, os estimadores de máxima verossimilhança são obtidos utilizando-se as expressões desenvolvidas no exemplo 1 da seção IV, quais sejam:

$$\hat{\mu}_j^{(\ell+1)} = \frac{\sum_{t=1}^{T} y_t \cdot P\{s_t = j | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}{\sum_{t=1}^{T} P\{s_t = i | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}, j = 1,2$$

$$\hat{\sigma}_j^{2;(\ell+1)} = \frac{\sum_{t=1}^{T} (y_t - \hat{\mu}_j^{(\ell+1)})^2 \cdot P\{s_t = j | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}{\sum_{t=1}^{T} P\{s_t = i | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}, j = 1,2$$

$$\hat{p}_{11}^{(\ell+1)} = \frac{\sum_{t=1}^{T} P\{s_t = 1, s_{t-1} = 1 | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}{\sum_{t=1}^{T} P\{s_{t-1} = 1 | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}$$

$$\hat{p}_{22}^{(\ell+1)} = \frac{\sum_{t=1}^{T} P\{s_t = 2, s_{t-1} = 2 | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}{\sum_{t=1}^{T} P\{s_{t-1} = 2 | \mathbf{Y}_T; \hat{\theta}^{(\ell)}\}}$$

A Tabela 2.1 a seguir apresenta os resultados alcançados, comparando com as estimativas que seriam obtidas se o modelo utilizado tivesse sido:

$y_t = \mu + u_t$

Com $u_t \sim N(0,\sigma^2)$ e $Cov(u_t, u_s) = 0$, para qualquer $t \neq s$.

Tabela 2.1 - Modelos Alternativos para a Série dos T-Bonds 30 anos

PARÂMETRO	HAMILTON	RANDOM WALK
μ	-	1.76×10^{-4} (1.36×10^{-4})
μ_1	5.87×10^{-4} (1.82×10^{-4})	-
μ_2	-4.60×10^{-4} (2.98×10^{-4})	-
σ^2	-	3.28×10^{-5} (1.10×10^{-6})
σ_1^2	2.46×10^{-5} (1.43×10^{-6})	-
σ_2^2	4.50×10^{-5} (2.75×10^{-6})	-
p_{11}	0.9910 (0.0045)	-
p_{22}	0.9867 (0.0067)	-
Log Verossimilhança	6679.06	6655.41
Teste LR [valor-p]	47.3 [0.000]	-
AIC	-13346.1	-13306.8
SC	-13313.2	-13295.9

Dos resultados obtidos, algumas considerações podem ser feitas:

1) O modelo de mudanças de regimes se ajusta melhor ao conjunto de dados quando comparado com o modelo de passeio aleatório, o qual seria ajustado se tivéssemos considerado apenas a classe dos modelos puramente lineares. Isto se evidencia quando comparamos o AIC e o SC dos dois modelos, uma vez que para o modelos de alternância de regimes tanto o AIC quanto o SC são menores que os obtidos para o modelo alternativo de passeio aleatório. Adicionalmente, o teste de razão das verossimilhanças rejeita a hipótese nula de passeio aleatório em favor do modelo de mudanças de regimes.

2) A parametrização do modelo em dois regimes - tendência de alta, caracterizada por uma predominância de retornos diários positivos; tendência de baixa, na qual a incidência de retornos diários negativos é maior - parece ser bastante adequada e coerente com nossas suposições iniciais. De fato, o retorno médio do regime 1 é positivo enquanto que o do regime 2 é negativo. Além disto, como podemos observar, as variâncias dos regimes são

bastante diferentes: a do regime 2 é quase o dobro daquela estimada para o regime 1. Isto sugere que o movimento da taxa de juros é mais abrupto quando estas sobem (vale lembrar que a taxa de juros e o preço do título variam no sentido inverso) do que quando caem. O que, de certa forma. é coerente com a experiência econômica: as autoridades monetárias quando promovem elevações dos juros visando, entre outras coisas, conter um crescimento excessivo da demanda agregada na economia, costumam fazê-lo "once for all". Entretanto, sempre que a trajetória é declinante, esta costuma ser bem mais suave. Assim, as distintas variâncias estimadas para ambos os regimes já sugerem uma assimetria na evolução da série temporal, fato que os modelos lineares tradicionais não consideram.

3) Outra maneira de se visualizar esta assimetria é olharmos para as estimativas referentes às probabilidades de transição da Cadeia de Markov. Vemos que, ambos os regimes possuem uma persistência bastante elevada: a probabilidade de permanecermos no regime 1 no instante t dado que no instante t-1 estávamos neste mesmo regime é de 99.10% enquanto que a probabilidade de permanecermos no regime 2 no instante t dado que no instante t-1 estávamos neste mesmo regime é de 98.67%. Assim, o tempo médio de permanência no regime 1 é maior que no regime 2:

$$T_1 = \frac{1}{1-p_{11}} = \frac{1}{1-0.9910} \cong 111 \text{ dias úteis}$$

$$T_2 = \frac{1}{1-p_{22}} = \frac{1}{1-0.9867} \cong 75 \text{ dias úteis}$$

4) Como suspeitávamos, as mudanças de regimes não ocorrem com muita freqüência. Uma forma de se visualizar esta característica é constatar as elevadas probabilidades de transição da Cadeia de Markov quando estas se referem à permanência dentro do regime, tal como foi exposto acima. Porém, podemos constatar este fato tomando-se a evolução das probabilidades suavizadas, conforme Fig. 2.3 a seguir.

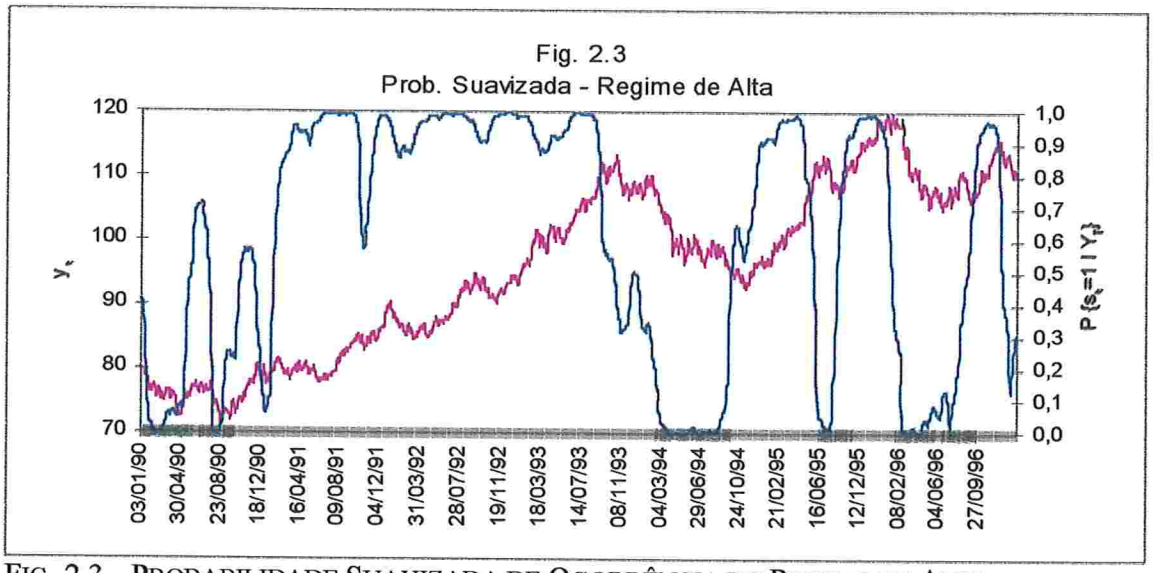

FIG. 2.3 - PROBABILIDADE SUAVIZADA DE OCORRÊNCIA DO REGIME DE ALTA

A Fig. 2.3 mostra a evolução da probabilidade suavizada de nos encontrarmos no regime 1. Se considerássemos como critério para identificar as mudanças de regimes o cruzamento no ponto de 50%, seja de baixo para cima ou de cima para baixo, da linha evolutória da probabilidade suavizada, teríamos então identificado 6 tendências de alta e 6 de baixa. Ou seja, ao longo dos 1785 dias úteis de observações, a nossa série teria realizado 12 mudanças de regimes, uma a cada 149 dias úteis em média, o que daria um tempo médio de 7 meses para cada regime (cada mês em média possui 21 dias úteis).

5) Por fim, o modelo de mudanças de regimes reduz significativamente o erro de predição. Para tanto, devemos neste ponto, derivar sua equação de predição, calcular os erros a horizontes de predição determinados e compará-los com os obtidos pelo ajuste do modelo de passeio aleatório.

Queremos, então, calcular:

$$E\left[y_{t+h}\middle|y_t,\ldots,y_1\right]=\int y_{t+h}\cdot f\left(y_{t+h}\middle|y_t,\ldots,y_1\right)dy_{t+h}=\int\sum_{j=1}^{k}y_{t+h}\cdot f\left(y_{t+h}.s_{t+h}=j\middle|y_t,\ldots,y_1\right)dy_{t+h}=$$

$$=\int\sum_{j=1}^{k}y_{t+h}\cdot f\left(y_{t+h}\middle|s_{t+h}=j,y_t,\ldots,y_1\right)\times P\left\{s_{t+h}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}dy_{t+h}=$$

Invertendo o sinal da integração com o somatório, obtemos:

$$E\left[y_{t+h}\middle|y_t,y_{t-1},\ldots,y_1\right]=\sum_{j=1}^{k}P\left\{s_{t+h}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}\times\int y_{t+h}\cdot f\left(y_{t+h}\middle|s_{t+h}=j,y_t,y_{t-1},\ldots,y_1\right)dy_{t+h}$$

$$=\sum_{j=1}^{k}E\left[y_{t+h}\middle|s_{t+h}=j,y_t,y_{t-1},\ldots,y_1\right]\times P\left\{s_{t+h}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}$$

Pelas hipóteses iniciais do modelo, temos que:

$$E\left[y_{t+h}\middle|s_{t+h}=j,y_t,y_{t-1},\ldots,y_1\right]=E\left[y_{t+h}\middle|s_{t+h}=j\right]=\mu_j$$

Então, chegamos a:

$$E\left[y_{t+h}\middle|y_t,y_{t-1},\ldots,y_1\right]=\sum_{j=1}^{k}E\left[y_{t+h}\middle|s_{t+h}=j\right]\times P\left\{s_{t+h}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}$$

Como, porém, neste exemplo k = 2 pois existem apenas 2 regimes, então:

$$E\left[y_{t+h}\middle|y_t,y_{t-1},\ldots,y_1\right]=\sum_{j=1}^{2}\mu_j\times P\left\{s_{t+h}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}=$$

$$=\mu_1\times P\left\{s_{t+h}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}+\mu_2\times\left(1-P\left\{s_{t+h}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}\right)=$$

$$=\mu_2+\left(\mu_1-\mu_2\right)\times P\left\{s_{t+h}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}$$

Note agora que:

$$P\left\{s_{t+h}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}=\sum_{j=1}^{2}P\left\{s_{t+h}=1,s_{t+h-1}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}=$$

$$=\sum_{j=1}^{2}P\left\{s_{t+h}=1\middle|s_{t+h-1}=j,y_t,y_{t-1},\ldots,y_1\right\}\times P\left\{s_{t+h-1}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}=$$

$$=\sum_{j=1}^{2}P\left\{s_{t+h}=1\middle|s_{t+h-1}=j\right\}\times P\left\{s_{t+h-1}=j\middle|y_t,y_{t-1},\ldots,y_1\right\}=$$

$$=p_{11}\times P\left\{s_{t+h-1}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}+p_{21}\times P\left\{s_{t+h-1}=2\middle|y_t,y_{t-1},\ldots,y_1\right\}=$$

$$=p_{11}\times P\left\{s_{t+h-1}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}+\left(1-p_{22}\right)\times\left(1-P\left\{s_{t+h-1}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}\right)=$$

$$=\left(-1+p_{11}+p_{22}\right)\times P\left\{s_{t+h-1}=1\middle|y_t,y_{t-1},\ldots,y_1\right\}+\left(1-p_{22}\right)$$

Substituindo-se na equação de predição, obtemos:

$$E\left[y_{t+h}\middle|y_t, y_{t-1}, \ldots, y_1\right] = \mu_2 + (\mu_1 - \mu_2) \times \left[(-1 + p_{11} + p_{22}) \times P\left\{s_{t+h-1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} + (1 - p_{22})\right]$$

$$= \mu_2 + (\mu_1 - \mu_2) \times \left[(-1 + p_{11} + p_{22}) \times P\left\{s_{t+h-1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} + \frac{(1 - p_{22})}{(2 - p_{11} - p_{22})} \times (2 - p_{11} - p_{22})\right]$$

$$= \mu_2 + (\mu_1 - \mu_2) \times \left[(-1 + p_{11} + p_{22}) \times P\left\{s_{t+h-1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} - \pi_1 \times (-1 - 1 + p_{11} + p_{22})\right]$$

$$= \mu_2 + (\mu_1 - \mu_2) \times \left\{\pi_1 + (-1 + p_{11} + p_{22}) \times \left[P\left\{s_{t+h-1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} - \pi_1\right]\right\}$$

Finalmente, por indução pode-se demonstrar que:

$$E\left[y_{t+h}\middle|y_t, y_{t-1}, \ldots, y_1\right] = \mu_2 + (\mu_1 - \mu_2) \times \left\{\pi_1 + (-1 + p_{11} + p_{22})^h \times \left[P\left\{s_t = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} - \pi_1\right]\right\}$$

De fato, fazendo para h = 1, obtemos:

$$E\left[y_{t+1}\middle|y_t, y_{t-1}, \ldots, y_1\right] = \mu_2 + (\mu_1 - \mu_2) \times \left\{\pi_1 + (-1 + p_{11} + p_{22}) \times \left[P\left\{s_{t+1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} - \pi_1\right]\right\}$$

Assumindo que a equação genérica é válida para o horizonte de previsão h, precisamos mostrar que a expressão também é válida para o horizonte h + 1. Para tanto, escrevemos:

$$E\left[y_{t+h+1}\middle|y_t, y_{t-1}, \ldots, y_1\right] = \sum_{j=1}^{2} \mu_j \times P\left\{s_{t+h+1} = j\middle|y_t, y_{t-1}, \ldots, y_1\right\} =$$

$$= \mu_1 \times P\left\{s_{t+h+1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} + \mu_2 \times \left(1 - P\left\{s_{t+h+1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\}\right) =$$

$$= \mu_2 + (\mu_1 - \mu_2) \times P\left\{s_{t+h+1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\}$$

Mas, por outro lado:

$$P\left\{s_{t+h+1} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} = \sum_{j=1}^{2} P\left\{s_{t+h+1} = 1, s_{t+h} = j\middle|y_t, y_{t-1}, \ldots, y_1\right\} =$$

$$= \sum_{j=1}^{2} P\left\{s_{t+h+1} = 1\middle|s_{t+h} = j, y_t, y_{t-1}, \ldots, y_1\right\} \times P\left\{s_{t+h} = j\middle|y_t, y_{t-1}, \ldots, y_1\right\} =$$

$$= p_{11} \times P\left\{s_{t+h} = 1\middle|y_t, y_{t-1}, \ldots, y_1\right\} + p_{21} \times P\left\{s_{t+h} = 2\middle|y_t, y_{t-1}, \ldots, y_1\right\} =$$

$$= p_{11} \times P\{s_{t+h} = 1|y_t, y_{t-1}, \ldots, y_1\} + (1 - p_{22}) \times (1 - P\{s_{t+h} = 1|y_t, y_{t-1}, \ldots, y_1\}) =$$

$$= (-1 + p_{11} + p_{22}) \times P\{s_{t+h} = 1|y_t, y_{t-1}, \ldots, y_1\} + (1 - p_{22})$$

Isto faz com que:

$$E[y_{t+h+1}|y_t, \ldots, y_1] = \mu_2 + (\mu_1 - \mu_2) \times \{(-1 + p_{11} + p_{22}) \times P\{s_{t+h} = 1|y_t, \ldots, y_1\} + (1 - p_{22})\}$$

(2.V.1)

Mas pela hipótese indutiva:

$$P\{s_{t+h} = 1|y_t, y_{t-1}, \ldots, y_1\} = \frac{E[s_{t+h}|y_t, y_{t-1}, \ldots, y_1] - \mu_2}{\mu_1 - \mu_2} =$$

$$= \frac{\mu_2 + (\mu_1 - \mu_2) \times \{\pi_1 + (-1 + p_{11} + p_{22})^h \times [P\{s_t = 1|y_t, y_{t-1}, \ldots, y_1\} - \pi_1]\} - \mu_2}{(\mu_1 - \mu_2)} =$$

$$= \{\pi_1 + (-1 + p_{11} + p_{22})^h \times [P\{s_t = 1|y_t, y_{t-1}, \ldots, y_1\} - \pi_1]\}$$

Substituindo esta última igualdade em (2.V.I) e rearranjando os termos obtemos:

$$E[y_{t+h+1}|y_t, \ldots, y_1] = \mu_2 + (\mu_1 - \mu_2) \times \{\pi_1 + (-1 + p_{11} + p_{22})^{h+1} \times [P\{s_t = 1|y_t, \ldots, y_1\} - \pi_1]\}$$

(c.q.d)

A expressão acima representa, portanto, a equação de predição para os retornos diários, calculados sob o regime de capitalização contínua, do preço no mercado futuro dos títulos de longo prazo do tesouro norte-americano.

Denominemos, agora, $E[y_{t+h}|y_t, y_{t-1}, \ldots, y_1] = \hat{y}_{t+h|t}$. Se estivermos interessados em projetar a evolução do preço dos títulos, ou seja, X_t, teremos então:

$$x_{t+h|t} = exp\{ln\, x_t + \hat{y}_{t+1|t} + \hat{y}_{t+2|t} + \ldots + \hat{y}_{t+h|t}\}$$

A Tabela 2.2. informa os erros quadráticos totais de predição calculados para os modelos de mudanças de regime e passeio aleatório, em horizontes de predição (h) variando de 1 a 20 dias úteis. Além disto, a Tabela 2.2 expõe o ganho em termos de redução percentual no

erro quadrático total do modelo de mudanças de regimes comparativamente ao modelo linear de passeio aleatório. Vale lembrar que, para este último, a equação de predição é:

$$x_{t+h|t} = exp\{ln\, x_t + h \times \bar{y}\}, \text{ onde } \bar{y} = \sum_{t=1}^{T} \frac{y_t}{T}$$

Tabela 2.2 - Erros Quadráticos de Projeção dos Modelos para o T-Bond 30 Anos

Horizonte (h)	Mudanças de Regimes	Passeio Aleatório	Ganho em %
1	548.6	549.5	0.16%
2	1107.3	1110.5	0.29%
3	1668.7	1675.5	0.40%
4	2200.7	2212.7	0.54%
5	2661.1	2681.1	0.75%
6	3114.1	3143.7	0.94%
7	3489.7	3536.5	1.32%
8	3907.8	3973.3	1.65%
9	4304.4	4396.7	2.10%
10	4648.1	4779.2	2.74%
11	5023.4	5192.7	3.26%
12	5441.9	5648.1	3.65%
13	5864.7	6109.7	4.01%
14	6347.1	6626.0	4.21%
15	6849.6	7165.2	4.41%
16	7404.9	7759.0	4.56%
17	7947.4	8339.9	4.71%
18	8509.9	8940.0	4.81%
19	9028.4	9501.4	4.98%
20	9536.1	10056.6	5.18%

Assim, notamos que a utilização do modelo de mudanças de regimes neste exemplo tende a melhorar o erro de predição, comparativamente ao modelo linear tradicional que seria ajustado à série em questão. Além disto, à medida em que se amplia o horizonte de projeção a superioridade do modelo não linear evidencia-se ainda mais.

VII- Exemplo 2: Spread Diário Entre Dois Títulos Da Dívida Externa Brasileira

Apresentaremos agora, um exemplo de como os modelos de mudanças de regimes se aplicam em situações em que a série temporal em questão possui uma dinâmica

autorregressiva. Para tanto, nos utilizaremos do "spread" diário entre dois títulos da dívida externa brasileira, o C-Bond e o Par-Bond.

Por "spread" diário entendemos como sendo o quociente entre os preços de mercado, apurados no fechamento dos mesmos, dos respectivos títulos. Assim se c_t é o preço de mercado apurado no fechamento do C-Bond no instante (dia) t e se p_t é o preço de mercado apurado no fechamento do Par-Bond no instante (dia) t, então o "spread" neste dia é dado por:

$$y_t = \frac{c_t}{p_t}$$

A Fig. 2.4 a seguir mostra a evolução do "spread" y_t durante o período compreendido entre 21/01/94 e 25/03/96, perfazendo um total de 488 observações.

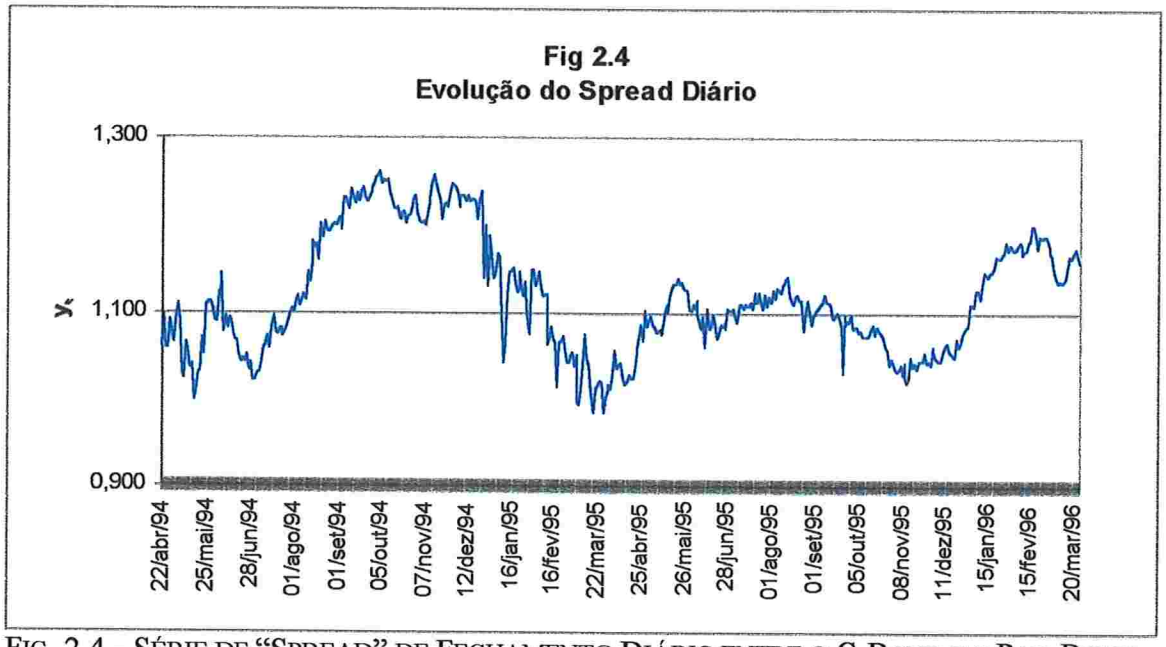

FIG. 2.4 - SÉRIE DE "SPREAD" DE FECHAMENTO DIÁRIO ENTRE O C-BOND E O PAR-BOND

Como podemos perceber pela Fig. 2.4, há uma tendência do "spread" oscilar em torno de um nível médio, porém apresentando uma dinâmica autorregressiva, na medida em que os afastamentos e os retornos ao nível médio da série ocorrem através de mini-tendências.

Esta característica de existir um nível médio "atrator" deriva da própria natureza dos dois títulos, uma vez que os fatores que promovem alterações em seus respectivos preços de mercado são geralmente comuns (nível da taxa internacional de juros, expectativas de desvalorização cambial interna, risco-país, etc.).

Entretanto, percebe-se também que quando ocorre um afastamento tanto para baixo como para cima do nível médio, o "spread" tende a oscilar em torno de novos níveis inferiores /

superiores, embora numa duração menor, antes de iniciar seu retorno ao nível médio global da série. Assim, aparentemente fica caracterizada a existência de 3 regimes diferentes:

- O primeiro seria quando a série estivesse oscilando num patamar inferior ao nível médio global da série, sendo este de pequena duração;

- O segundo seria caracterizado por oscilações ao redor do nível médio global da série, sendo este o de maior duração;

- O terceiro quando as oscilações estivesse ao redor de um patamar superior ao nível médio global, sendo também de duração pequena.

Adicionalmente, poderíamos supor que a transição entre os regimes seguisse uma Cadeia de Markov e, mais ainda, que o coeficiente autorregressivo dos 3 regimes fossem diferentes.

Isto posto, podemos escrever o modelo acima descrito da seguinte forma:

$$\left(y_t - \mu_{s_t}\right) = \phi_{s_t} \times \left(y_{t-1} - \mu_{s_{t-1}}\right) + u_t, \text{ onde } u_t \sim N\left(0;\sigma^2\right) \text{ e } s_t \in \{1,2,3\} \text{ para todo t}$$

Além disto, temos as seguintes probabilidades de transição:

$$P\{s_t = 1|s_{t-1} = 1\} = p_{11} \qquad P\{s_t = 1|s_{t-1} = 2\} = p_{21} \qquad P\{s_t = 1|s_{t-1} = 3\} = p_{31}$$

$$P\{s_t = 2|s_{t-1} = 1\} = p_{12} \qquad P\{s_t = 2|s_{t-1} = 2\} = p_{22} \qquad P\{s_t = 2|s_{t-1} = 3\} = p_{32}$$

$$P\{s_t = 3|s_{t-1} = 1\} = p_{13} \qquad P\{s_t = 3|s_{t-1} = 2\} = p_{23} \qquad P\{s_t = 3|s_{t-1} = 3\} = p_{33}$$

Vale lembrar ainda que, dadas as propriedades das probabilidades de transição:

$$p_{11} + p_{12} + p_{13} = 1 \qquad p_{21} + p_{22} + p_{32} = 1 \qquad p_{31} + p_{32} + p_{33} = 1$$

Devido ao elevado número de parâmetros a serem estimados (13 no total) e sabendo-se da complicações introduzidas pela dinâmica autorregressiva na obtenção analítica dos estimadores de máxima verossimilhança, optaremos pelo métodos numérico de estimação dos parâmetros, aplicando-se o algoritmo descrito na seção V.

Testaremos o modelo acima descrito em duas versões: a primeira onde o coeficiente autorregressivo é o mesmo para os diferentes regimes, ou seja, $\phi_{s_t} = \phi, \forall s_t$. Desta maneira, reduzimos de 13 para 11 o número de parâmetros a serem estimados. A segunda versão é a que explicitamos nos parágrafos anteriores: o coeficiente autorregressivo também varia conforme os regimes. Chamaremos a primeira versão de SW(3,1) e a segunda versão de SW(3,3).

Os resultados assim obtidos serão comparados com aqueles que seriam gerados utilizando o modelo linear clássico para este caso, ou seja:

$$(y_t - \mu) = \phi \times (y_{t-1} - \mu) + u_t, \text{ onde } u_t \sim N(0; \sigma^2).$$

A Tabela 2.3 a seguir apresenta os resultados obtidos para os dois modelos em questão:

Tabela 2.3 - Modelos Alternativos para o Spread entre os Títulos da Dívida Externa Brasileira

	SW (3,1)	SW (3,3)	AR(1)
μ	-	-	1.1174 (0.0239)
μ_1	1.0677 (0.0290)	1.0358 (0.0068)	-
μ_2	1.1043 (0.0291)	1.0771 (0.0073)	-
μ_3	1.1439 (0.0293)	1.1311 (0.0162)	-
ϕ	0.983 (0.0095)	-	0.965 (0.0137)
ϕ_1	-	0.6945 (0.0444)	-
ϕ_2	-	0.9977 (0.0075)	-
ϕ_3	-	0.4994 (0.4312)	-
σ	0.0108 (0.0004)	0.0113 (0.0003)	0.0174 (0.0003)
p_{11}	0.3785 (0.1203)	0.4582 (0.1389)	-
p_{12}	0.5135 (0.1268)	0.5078 (0.1607)	-
p_{21}	0.0639 (0.0217)	0.0366 (0.0106)	-
p_{22}	0.9215 (0.0263)	0.9634 (0.0438)	-
p_{31}	0.0164 (0.0090)	0	-
p_{32}	0.0089 (0.0122)	0.1139 (0.1416)	-
Log Verossimilhança	1391.8	1395.9	1289.1
Teste LR [valor-p]	205.4 [0.000]	213.6 [0.000]	-
AIC	-2761.6	-2765.8	-2572.2
SC	-2715.6	-2711.4	-2559.6

Conforme os resultados obtidos, o modelo autorregressivo com alternância de regimes ajusta-se melhor ao conjunto de dados da série temporal analisada, uma vez que os critérios de ajuste AIC e SC geram valores menores para estes modelos. O teste das razões de

verossimilhança também apontam nesta direção, rejeitando a hipótese nula de passeio aleatório.

No entanto, os elevados desvios padrões associados aos parâmetros referentes ao estado 3, tanto nas probabilidades de transição quanto no coeficiente autorregressivo sinalizam um possível superdimensionamento do modelo. Ou seja, que modelos autorregressivos com transições ocorrendo não em 3, mas em 2 regimes seriam mais adequados.

Assim, reduzindo de 3 para 2 o número dos possíveis regimes, construímos os seguintes modelos:

a) AR(1) - SW(2,1):

$$\left(y_t - \mu_{s_t}\right) = \phi \times \left(y_{t-1} - \mu_{s_{t-1}}\right) + u_t \text{, onde } u_t \sim N\left(0;\sigma^2\right) \text{ e } s_t \in \{1,2\} \text{ para todo t}$$

Além disto, temos as seguintes probabilidades de transição:

$$P\{s_t = 1 | s_{t-1} = 1\} = p_{11} \qquad P\{s_t = 2 | s_{t-1} = 2\} = p_{22}$$

b) AR(1) - SW(2,2):

$$\left(y_t - \mu_{s_t}\right) = \phi_{s_t} \times \left(y_{t-1} - \mu_{s_{t-1}}\right) + u_t \text{, onde } u_t \sim N\left(0;\sigma^2\right) \text{ e } s_t \in \{1,2\} \text{ para todo t}$$

As probabilidades de transição, são semelhantes ao do modelo anterior.

$$P\{s_t = 1 | s_{t-1} = 1\} = p_{11} \qquad P\{s_t = 2 | s_{t-1} = 2\} = p_{22}$$

As Tabelas 2.4 e 2.5 em seguida, apresentam as estimativas para os parâmetros dos modelos acima descritos e efetuam a comparação com os modelos anteriormente ajustados.

Tabela 2.4 - Ajuste dos Modelos de 2 Regimes para o Spread entre os Títulos da Dívida Externa Brasileira

Parâmetro	SW (2,1)	SW (2,2)
μ_1	1.0753 (0.0310)	1.0392 (0.0064)
μ_2	1.1231 (0.0311)	1.0814 (0.0070)
ϕ	0.982 (0.0098)	-
ϕ_1	-	0.6917 (0.0443)
ϕ_2	-	0.9964 (0.0082)
σ	0.0122 (0.0003)	0.0116 (0.0004)
p_{11}	0.4398 (0.0914)	0.4523 (0.1052)
p_{22}	0.9647 (0.0104)	0.9624 (0.0102)
Log Verossimilhança	1371.5	1392.0
Teste LR [valor-p]	164.8 [0.000]	205.8 [0.000]
AIC	-2731.0	-2770.0
SC	-2705.8	-2740.6

Tabela 2.5 - Critérios de Seleção de Modelos para o Spread entre os Títulos da Dívida Externa Brasileira

Modelo	Log-Verossimilhança	Nº de parâmetros	AIC	SC
AR(1) - SW(2,1)	1371.5	6	-2731.0	-2705.8
AR(1) - SW(2,2)	**1392.0**	**7**	**-2770.0**	**-2740.6**
AR(1) - SW(3,1)	1391.8	11	-2761.6	-2715.6
AR(1) - SW(3,3)	1395.9	13	-2765.8	-2711.4
AR(1)	1289.1	3	-2572.2	-2559.6

Fica evidenciado, novamente, que os modelos de mudança de regimes produzem um ajuste mais adequado à série temporal, comparativamente aos modelos tradicionais. No exemplo em estudo, tanto os modelos de alternância entre dois como em três regimes possuem AIC e SC menores que o ajuste produzido através de um modelo autorregressivo de primeira ordem sem alternância de regime.

Observando-se o AIC e o SC dos modelos alternativos, a escolha recai sobre o modelo AR(1)-SW(2,2), uma vez que para este modelo os valores relativos ao AIC e SC são menores que os demais. Isto confirma a nossa suspeita levantada anteriormente de um provável superdimensionamento ao ajustarmos os modelos de 3 regimes para a série do "spread".

A Fig. 2.5 a seguir explicita a relação entre a série temporal e a probabilidade suavizada de ocorrência do regime 1, o de menor freqüência de acordo com as estimativas, associado a um nível médio inferior.

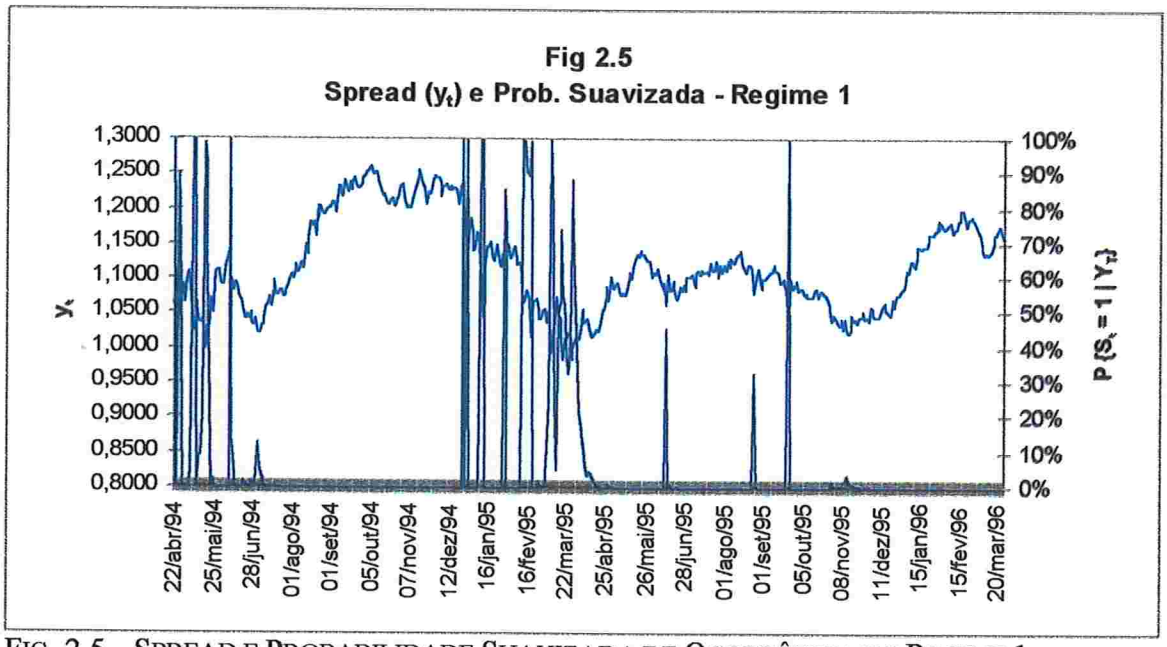

FIG. 2.5 - SPREAD E PROBABILIDADE SUAVIZADA DE OCORRÊNCIA DO REGIME 1

Através dos resultados obtidos neste exemplo, algumas considerações adicionais merecem ser feitas:

1) Da mesma forma que encontramos quando analisamos o exemplo dos títulos de 30 anos do tesouro norte-americano, o "spread" entre os títulos da dívida externa brasileira em questão possuem um caráter assimétrico. Assim, na maior parte do tempo, conforme mostra a Fig. 2.5, o regime 2 é predominante, o qual é caracterizado pela seguinte expressão:

$$(y_t - 1.0814) = 0.9964 \times (y_{t-1} - 1.0814) + u_t$$

De fato, os dados indicam que em 93% das vezes a probabilidade de nos encontrarmos no regime 2 supera os 50%.

A ocorrência do regime 1, dado pela expressão $(y_t - 1.0392) = 0.6917 \times (y_{t-1} - 1.0392) + u_t$, restringe-se a momentos esporádicos, associados com quedas bruscas no valor do "spread".

Outra maneira de se visualizar esta assimetria são as estimativas referentes às probabilidades de transição da Cadeia de Markov. Vemos que a probabilidade de permanecermos no regime 1 no instante t dado que no instante t-1 estávamos neste mesmo regime é de 45.23% enquanto que a probabilidade de permanecermos no regime 2 no instante t dado que no instante t-1 estávamos neste mesmo regime é de 96.24%. Assim, o tempo médio de permanência no regime 2 é bem maior que no regime 1:

$$T_1 = \frac{1}{1-p_{11}} = \frac{1}{1-0.4523} \cong 2 \text{ dias úteis}$$

$$T_2 = \frac{1}{1-p_{22}} = \frac{1}{1-0.9624} \cong 27 \text{ dias úteis}$$

2) As mudanças de regimes novamente não ocorrem com muita freqüência. Diante do que foi evidenciado, o regime 1 possui as características de um caso excepcional. Por algum motivo (mudanças bruscas no risco-país, ou até mesmo a falta de liquidez em algum dos papéis) o "spread" acaba se distanciando do seu ponto de equilíbrio, gerando uma distorção no mercado. No entanto esta é corrigida de maneira rápida, uma vez que o tempo médio de permanência neste regime não chega a 2 dias úteis.

3) O modelo de mudanças de regimes reduz significativamente o erro de predição. Para tanto, devemos neste ponto, derivar sua equação de predição, calcular os erros a horizontes de predição e um passo e compará-los com os obtidos pelo ajuste do modelo autorregressivo tradicional.

Nosso intuito é, calcular $E\left[y_{t+1} \middle| y_t, y_{t-1}, \ldots, y_1\right]$ para todo o instante t e, de posse destas quantidades, calcular a soma dos erros quadráticos para ambos os modelos.

Para o caso do modelo AR(1) a equação de predição a um passo, largamente conhecida, é dada por:

$$E\left[y_{t+1} \middle| y_t, y_{t-1}, \ldots, y_1\right] = \mu + \phi \cdot (y_t - \mu)$$

Já para o modelo autorregressivo de primeira ordem com alternância de regimes, a derivação da equação de predição a um passo segue abaixo.

$$E\left[y_{t+1} \middle| y_t, y_{t-1}, \ldots, y_1\right] = \int y_{t+1} \cdot f\left(y_{t+1} \middle| y_t, \ldots, y_1\right) dy_{t+1} = \int y_{t+1} \cdot \sum_{i=1}^{k} f\left(y_{t+1}, s_t = i \middle| y_t, \ldots, y_1\right) dy_{t+1} =$$

$$= \int y_{t+1} \cdot \sum_{i=1}^{k} f\left(y_{t+1} \middle| s_t = i, y_t, \ldots, y_1\right) \cdot P\left\{s_t = i \middle| y_t, \ldots, y_1\right\} dy_{t+1}$$

Invertendo a ordem de integração, obtemos:

$$E\left[y_{t+1}|y_t, y_{t-1}, \ldots, y_1\right] = \sum_{i=1}^{k} \int y_{t+1} \cdot f\left(y_{t+1}|s_t = i, y_t, \ldots, y_1\right) dy_{t+1} \cdot P\left\{s_t = i|y_t, \ldots, y_1\right\}$$

Ou seja, $E\left[y_{t+1}|y_t, y_{t-1}, \ldots, y_1\right] = \sum_{i=1}^{k} E\left[y_{t+1}|s_t = i, y_t, \ldots, y_1\right] \cdot P\left\{s_t = i|y_t, \ldots, y_1\right\}$

Por raciocínio análogo, é possível demonstrar que:

$$E\left[y_{t+1}|s_t = i, y_t, y_{t-1}, \ldots, y_1\right] = \sum_{j=1}^{k} E\left[y_{t+1}|s_{t+1} = j, s_t = i, y_t, \ldots, y_1\right] \cdot P\left\{s_{t+1} = j|s_t = i, y_t, \ldots, y_1\right\}$$

Mas, por hipótese do nosso modelo $P\left\{s_{t+1} = j|s_t = i, y_t, \ldots, y_1\right\} = p_{ij}$

Assim, chegamos à expressão final:

$$E\left[y_{t+1}|y_t, y_{t-1}, \ldots, y_1\right] = \sum_{i=1}^{k} \left(\sum_{j=1}^{k} E\left[y_{t+1}|s_{t+1} = j, s_t = i, y_t, \ldots, y_1\right] \cdot p_{ij} \right) \cdot P\left\{s_t = i|y_t, y_{t-1}, \ldots, y_1\right\},$$

sendo que todas as quantidades envolvidas são conhecidas:

a) $P\left\{s_t = i|y_t, y_{t-1}, \ldots, y_1\right\}$ é a inferência filtrada no instante t

b) p_{ij} é o elemento da i-ésima coluna e j-ésima linha da matriz das probabilidades de transição estimadas.

c) $E\left[y_{t+1}|s_{t+1} = j, s_t = i, y_t, y_{t-1}, \ldots, y_1\right] = \mu_j + \phi_j \cdot \left(y_t - \mu_i\right)$

Isto posto, resta-nos apresentar os erros quadráticos de predição associados a cada um dos modelos, o que é dado pela Tabela 2.6 a seguir.

Tabela 2.6 - Erros de Projeção a um Passo dos Modelos para "Spread"

Modelo	Soma dos Erros Quadráticos de Predição a um Passo
AR (1)	0.1446
AR(1) - SW (2,2)	0.1370

Desta forma, o modelo autorregressivo com alternância de regimes reduz em 5.3% o erro médio de predição para o "spread" entre os títulos da dívida externa brasileira, C-Bond e Par-Bond, oferecendo um ajuste melhor à série temporal em relação ao modelo autorregressivo tradicional.

No próximo capítulo veremos como os modelos de alternância de regimes podem ser também aplicados a séries temporais estacionárias cuja variância condicional não é constante, hipótese muito utilizada para se descrever o comportamento de séries financeiras.

CAPÍTULO 3 - MUDANÇAS DE REGIME EM VARIÂNCIAS CONDICIONAIS

I - Introdução

Como já mencionamos anteriormente, nestas últimas duas décadas os estatísticos têm encontrado cada vez mais evidências de que séries temporais financeiras geralmente seguem algum tipo de processo não linear.

Dentre as causas que produziriam esta característica a mais popular delas tem sido a suposição de que a variância condicional, ao contrário da variância não condicional, não é constante ao longo do tempo. Esta idéia tomou corpo mais precisamente no início da década de 80 a partir do trabalho de Engle (1982) sobre a inflação no Reino Unido, momento em que foi apresentado pelo primeira vez o modelo ARCH. Desde então, a popularidade deste modelo e suas mais diversas variantes cresceu assustadoramente, principalmente em aplicações sobre séries financeiras onde as especificações do tipo ARCH oferecem um poder de explicação muito superior aos modelos tradicionais de séries temporais.

Uma outra razão contribuiu para a difusão dos modelos do tipo ARCH: o mercado de opções. Uma opção de compra / venda é um título negociado no mercado financeiro que dá o direito ao seu portador o direito, ou a opção (daí a origem do nome), de comprar / vender um determinado ativo financeiro (ativo-base) até/ou em uma data específica (data de exercício ou prazo de maturidade) por um preço previamente estabelecido (preço de exercício). Para obter este direito, o portador paga um preço ao adquirir a opção, é o denominado prêmio da opção.

Dada as características e a flexibilidade deste título (ao mesmo tempo são negociadas opções de diferentes preços e datas de exercícios) o investidor consegue estruturar uma carteira de opções maximizando seu resultado dependendo das suas expectativas quanto a evolução futura do preço do ativo-base. Por exemplo, suponhamos que o preço de uma determinada ação seja hoje de R$ 100,00 e se um investidor acredita que na data de exercício o preço desta ação estará distante de R$ 100,00, não importa se seja acima ou abaixo deste valor, o investidor então simultaneamente adquire hoje uma opção de compra e uma opção de venda do ativo-base ambas ao preço de exercício seja de R$ 100,00 e, desta forma, o portfólio destas opções será tal que quanto mais distante de R$ 100,00 estiver o preço do ativo-base (ação) na data de exercício maior será o retorno para o investidor.

A popularidade do mercado de opções cresceu fundamentalmente a partir de 1973 devido a duas razões: a criação da Chicago Board Options Exchange (CBOE) e a publicação do trabalho de Black & Scholes (1973), os quais desenvolveram o mais utilizado até hoje método de cálculo do prêmio de opções, derivado da solução de uma equação diferencial estocástica, cuja única variável desconhecida é a variância esperada dos retornos calculados

em regime de capitalização contínua do preço do ativo-base (comumente conhecida como volatilidade), do momento em que a opção é negociada até a data de exercício. Ou seja, para se precificar corretamente uma opção, o investidor necessita de um modelo de projeção para a variância dos retornos do ativo-base, caso contrário corre o risco de pagar um preço superior pela opção ou de vendê-la abaixo do preço justo.

Entretanto, o modelo de Black & Scholes pressupõe que a variância seja uma função determinística do tempo. Desta maneira, para se incorporar a hipótese de heterocedasticidade, é necessário utilizar outros métodos de precificação de opções. Neste sentido, o modelo binomial, que pode ser encontrado em Hull (1997), contemplando variância não constante é uma das alternativas para se precificar opções num ambiente em que a volatilidade é estocástica.

Neste âmbito, os modelos ARCH, principalmente o GARCH (1,1) ainda são os mais utilizados na projeção das volatilidades futuras, uma vez que são de aplicação relativamente simples e, por permitir alterações na variância condicional, ao mesmo tempo vão de encontro com a percepção dos investidores de que a volatilidade nem sempre parece constante, pelo contrário.

Nas próximas seções deste capítulo, apresentaremos de maneira resumida o modelo GARCH (p,q), suas principais características, e como este pode ser incorporardo nos modelos de mudanças markovianas de regime dentro da especificação da variância condicional. Para tanto, veremos as abordagens de Cai (1994), de Hamilton & Susmel (1994) e de Dueker (1997).

II - O Modelo GARCH (p,q)

II.1 - Definições Básicas e Principais Resultados

Seja $y_t = \mu + u_t$ um processo estocástico definido para os retornos diários calculados em regime de capitalização contínua de um ativo financeiro qualquer.

Suponha que u_t seja governado por:

$$u_t = \sigma_t \cdot v_t$$

onde (v_t) é uma seqüência i.i.d. com média zero e variância unitária. Além disto, σ_t é independente de v_t para todo t.

A variância condicional de u_t, representada por σ_t^2, é especificada como sendo uma função dos valores passados de u_t e σ_t^2.

$$\sigma_t^2 = g(u_{t-1}, u_{t-2}, \ldots, \sigma_{t-1}^2, \sigma_{t-2}^2, \ldots)$$

Assumiremos que v_t segue uma distribuição normal com média zero e variância unitária e que $g(\cdot)$ depende linearmente dos σ_t^2 anteriores e do quadrado das realizações passadas de u_t conforme especificado abaixo:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \cdot u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \cdot \sigma_{t-j}^2$$

A especificação assim obtida acima é conhecida como GARCH (p,q), introduzida por Bollerslev (1986). Quando p = 0, o processo transforma-se em ARCH (q).

Podemos rescrever a expressão acima como sendo:

$$\sigma_t^2 = \alpha_0 + \alpha(B)u_t^2 + \beta(B)\sigma_t^2$$

Analogamente aos processos ARMA tradicionais, o processo $\{u_t\}$ será estacionário se $\alpha(1) + \beta(1) < 1$ e terá como variância não-condicional:

$$\mathrm{Var}(u_t) = \frac{\alpha_0}{1 - \alpha(1) - \beta(1)}$$

Desta forma, obtemos que a largamente utilizada especificação GARCH (1,1) é dada por:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \cdot u_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2, \text{ com } \mathrm{Var}(u_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

II.2 - Persistência

A questão da persistência em modelos GARCH está intimamente ligada à equação de projeção da variância condicional.

Considere, por exemplo, o modelo GARCH (1,1), onde a expressão da variância condicional σ_t^2 seja dada por:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \cdot u_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2$$

De acordo com as definições básicas dos modelos GARCH, podemos escrever para o modelo GARCH (1,1):

$$E\left[u_{t+m}^2 \middle| u_{t+m-1}, u_{t+m-2}, \ldots\right] = E\left[\sigma_{t+m}^2 \cdot v_{t+m}^2 \middle| u_{t+m-1}, u_{t+m-2}, \ldots\right] = E\left[\sigma_{t+m}^2 \middle| u_{t+m-1}, u_{t+m-2}, \ldots\right] =$$

$$= E\left[\alpha_0 + \alpha_1 \cdot u_{t+m-1}^2 + \beta_1 \cdot \sigma_{t+m-1}^2 \middle| u_{t+m-1}, u_{t+m-2}, \ldots\right] =$$

$$= \alpha_0 + \alpha_1 \cdot E\left[u_{t+m-1}^2 \middle| u_{t+m-1}, u_{t+m-2}, \ldots\right] + \beta_1 \cdot E\left[\sigma_{t+m-1}^2 \middle| u_{t+m-1}, u_{t+m-2}, \ldots\right] =$$

$$= \alpha_0 + \alpha_1 \cdot u_{t+m-1}^2 + \beta_1 \cdot E\left[u_{t+m-1}^2 \middle| u_{t+m-2}, u_{t+m-3}, \ldots\right]$$

Tomando-se em ambos os membros as esperanças condicionais, obtemos:

$$E\left[u_{t+m}^2 \middle| u_t, u_{t-1}, u_{t-2}, \ldots\right] = \alpha_0 + \alpha_1 \cdot E\left[u_{t+m-1}^2 \middle| u_t, u_{t-1}, u_{t-2}, \ldots\right] + \beta_1 \cdot E\left[u_{t+m-1}^2 \middle| u_t, u_{t-1}, u_{t-2}, \ldots\right] =$$

$$= \alpha_0 + \left(\alpha_1 + \beta_1\right) \cdot E\left[u_{t+m-1}^2 \middle| u_t, u_{t-1}, u_{t-2}, \ldots\right]$$

Chamando então de $\sigma_{t+m|t}^2$ a previsão no horizonte m da variância condicional, da expressão acima obtemos:

$$\sigma_{t+m|t}^2 = \alpha_0 + \left(\alpha_1 + \beta_1\right) \cdot \sigma_{t+m-1|t}^2$$

Temos então que a previsão no horizonte m da variância condicional segue um a equação a diferenças finitas de primeira ordem, cujo parâmetro de decaimento δ é igual a $\alpha_1 + \beta_1$.

Além disto é possível mostrar que quando m tende ao infinito, $\sigma^2_{t+m|t}$ converge para a variância não-condicional do processo, $\mathrm{Var}(u_t) = \dfrac{\alpha_0}{1-\alpha_1-\beta_1}$.

O principal problema que surge é que, geralmente o parâmetro δ é muito próximo de 1. O que significa dizer que nestes casos, qualquer alteração mais significativa da volatilidade produzirá efeitos não neglicenciáveis por um horizonte de tempo elevado, o que nem sempre na prática se verifica. Por exemplo, o "crash" em outubro de 1987 das bolsas norte-americanas ou os efeitos nas bolsas latinas no final de 1994 em decorrência da crise mexicana provocou, repentinamente, uma elevação na volatilidade das ações. Porém, o tempo em que o mercado levou para se normalizar foi muito menor do que aqueles que teoricamente apontavam os modelos GARCH para as variâncias condicionais destes dois momentos, devido ao parâmetro δ ser muito próximo de 1.

Daí então a necessidade de se adotar um modelo que permita um ajuste mais rápido nas variâncias condicionais após a ocorrências de grandes choques tais como os acima citados. Uma hipótese plausível para tal talvez seja a de que a natureza de grandes choques é diferente daquela dos pequenos choques. Assim, um modelo de mudança markoviana de regimes, desta vez aplicado sobre o processo gerador das variâncias condicionais, alternando regimes de alta, média e baixa volatilidade, seja o mais adequado no tratamento das séries financeiras, nas quais a possibilidade de ocorrência de grandes choques é bem maior do que nas séries econômicas em geral.

III - Os Modelos SWARCH

Os modelos que introduzem as mudanças de regimes nas variâncias condicionais ficaram conhecidos como Modelos SWARCH, sendo que apresentaremos dois exemplos deste modelos: o proposto por Cai (1994) e o utilizado por Hamilton & Susmel (1994). Lembramos que, exclusivamente nesta seção, estaremos preservando as notações originais introduzidas pelos dois autores, o que de certa forma, diferem das quais estamos utilizando ao longo deste trabalho.

III.1 - O Modelo de Cai

Em seu estudo sobre o excesso de retorno dos títulos do tesouro norte-americano de três meses sobre os títulos de um mês, Cai (1994) propõe o seguinte modelo:

$$r_t = c_o + c_1 \cdot S_t + z_t$$

$$z_t = b_1 \cdot z_{t-1} + b_2 \cdot z_{t-2} + \ldots + b_k \cdot z_{t-k} + e_t$$

$$e_t = u_t \cdot \sqrt{h_t}\ ;\ \text{onde } u_t \text{ é i.i.d. } N(0,1)$$

$$h_t = \gamma(S_t) + \sum_{i=1}^{g} \alpha_i \cdot e_{t-i}^2\ ;\ \text{com } \alpha_i \geq 0 \text{ e } \gamma(S_t) = \gamma_o + \gamma_1 \cdot S_t\ ;\ \text{com } \gamma_0 > 0; \gamma_1 > 0$$

Assim, fica caracterizado um processo autorregressivo de ordem k, sendo que tanto a média quanto a variância condicional, esta última evoluindo segundo um processo ARCH(g), ficam condicionadas aos regimes. Cai (1994) adota dois regimes. Assim. S_t assume os valores 0 ou 1, sendo que S_t segue uma Cadeia de Markov de ordem 1 com as seguintes probabilidades de transição:

$$P\{S_t = 0 | S_{t-1} = 0\} = p$$

$$P\{S_t = 1 | S_{t-1} = 0\} = 1 - p$$

$$P\{S_t = 1 | S_{t-1} = 1\} = q$$

$$P\{S_t = 0 | S_{t-1} = 1\} = 1 - q$$

Esta especificação permite mudanças estruturais tanto na média quanto na constante (intercepto) da especificação ARCH. Este sistema é então dito estar no estado de baixa volatilidade quando $S_t = 0$ e no estado de alta volatilidade quando $S_t = 1$.

É demonstrado também que, a variância não condicional deste processo é dada por:

$E[e_t^2] = \dfrac{\gamma_0 + \pi \cdot \gamma_1}{1 - \sum_{i=1}^{g} \alpha_i}$, onde $\pi = \dfrac{1-p}{2-p-q}$ é a probabilidade não condicional de se encontrar no estado 1.

III.2 - O Modelo de Hamilton

Uma outra forma de se incorporar a alternância dos regimes na variância condicional foi apresentada por Hamilton & Susmel (1994) em seu estudo sobre o mercado acionário. Em linhas gerais, é muito semelhante ao proposto por Cai (1994), diferindo apenas na forma de

se especificar a variância condicional do processo: enquanto Cai (1994) permite mudança no intercepto da especificação ARCH, Hamilton & Susmel (1994) opta por uma mudança de escala. Assim, este último modelo é descrito na seguinte forma:

$$y_t = \mu + \phi \cdot y_{t-1} + u_t$$

$$u_t = \sqrt{g_{s_t}} \times \tilde{u}_t$$

$\tilde{u}_t = h_t \cdot v_t$, onde v_t é i.i.d. N(0,1)

$$h_t^2 = a_o + a_1 \cdot \tilde{u}_{t-1}^2 + \ldots + a_q \cdot \tilde{u}_{t-q}^2$$

Assim, temos para a variância condicional um processo ARCH (q) sendo que a variável $\tilde{u}_t$ é multiplicada por $\sqrt{g_1}$ quando o processo se encontra no regime 1, por $\sqrt{g_2}$ quando estiver no regime 2, e assim por diante. A idéia básica é modelar as mudanças de regime como alterações na escala do processo da variância condicional.

Vale mencionar que a transição entre os regimes também é governada por uma Cadeia de Markov de primeira ordem. No entanto, Hamilton & Susmel (1994) utiliza a existência de até quatro possíveis regimes em sua análise, normalizando para a unidade o fator do primeiro estado, fazendo $g_1 = 1$ e, assim, $g_j \geq 1$, para j=2,3,...,k.

Adicionalmente, Hamilton & Susmel (1994) também introduz na especificação da variância condicional o que é conhecido como "efeito alavancagem": a variância condicional reage diferentemente ao sinal de u_t. Retornos inesperados negativos tendem a afetar mais intensamente a variância condicional do que retornos inesperados positivos da mesma magnitude. Esta hipótese, embora não a trataremos nesta ocasião, corresponde à seguinte especificação:

$h_t^2 = a_o + a_1 \cdot \tilde{u}_{t-1}^2 + \ldots + a_q \cdot \tilde{u}_{t-q}^2 + \xi \cdot d_{t-1} \cdot \tilde{u}_{t-1}^2$, onde $d_{t-1} = 1$ se $\tilde{u}_{t-1} \leq 0$ e $d_{t-1} = 0$ se $\tilde{u}_{t-1} > 0$, com $\xi > 0$.

A hipótese de normalidade para os resíduos não é restritiva. Seguindo Hamilton & Susmel (1994) permitiremos também que os resíduos v_t (u_t no caso do modelo de Cai (1994)) possam ser gerados por uma distribuição t-student com η graus de liberdade (normalizados para ter variância unitária). Assim, a normalidade seria obtida na medida em que η se tornar suficientemente grande.

IV - Os Modelos SWGARCH

Os modelos da seção anterior, conforme apresentados, não são do tipo GARCH. Ou seja, na expressão da variância condicional h_t^2, não faz parte a variância condicional do instante anterior.

Na verdade, há um sério problema se considerarmos a variância condicional do instante anterior na especificação GARCH: h_t^2 fica dependente de todo o processo $\{S_t\}$ desde o instante 0 até o instante t. Teríamos, então, para o exemplo do modelo de Cai (1994):

$$h_t^2(s_t, s_{t-1}, \ldots, s_0) = \gamma_{s_t} + a_1 \cdot u_{t-1}^2 + b_1 \cdot h_{t-1}^2(s_{t-1}, s_{t-2}, \ldots, s_0)$$

Desta forma, o processo de estimação fica extremamente comprometido, principalmente no caso das séries temporais financeiras, geralmente de tamanhos maiores que 1.000 observações. No caso de apenas dois regimes alternantes, seria necessário o cálculo de $2^{1.000}$ probabilidades para se avaliar a variância condicional no instante t = 1.000.

Dueker (1997) consegue contornar este problema utilizando um procedimento proposto por Kim (1994), qual seja, de desconsiderar um certo número de s_t's passados na especificação GARCH. Segundo Kim (1994), se h_t^2 é tido como um processo autorregresivo de ordem p, sem perda significativa de acurácia no cálculo da log-verossimilhança, pode-se considerar h_t^2 como uma função de p + 1 valores mais recentes de s_t. Assim, no caso GARCH (1,1), onde p = 1, h_t^2 fica sendo função de s_t e s_{t-1}. Valores de s_t anteriores além de t - 2 são negligenciados:

$$h_t^{2(i,j)} = h_t^2(s_t = i, s_{t-1} = j)$$

Com isto, a expressão da variância condicional acaba sendo:

$$h_t^{2(i,j)} = \gamma_{s_t} + a_1 \cdot u_{t-1}^2 + b_1 \cdot h_{t-1}^{2(j)} \qquad (3.IV.1)$$

Além disto, de posse das probabilidades utilizadas no algoritmo de maximização da log-verossimilhança, podemos calcular:

$$h_t^{2(i)} = \sum_{j=1}^{2} P\{s_{t-1} = j | s_t = i, Y_t\} \times h_t^{2(i,j)} \qquad (3.IV.2)$$

As expressões (3.IV.1) e (3.IV.2) conforme apresentadas acima permitem o cálculo das variâncias condicionais desde o instante 0 até o instante t, bastando apenas introduzir um valor inicial em $h_{t-1}^{2(j)}$, na expressão (3.IV.1) para t = 1, e que pode ser a variância não condicional da série.

Este procedimento simplifica sobremaneira a estimação de modelos SWGARCH, uma vez que reduz a dependência de h_t^2 em relação a s_t apenas aos dois últimos valores de s_t.

Para o modelo de Hamilton & Susmel (1994), diferentemente do que acontece no modelo de Cai (1994) h_t^2 depende de s_t somente através de s_{t-1}. Isto acaba por simplificar ainda mais as expressões (3.IV.1) e (3.IV.2), que ficam caracterizadas por:

$$h_t^{2(j)} = a_0 + a_1 \cdot \tilde{u}_{t-1}^2 + b_1 \cdot \hat{h}_{t-1}^2, \text{ ou ainda } h_t^{2(j)} = a_0 + a_1 \cdot \frac{u_{t-1}^2}{g_{s_{t-1}}} + b_1 \cdot \hat{h}_{t-1}^2$$

$$\hat{h}_t^2 = \sum_{j=1}^{2} P\{s_{t-1} = j | Y_t\} \times h_t^{2(j)}$$

Na seção a seguir, trataremos de unificar as notações dos modelos de Cai (1994) e Hamilton & Susmel (1994), incorporando h_{t-1}^2 na expressão da variância condicional, a fim de utilizar estes modelos no ajuste da série dos retornos diários da ação Telebrás PN.

V - Modelos Utilizados neste Capítulo

A fim de exemplificar como são utilizados os modelos propostos nas seções III e IV, com o intuito de se realizar previsões sobre o comportamento futuro da variância condicional de ativos financeiros, promoveremos algumas modificações nos modelos originais. Mais precisamente, desconsideraremos o "efeito alavancagem" na variância condicional e, finalmente, assumiremos que a mudança entre os regimes só afeta esta última, conforme Hamilton & Susmel (1994), ou seja, o parâmetro μ não varia conforme os regimes.

De acordo com estas adaptações, os modelos de Cai (1994) e Hamilton & Susmel (1994) ficam restritos a 6 classes

<u>Classe 1: Modelo ARCH(1) - Cai Adaptado</u>

$$y_t = \mu + \phi \cdot y_{t-1} + u_t$$

$u_t = v_t \cdot h_t$; onde v_t é i.i.d. N(0,1)

$h_t^2 = \gamma_{s_t} + a_1 \cdot u_{t-1}^2$; com $a_1 \geq 0$ e com $\gamma_{s_t} > 0; \forall s_t$

<u>Classe 2: Modelo ARCH(2) - Cai Adaptado</u>

$$y_t = \mu + \phi \cdot y_{t-1} + u_t$$

$u_t = v_t \cdot h_t$; onde v_t é i.i.d. N(0,1)

$h_t^2 = \gamma_{s_t} + a_1 \cdot u_{t-1}^2 + a_2 \cdot u_{t-2}^2$; com $a_1 \geq 0; a_2 \geq 0$ e com $\gamma_{s_t} > 0; \forall s_t$

Classe 3: Modelo GARCH(1,1) - Cai Adaptado

$y_t = \mu + \phi \cdot y_{t-1} + u_t$

$u_t = v_t \cdot h_t$; onde v_t é i.i.d. N(0,1)

$h_t^2 = \gamma_{s_t} + a_1 \cdot u_{t-1}^2 + b_1 \cdot h_{t-1}^2$; com $a_1 \geq 0, b_1 \geq 0$ e com $\gamma_{s_t} > 0; \forall s_t$

Classe 4: Modelo ARCH(1) - Hamilton Adaptado

$y_t = \mu + \phi \cdot y_{t-1} + u_t$

$u_t = \sqrt{g_{s_t}} \times \tilde{u}_t$

$\tilde{u}_t = h_t \cdot v_t$, onde v_t é i.i.d. N(0,1)

$h_t^2 = a_o + a_1 \cdot \tilde{u}_{t-1}^2$; com $a_0 > 0, a_1 \geq 0$

Classe 5: Modelo ARCH(2) - Hamilton Adaptado

$y_t = \mu + \phi \cdot y_{t-1} + u_t$

$u_t = \sqrt{g_{s_t}} \times \tilde{u}_t$

$\tilde{u}_t = h_t \cdot v_t$, onde v_t é i.i.d. N(0,1)

$h_t^2 = a_o + a_1 \cdot \tilde{u}_{t-1}^2 + a_2 \cdot \tilde{u}_{t-2}^2$; com $a_0 > 0, a_1 \geq 0, a_2 \geq 0$

Classe 6: Modelo GARCH(1,1) - Hamilton Adaptado

$y_t = \mu + \phi \cdot y_{t-1} + u_t$

$u_t = \sqrt{g_{s_t}} \times \tilde{u}_t$

$\tilde{u}_t = h_t \cdot v_t$, onde v_t é i.i.d. N(0,1)

$h_t^2 = a_o + a_1 \cdot \tilde{u}_{t-1}^2 + b_1 \cdot h_{t-1}^2$; com $a_0 > 0, a_1 \geq 0, b_1 \geq 0$

Em todas as classes, S_t segue uma Cadeia de Markov com k estados, representando os k diferentes regimes, e Matriz de Transição $\mathbf{P}$.

Além disto, definiremos para cada uma das classes um modelo correspondente. Efetuaremos uma alteração na distribuição de u_t, passando-a de $N(0, h_t^2)$ para t-student, normalizada para variância unitária. Assim, teremos:

$$f(u_t|u_{t-1}, u_{t-2}, \ldots) = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi}\cdot\Gamma\left(\frac{\eta}{2}\right)}\cdot(\eta-2)^{-1/2}\cdot h_t^{-1}\cdot\left[1+\frac{u_t^2}{h_t^2\cdot(\eta-2)}\right]^{-\left(\frac{\eta+1}{2}\right)}$$

Neste caso, o número de graus de liberdade (η) da distribuição t- student também será incluído no vetor de parâmetros θ a ser estimado.

A fim de se verificar a adequação dos modelos acima descritos, utilizaremos como exemplo ajuste obtido dos mesmos sobre os retornos diários da ação Telebrás PN negociada na Bolsas de Valores de SP, calculados em regime de capitalização contínua através dos preços de fechamento, durante o período de 05/1/1993 e 06/08/1997. Os resultados obtidos serão comparados, tanto em nível de "goodness-of-fit" quanto de erros de previsão com o modelo GARCH (1,1) tradicionalmente utilizado na modelagem de volatilidades de séries financeiras.

VI - Ajuste dos Modelos SWARCH e SWGARCH sobre Telebrás PN

Até agora, nada foi dito a respeito do espaço de estados da Cadeia de Markov dos modelos acima expostos, o que caracteriza quantos regimes serão utilizados.

Na verdade, iremos ajustar os modelos da seguinte forma: para os modelos do tipo ARCH(1) e ARCH(2), trabalharemos com k = 2 e k = 3 (2 e 3 regimes), enquanto que para os modelos do tipo GARCH(1,1) somente consideraremos k = 2 (2 regimes). Logo, devemos ajustar a totalidade de 20 modelos, classificados segundo a Tabela 3.1 a seguir:

Tabela 3.1 - Modelos de Alternância de Regimes Propostos para os Retornos Diários de Telebrás.

Modelo	Regimes	Distribuição dos Resíduos	Parâmetros a Serem Estimados
ARCH(1) Cai	2	normal	ϕ, μ, γ_1, γ_2, a_1, p_{11}, p_{22}
ARCH(1) Cai	2	t-student	ϕ, μ, γ_1, γ_2, a_1, p_{11}, p_{22}, η
ARCH(1) Cai	3	normal	ϕ, μ, γ_1, γ_2, γ_3, a_1, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}
ARCH(1) Cai	3	t-student	ϕ, μ, γ_1, γ_2, γ_3, a_1, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}, η
ARCH(1) Hamilton	2	normal	ϕ, μ, g_2, a_0, a_1, p_{11}, p_{22}
ARCH(1) Hamilton	2	t-student	ϕ, μ, g_2, a_0, a_1, p_{11}, p_{22}, η
ARCH(1) Hamilton	3	normal	ϕ, μ, g_2, g_3, a_0, a_1, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}
ARCH(1) Hamilton	3	t-student	ϕ, μ, g_2, g_3, a_0, a_1, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}, η
ARCH(2) Cai	2	normal	ϕ, μ, γ_1, γ_2, a_1, a_2, p_{11}, p_{22}
ARCH(2) Cai	2	t-student	ϕ, μ, γ_1, γ_2, a_1, a_2, p_{11}, p_{22}, η
ARCH(2) Cai	3	normal	ϕ, μ, γ_1, γ_2, γ_3, a_1, a_2, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}
ARCH(2) Cai	3	t-student	ϕ, μ, γ_1, γ_2, γ_3, a_1, a_2, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}, η
ARCH(2) Hamilton	2	normal	ϕ, μ, g_2, a_0, a_1, a_2, p_{11}, p_{22}
ARCH(2) Hamilton	2	t-student	ϕ, μ, g_2, a_0, a_1, a_2, p_{11}, p_{22}, η
ARCH(2) Hamilton	3	normal	ϕ, μ, g_2, g_3, a_0, a_1, a_2, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}
ARCH(2) Hamilton	3	t-student	ϕ, μ, g_2, g_3, a_0, a_1, a_2, p_{11}, p_{12}, p_{21}, p_{22}, p_{32}, p_{33}, η
GARCH(1,1) Cai	2	normal	ϕ, μ, γ_1, γ_2, a_1, b_1, p_{11}, p_{22}
GARCH(1,1) Cai	2	t-student	ϕ, μ, γ_1, γ_2, a_1, b_1, p_{11}, p_{22}, η
GARCH(1,1) Hamilton	2	normal	ϕ, μ, g_2, a_0, a_1, b_1, p_{11}, p_{22}
GARCH(1,1) Hamilton	2	t-student	ϕ, μ, g_2, a_0, a_1, b_1, p_{11}, p_{22}, η

Vale lembrar ainda que temos as seguintes restrições em cada um dos modelos, segundo a Tabela 3.2 a seguir:

Tabela 3.2 - Restrições aos Parâmetros dos Modelos de Mudanças de Regimes Propostos

Parâmetros	Restrições
ϕ	$-1 < \phi < 1$
$\gamma_1, \gamma_2, \gamma_3$	$\gamma_1 > 0, \gamma_2 > 0, \gamma_3 > 0$
a_0, a_1, a_2	$a_0 > 0, a_1 \geq 0, a_2 \geq 0$
b_1	$b_1 \geq 0$
p_{ij}	$0 \leq p_{ij} \leq 1, \sum_{j=1}^{k} p_{ij} = 1$

Surge aqui um problema adicional: dependendo dos valores iniciais adotados para o processo de otimização numérica, é possível que o algoritmo convirja para um ponto de inadmissibilidade em alguns parâmetros. Por exemplo, nada nos garante que na solução final encontrada todas as probabilidades da matriz de transição estejam no intervalo [0,1]. A possibilidade de ocorrência de valores negativos para as probabilidades de transição não está descartada.

Para contornar este problema é necessário elaborar uma reparametrização da função de verossimilhança em termos de um vetor ω $(a \times 1)$ para o qual $\theta = g(\omega)$, sendo que a função $g: \Re^a \rightarrow \Re^a$ incorpora as restrições desejadas.

Para o caso das probabilidades de transição, a forma de se reparametrizar a função de verossimilhança é a seguinte:

Suponhamos que possuímos k probabilidades $p_1, p_2, \ldots, p_k$, as quais devem satisfazer às seguintes restrições:

$$p_1 + p_2 + \ldots + p_k = 1$$

$$0 \leq p_j \leq 1, \forall j = 1,2,\ldots,k$$

Assim, construímos $\omega_1, \omega_2, \ldots, \omega_k$ tais que:

$$p_j = \omega_j^2 / \left(1 + \omega_1^2 + \omega_2^2 + \ldots + \omega_{k-1}^2\right) \text{ para } j = 1, 2, \ldots, k\text{-}1.$$

$$p_j = 1 / \left(1 + \omega_1^2 + \omega_2^2 + \ldots + \omega_{k-1}^2\right) \text{ para } j = k.$$

Note que, no caso em que a matriz de transição é de ordem 2 (4 probabilidades de transição) somente 2 parâmetros ω são necessários. Isto provém do fato de que, nas matrizes de transição, a soma das colunas é igual a 1 e, desta forma, são necessárias k-1 probabilidades de transição para, em cada uma das colunas, se determinar as k probabilidades.

Sobre as restrições relativas à não negatividade dos parâmetros, a reparametrização é mais simples. Se ψ é um parâmetro necessariamente não-negativo, reparametrizaremos a função de verossimilhança em termos de ψ' através de $\psi = \sqrt{\psi'^2}$. Aplicamos esta transformação a todos os parâmetros que compõem a expressão da variância condicional.

Finalmente, para o coeficiente autorregressivo ϕ, o qual deve estar compreendido dentro do intervalo]-1,1[, a transformação em termos de ϕ' é dada por $\phi = \frac{\phi'}{1+|\phi'|}$

Isto posto, convém salientar que na obtenção das estimativas finais para os parâmetros dos modelos alternativos de mudanças de regimes propostos dois procedimentos foram adotados:

a) Modelos cuja distribuição dos resíduos é assumida normal

Estimação em três etapas:

1) Reparametrizamos a função de verossimilhança e otimizamos a função utilizando o algoritmo BHHH através do software RATS.

2) De posse dos valores obtidos na etapa acima, encontramos os valores originais dos parâmetros correspondentes e a função de verossimilhança, também em sua forma original, foi otimizada utilizando-se a rotina Solver embutida no MS-EXCEL 5.0, rotina esta que permite otimizar funções com restrições em seus parâmetros.

3) As estimativas obtidas na etapa anterior foram novamente alimentadas no RATS e a função de verossimilhança, agora em sua forma original, foi otimizada através do algoritmo BHHH, o qual fornece também os desvios-padrões associados às estimativas finais.

Vale também mencionar que esta estimação conduzida em três etapas foi também utilizada para se obter as estimativas finais do modelo de "spread" entre os títulos da dívida externa apresentado no capítulo anterior.

b) Modelos cuja distribuição dos resíduos é assumida t-student

Antes de mais nada é importante notar que para cada modelo em que a distribuição dos resíduos é assumida t-student, existe um modelo correspondente cuja distribuição dos resíduos é, por hipótese, normal. Portanto, a estimação dos parâmetros para este segundo conjunto de modelos foi conduzida da seguinte forma:

- Os valores iniciais foram obtidos observando-se as estimativas finais do modelo de distribuição normal associado;

- Utilizou-se novamente a rotina Solver do MS-EXCEL 5.0 para otimizar a função de verossimilhança na sua forma não reparametrizada, obtendo assim, as estimativas finais dos parâmetros;

- Os desvios-padrões associados foram obtidos numericamente através do método do "Score";

Feitas estas considerações, apresentamos a seguir os resultados finais obtidos, conforme as Tabelas 3.3, 3.4, 3.5 e 3.6 a seguir:

Tabela 3.3 - Critérios de Seleção para os Modelos Alternativos

Modelo	Regimes	Dist. dos Resíduos	Log Verossim.	Número de Parâmetros	AIC	SC
ARCH(1) Cai	2	normal	2312.430	7	-4610.86	-4575.67
ARCH(1) Cai	2	t-student	2321.536	8	-4627.07	-4586.85
ARCH(1) Cai	3	normal	2339.139	12	-4654.28	-4593.95
ARCH(1) Cai	3	t-student	2341.176	13	-4656.35	-4591.00
ARCH(1) Hamilton	2	normal	2312.712	7	-4611.42	-4576.23
ARCH(1) Hamilton	2	t-student	2321.590	8	-4627.18	-4586.96
ARCH(1) Hamilton	3	normal	2339.141	12	-4654.28	-4593.95
ARCH(1) Hamilton	3	t-student	2341.178	13	-4656.36	-4591.00
ARCH(2) Cai	2	normal	2328.631	8	-4641.26	-4601.04
ARCH(2) Cai	2	t-student	2336.398	9	-4654.80	-4609.55
ARCH(2) Cai	3	normal	2341.005	13	-4656.01	-4590.65
ARCH(2) Cai	3	t-student	2344.137	14	-4660.27	-4589.89
ARCH(2) Hamilton	2	normal	2328.959	8	-4641.92	-4601.70
ARCH(2) Hamilton	2	t-student	2336.541	9	-4655.08	-4609.84
ARCH(2) Hamilton	3	normal	2340.847	13	-4655.69	-4590.34
ARCH(2) Hamilton	3	t-student	2344.104	14	-4660.21	-4589.83
GARCH(1,1) Cai	2	normal	2337.703	8	-4659.41	-4619.19
GARCH(1,1) Cai	2	t-student	2346.976	9	**-4675.95**	-4630.71
GARCH(1,1) Hamilton	2	normal	2336.799	8	-4657.60	-4617.38
GARCH(1,1) Hamilton	2	t-student	2345.865	9	-4673.73	-4628.48
GARCH (1,1)	-	normal	2326.106	5	-4642.21	-4617.08
GARCH (1,1)	-	t-student	2343.155	6	-4674.31	**-4644.15**

Tabela 3.4 - Estimativas dos Parâmetros para os Modelos SWARCH (1) Alternativos

	Cai 2 Regimes Normal	Cai 2 Regimes t-student	Cai 3 Regimes Normal	Cai 3 Regimes t-student	Hamilton 2 Regimes Normal	Hamilton 2 Regimes t-student	Hamilton 3 Regimes Normal	Hamilton 3 Regimes t-student
μ	3.35×10^{-3} (8.05×10^{-4})	3.17×10^{-3} (8.03×10^{-4})	3.16×10^{-3} (7.75×10^{-4})	3.05×10^{-3} (7.74×10^{-4})	3.37×10^{-3} (8.01×10^{-4})	3.17×10^{-3} (8.03×10^{-4})	3.15×10^{-3} (7.75×10^{-4})	3.05×10^{-3} (7.71×10^{-4})
ϕ	6.17×10^{-3} (3.23×10^{-2})	1.43×10^{-3} (3.33×10^{-2})	4.52×10^{-3} (3.22×10^{-2})	8.31×10^{-3} (3.10×10^{-2})	5.79×10^{-3} (3.22×10^{-2})	1.42×10^{-3} (3.33×10^{-2})	4.25×10^{-3} (3.22×10^{-2})	5.00×10^{-3} (3.09×10^{-2})
γ_0	3.13×10^{-4} (2.57×10^{-5})	3.48×10^{-4} (3.40×10^{-5})	3.01×10^{-4} (2.75×10^{-5})	3.03×10^{-4} (2.95×10^{-5})	n.d.	n.d.	n.d.	n.d.
γ_1	1.47×10^{-3} (1.06×10^{-4})	1.52×10^{-3} (1.43×10^{-4})	1.05×10^{-3} (9.20×10^{-5})	1.02×10^{-3} (9.82×10^{-5})	n.d.	n.d.	n.d.	n.d.
γ_2	n.d.	n.d.	4.35×10^{-3} (5.87×10^{-4})	3.82×10^{-3} (6.12×10^{-4})	n.d.	n.d.	n.d.	n.d.
a_0	n.d.	n.d.	n.d.	n.d.	3.12×10^{-4} (2.57×10^{-5})	3.48×10^{-4} (3.40×10^{-5})	3.01×10^{-4} (2.75×10^{-5})	3.04×10^{-4} (3.10×10^{-5})
a_1	0.2259 (0.0319)	0.1937 (0.0483)	0.0186 (0.0380)	0.0135 (0.0381)	0.2283 (0.0322)	0.1953 (0.0487)	0.0193 (0.0387)	0.0108 (0.0373)
g_2	n.d.	n.d.	n.d.	n.d.	4.7062 (0.4844)	4.3704 (0.5284)	3.4832 (0.3786)	3.3673 (0.3966)
g_3	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	14.4109 (2.1084)	12.6185 (2.2250)
p_{11}	0.9882 (0.0340)	0.9920 (0.0328)	0.9857 (0.0485)	0.9874 (0.0473)	0.9878 (0.0341)	0.9920 (0.0328)	0.9857 (0.0485)	0.9874 (0.0491)
p_{12}	-	-	0.0085 (0.0085)	0.0084 (0.0093)	-	-	0.0086 (0.0085)	0.0084 (0.0092)
p_{21}	-	-	0.0088 (0.0071)	0.0075 (0.0064)	-	-	0.0088 (0.0071)	0.0075 (0.0068)
p_{22}	0.9921 (0.0321)	0.9947 (0.0319)	0.9838 (0.0379)	0.9850 (0.0381)	0.9918 (0.0322)	0.9947 (0.0319)	0.9838 (0.0378)	0.9850 (0.0379)
p_{32}	n.d.	n.d.	0.0404 (0.0289)	0.0307 (0.0258)	n.d.	n.d.	0.0413 (0.0288)	0.0308 (0.0257)
p_{33}	n.d.	n.d.	0.9596 (0.0548)	0.9693 (0.0503)	n.d.	n.d.	0.9583 (0.0547)	0.9692 (0.0511)
η	n.d.	10.1197 (2.5203)	n.d.	22.0829 (8.1937)	n.d.	10.1724 (2.5180)	n.d.	22.0571 (8.1854)

Tabela 3.5 - Estimativas dos Parâmetros para os Modelos SWARCH (2) Alternativos

	Cai 2 Regimes Normal	Cai 2 Regimes t-student	Cai 3 Regimes Normal	Cai 3 Regimes t-student	Hamilton 2 Regimes Normal	Hamilton 2 Regimes t-student	Hamilton 3 Regimes Normal	Hamilton 3 Regimes t-student
μ	3.62×10^{-3} (8.31×10^{-4})	3.46×10^{-3} (8.03×10^{-4})	3.30×10^{-3} (7.76×10^{-4})	3.22×10^{-3} (7.71×10^{-4})	3.64×10^{-3} (8.28×10^{-4})	3.46×10^{-3} (8.02×10^{-4})	3.29×10^{-3} (7.74×10^{-4})	3.22×10^{-3} (7.76×10^{-4})
ϕ	3.10×10^{-3} (3.16×10^{-2})	1.30×10^{-3} (3.17×10^{-2})	2.13×10^{-3} (3.18×10^{-2})	2.29×10^{-3} (3.07×10^{-2})	3.09×10^{-3} (3.17×10^{-2})	1.29×10^{-2} (3.17×10^{-2})	-5.48×10^{-4} (3.18×10^{-2})	9.38×10^{-3} (3.07×10^{-2})
γ_0	2.99×10^{-4} (2.74×10^{-5})	2.98×10^{-4} (3.44×10^{-5})	2.76×10^{-4} (2.90×10^{-5})	2.70×10^{-4} (3.40×10^{-5})	n.d.	n.d.	n.d.	n.d.
γ_1	1.19×10^{-3} (9.80×10^{-5})	1.15×10^{-3} (1.18×10^{-4})	9.74×10^{-4} (1.06×10^{-4})	7.89×10^{-4} (9.78×10^{-5})	n.d.	n.d.	n.d.	n.d.
γ_2	n.d.	n.d.	3.70×10^{-3} (6.73×10^{-4})	2.42×10^{-3} (3.97×10^{-4})	n.d.	n.d.	n.d.	n.d.
a_0	n.d.	n.d.	n.d.	n.d.	2.98×10^{-4} (2.75×10^{-5})	2.99×10^{-4} (3.44×10^{-5})	2.74×10^{-4} (2.91×10^{-5})	2.71×10^{-4} (3.40×10^{-5})
a_1	0.1799 (0.0361)	0.1504 (0.0460)	0.0227 (0.0390)	0.0364 (0.0387)	0.1806 (0.0364)	0.1503 (0.0460)	0.0272 (0.0392)	0.0385 (0.0392)
a_2	0.1761 (0.0439)	0.2066 (0.0558)	0.0734 (0.0432)	0.1131 (0.0477)	0.1759 (0.0443)	0.2059 (0.0560)	0.0721 (0.0436)	0.1127 (0.0481)
g_2	n.d.	n.d.	n.d.	n.d.	4.0116 (0.4693)	3.8558 (0.5134)	3.5094 (0.4249)	2.8904 (0.4303)
g_3	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	12.8607 (2.3346)	8.8143 (1.6981)
p_{11}	0.9950 (0.0029)	0.9965 (0.0030)	0.9851 (0.0336)	0.9917 (0.0500)	0.9949 (0.0030)	0.9964 (0.0030)	0.9851 (0.0489)	0.9919 (0.0501)
p_{12}	-	-	0.0096 (0.0092)	0.0052 (0.0081)	-	-	0.0101 (0.0096)	0.0050 (0.0080)
p_{21}	-	-	0.0091 (0.0074)	0.0049 (0.0063)	-	-	0.0091 (0.0075)	0.0048 (0.0062)
p_{22}	0.9970 (0.0021)	0.9983 (0.0016)	0.9829 (0.0306)	0.9858 (0.0378)	0.9969 (0.0021)	0.9982 (0.0016)	0.9832 (0.0382)	0.9859 (0.0378)
p_{32}	n.d.	n.d.	0.0391 (0.0325)	0.0173 (0.0183)	n.d.	n.d.	0.0340 (0.0290)	0.0168 (0.0175)
p_{33}	n.d.	n.d.	0.9608 (0.0515)	0.9827 (0.0426)	n.d.	n.d.	0.9659 (0.0512)	0.9832 (0.0425)
η	n.d.	10.0148 (3.0988)	n.d.	15.4001 (4.9196)	n.d.	11.0864 (3.0575)	n.d.	15.2838 (4.9475)

Tabela 3.6 - Estimativas dos Parâmetros para os Modelos SWGARCH (1,1) Alternativos

	Cai 2 Regimes normal	Cai 2 Regimes t-student	Hamilton 2 Regimes normal	Hamilton 2 Regimes t-student	Garch normal	Garch t-student
μ	3.041×10^{-3} (8.319×10^{-4})	3.017×10^{-3} (7.971×10^{-4})	3.127×10^{-3} (8.352×10^{-4})	3.036×10^{-3} (7.910×10^{-4})	2.813×10^{-3} (8.550×10^{-4})	2.978×10^{-3} (7.817×10^{-4})
ϕ	1.179×10^{-3} (3.203×10^{-2})	1.263×10^{-3} (3.230×10^{-2})	1.006×10^{-3} (3.288×10^{-2})	1.185×10^{-3} (3.217×10^{-2})	1.251×10^{-3} (3.041×10^{-2})	1.244×10^{-3} (3.113×10^{-2})
γ_1	3.732×10^{-5} (1.176×10^{-5})	2.872×10^{-5} (1.164×10^{-5})	n.d.	n.d.	n.d.	n.d.
γ_2	1.413×10^{-4} (4.101×10^{-5})	1.010×10^{-4} (3.987×10^{-5})	n.d.	n.d.	n.d.	n.d.
g_2	n.d.	n.d.	3.2515 (0.4416)	2.6643 (0.6077)	n.d.	n.d.
a_0	n.d.	n.d.	4.739×10^{-5} (1.574×10^{-5})	2.889×10^{-5} (1.150×10^{-5})	1.803×10^{-5} (5.842×10^{-6})	1.579×10^{-5} (7.052×10^{-6})
a_1	0.1288 (0.0213)	0.1253 (0.0275)	0.1363 (0.0256)	0.1276 (0.0288)	0.1181 (0.0144)	0.1262 (0.0236)
b_1	0.7962 (0.0402)	0.8201 (0.0418)	0.7735 (0.0512)	0.8227 (0.0408)	0.8724 (0.0149)	0.8677 (0.0227)
p_{11}	0.9976 (0.0252)	0.9983 (0.0281)	0.9954 (0.0314)	0.9983 (0.0373)	n.d.	n.d.
p_{22}	0.9986 (0.0172)	0.9989 (0.0207)	0.9979 (0.0152)	0.9990 (0.0210)	n.d.	n.d.
η	n.d.	10.5395 (2.8722)	n.d.	10.4306 (2.6715)	n.d.	8.9124 (1.8315)

Tabela 3.7 - Teste LR para os Modelos SWGARCH (1,1) Alternativos

Modelo	Log-Verossimilhança	Teste LR	Valor-p
GARCH(1,1) t-student	2343.155	-	-
SWGARCH (1,1) Cai - t-student	2346.976	7.642	0.054
SWGARCH (1,1) Hamilton - t-student	2345.865	5.420	0.144

À luz dos resultados obtidos, certas considerações merecem ser feitas:

1) Em todas as 6 classes, os modelos que assumem distribuição t-student para os resíduos ofereceram diferenças significativas em relação aos modelos que assumem distribuição normal. Isto é uma evidência de que a existência de caudas pesadas na distribuição dos retornos diários de Telebrás está também associada a possíveis diferentes distribuições aplicadas aos resíduos, além das alterações bruscas na variância condicional da série.

Focalizando apenas os modelos das classes ARCH(1) e ARCH(2), notamos que estes últimos produziram ajustes melhores aos dados, entretanto nos modelos ARCH(2), a passagem de 2 para 3 regimes não provocou melhoras segundo o critério SC. Além disto, como mostra a Tabela 3.5, os desvios-padrões associados às probabilidades de transição que envolvem o terceiro regime são elevadas em relação às estimativas dos parâmetros.

Ainda sobre os modelos que apresentam mudanças de regimes, as classes GARCH(1,1) foram as que apresentaram melhores resultados tanto no critério AIC quanto no SC. Porém, não foi possível a identificação de 3 regimes nestes modelos.

Por último, considerando-se a totalidade dos modelos ajustados, temos que, pelo critério AIC, o modelo mais apropriado seria o GARCH(1,1)-Cai-t-student, ficando o modelo GARCH(1,1)-t-student como segunda opção. Esta situação se inverte quando o critério utilizado é o SC. Já o modelo GARCH(1,1)-Hamilton-t-student fica como terceira opção em ambos os critérios. Entretanto, o teste da razão das verossimilhanças, não rejeitou a hipótese nula de que o processo gerador dos dados fosse um GARCH(1,1) t-student sem mudanças de regimes. Muito embora, conforme veremos mais adiante, nossa escolha final recairá sobre os modelos que apresentam mudanças de regimes devido ao menor erro de predição que apresentam, comparativamente aos modelos GARCH(1,1) usuais.

2) Como era de se esperar, as mudanças de regimes novamente não ocorrem com muita freqüência. As figuras 3.1 e 3.2 a seguir exibem a série dos retornos diários de Telebrás e as probabilidades suavizadas de nos encontrarmos em cada um dos dois regimes diferentes, associadas ao modelo GARCH(1,1)-Cai-t-student. Partindo-se do pressuposto de que um regime é predominante sobre os demais num instante t quando sua probabilidade suavizada neste instante supera 50%, temos que durante toda a evolução da série ocorreu apenas 1 mudança de regime, em julho de 1995. Se o critério for considerar os cruzamentos na linha de 50% da probabilidade filtrada, identificamos 7 mudanças de regimes.

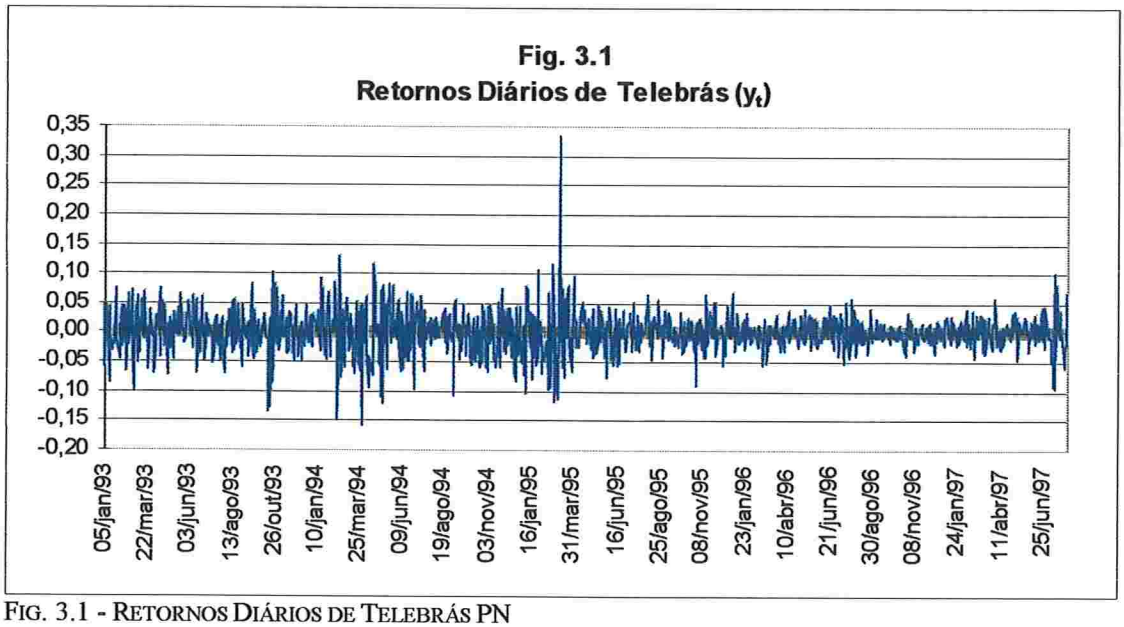

FIG. 3.1 - RETORNOS DIÁRIOS DE TELEBRÁS PN

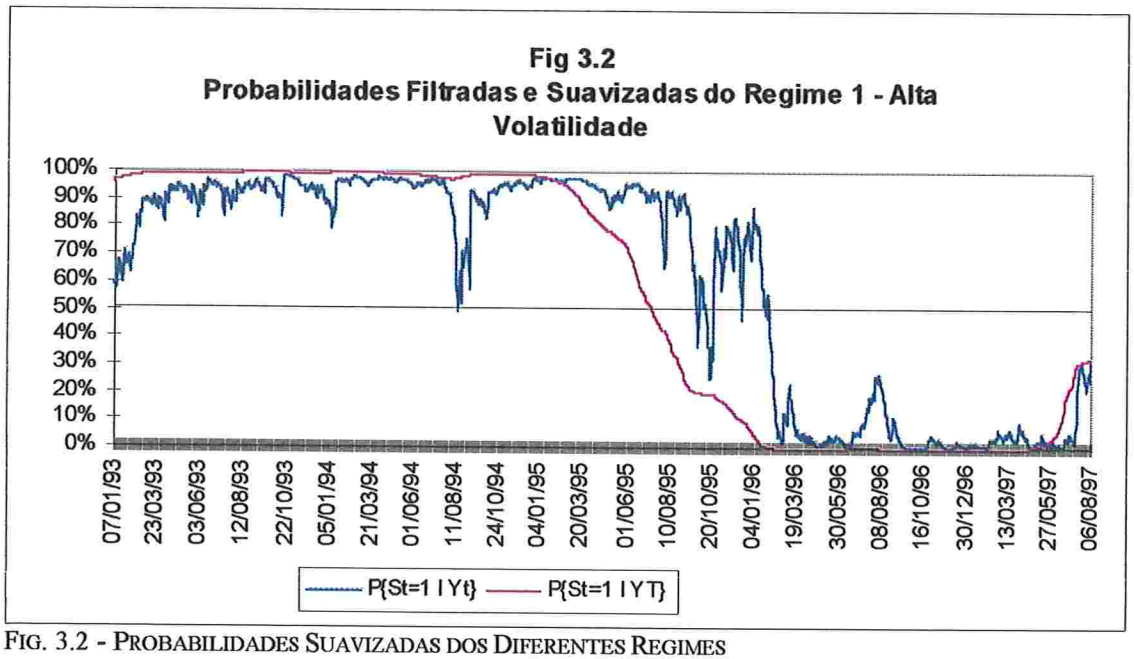

FIG. 3.2 - PROBABILIDADES SUAVIZADAS DOS DIFERENTES REGIMES

3) A assimetria também se verifica para o caso de Telebrás. Se olharmos para as estimativas referentes às probabilidades de transição da Cadeia de Markov, vemos que os regimes possuem uma permanência bastante elevada: a probabilidade de permanecermos no regime 1 no instante t dado que no instante t-1 estávamos neste mesmo regime é de 99.83%. A probabilidade de permanecermos no regime 2 no instante t dado que no instante t-1 estávamos neste mesmo regime é de 99.89%. Assim, o tempo médio de permanência em cada um dos regimes é:

$$T_1 = \frac{1}{1-p_{11}} = \frac{1}{1-0.9983} \cong 589 \text{ dias úteis}$$

$$T_2 = \frac{1}{1-p_{22}} = \frac{1}{1-0.9989} \cong 870 \text{ dias úteis}$$

Adicionalmente, as probabilidades não condicionais de nos encontrarmos em cada um dos regimes, calculada como sendo a medida invariante da Cadeia de Markov são dadas por:

$\pi_1 = 0.404$ $\qquad$ $\pi_2 = 0.596$

Daí conclui-se que o regime 2 (alta volatilidade) costuma ser o mais freqüente.

4) Por fim, o modelo de alternância de regimes reduz o erro de previsão quando comparado com o modelo GARCH(1,1). Porem, antes de chegarmos a este resultado, necessitamos desenvolver as equações de projeção para o modelo de alternância de regimes, o que será

feito nas próximas páginas. Vale lembrar que para o caso do modelo GARCH(1,1) a equação de predição já foi desenvolvia na seção II.2 deste capítulo.

Queremos, então calcular:

$$E\left[u_{t+1}^2 | Y_t\right] = \sum_{i=1}^{2}\sum_{j=1}^{2} E\left[u_{t+1}^2 | s_{t+1} = i, s_t = j, Y_t\right] \times P\left\{s_{t+1} = i, s_t = j, Y_t\right\} =$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} E\left[u_{t+1}^2 | s_{t+1} = i, s_t = j, Y_t\right] \times P\left\{s_t = j | Y_t\right\} \times p_{ji}$$

Para o modelo Cai - Adaptado, temos:

$E\left[u_{t+1}^2 | s_{t+1} = i, s_t = j, Y_t\right] = E\left[h_{t+1}^2 \cdot v_{t+1}^2 | s_{t+1} = i, s_t = j, Y_t\right] = h_{t+1}^{2(i,j)}$, devido à independência de v_t.

Para o modelo Hamilton - Adaptado, temos:

$E\left[u_{t+1}^2 | s_{t+1} = i, s_t = j, Y_t\right] = E\left[g_i \cdot h_{t+1}^2 \cdot v_{t+1}^2 | s_{t+1} = i, s_t = j, Y_t\right] = g_i \cdot h_{t+1}^{2(j)}$, pelo mesmo motivo.

Resta-nos, agora, aplicar as expressões acima sobre os dados amostrais e avaliar os erros associados para cada um dos dois modelos alternativos, sob algum critério de perda.

Até agora, viemos utilizando a função de perda quadrática, ou seja, o erro quadrático médio foi obtido segundo a expressão:

$$MSE = \frac{\sum_{t=1}^{T}\left(\hat{u}_t^2 - \sigma_t^2\right)^2}{T}$$, onde σ^2_t é a estimativa de u^2_t para o instante t.

Porém, o erro quadrático médio pressupõe a existência do quarto momento para y_t, o que muitas vezes podem não existir para os modelos do tipo ARCH e GARCH, dependendo dos valores dos parâmetros.

Assim, para compararmos as performances dos modelos em termos dos erros de predição, utilizaremos além da função de perda quadrática outras variantes para a função-perda:

$$MAE = \frac{\sum_{t=1}^{T}\left|\left(\hat{u}_t^2 - \sigma_t^2\right)\right|}{T}$$

$$[LE]^2 = \frac{\sum_{t=1}^{T}\left(ln\left(\hat{u}_t^2\right) - ln\left(\sigma_t^2\right)\right)^2}{T}$$

$$|LE| = \frac{\sum_{t=1}^{T}\left|\left(ln\left(\hat{u}_t^2\right) - ln\left(\sigma_t^2\right)\right)\right|}{T}$$

A Tabela 3.8 a seguir nos fornece os valores das quatro funções-perda associadas aos erros de predição a um passo para cada um dos modelos alternativos:

Tabela 3.8 - Comparativo entre os Erros de Projeção dos Modelos para Telebrás PN

	MSE	MAE	$[LE]^2$	| LE |
GARCH(1,1) t-student	1.51 x 10-5	1.34 x 10-3	6.75	1.84
GARCH(1,1) Hamilton t-student	1.47 x 10-5	1.31 x 10-3	6.49	1.83
% de melhora	+ 2.7 %	+ 2.5 %	+ 3.9 %	+ 0.4 %
GARCH(1,1) Cai t-student	1.47 x 10-5	1.31 x 10-3	6.52	1.83
% de melhora	+ 3.0 %	+ 2.5 %	+ 3.5 %	+ 0.4 %

Pelos resultados obtidos, temos que os modelos de alternância de regimes conseguem reduzir em até 3.9 % o erro médio de previsão comparativamente ao modelo GARCH(1,1). A razão desta melhor performance do modelo de mudança de regimes está relacionada com a habilidade que este possui em ajustar-se mais rapidamente às variações bruscas na volatilidade condicional. De fato, devido à persistência elevada dos modelos GARCH(1,1) e a não flexibilização dos parâmetros que compõem a expressão da variância condicional, seu ajuste é mais gradual à presença de choques extremos. Enquanto isto, os modelos de alternância de regimes rapidamente se ajustam ao novo patamar de volatilidade, justamente pela alteração dos parâmetros associados à especificação da variância condicional, embora nestes a persistência também seja bastante elevada, porém menor que a do modelo GARCH(1,1), conforme observamos na Tabela 3.9 a seguir.

Tabela 3.9 - Comparativo entre a persistência (δ) dos modelos

	a_1	b_1	δ
GARCH(1,1) t-student	0.1262	0.8677	0.9939
GARCH(1,1) Hamilton t-student	0.1276	0.8227	0.9503
GARCH(1,1) Cai t-student	0.1253	0.8201	0.9454

NOTAS CONCLUSIVAS

O presente trabalho objetivou apresentar como os modelos de alternância markovianas de regimes podem ser aplicados em séries temporais financeiras. No capítulo 1 lançamos algumas idéias básicas que estão por detrás da metodologia dos modelos de mudanças de regimes e apresentamos, naquele capítulo, a versão mais simplificada desta abordagem, conhecida como "Mistura de Distribuições I. I. D." Aplicando-se a técnica em questão sobre a séries histórica dos retornos do Ibovespa disponível concluímos que, mesmo na sua forma mais simples, os modelos de alternância de regimes já oferecem uma melhor descrição do comportamento da série.

No capítulo 2, relaxamos algumas hipóteses restritivas que se faziam presentes no capítulo anterior:

1) Substituímos a independência entre os regimes pela hipótese de que os regimes sejam governados por uma Cadeia de Markov;

2) Permitimos dinâmicas autorregressivas na série temporal;

Desta maneira, características importantes das séries financeiras puderam ser tratadas pela abordagem hamiltoniana, tais como assimetria e movimentações em tendências. Para tanto, desenvolvemos o arcabouço teórico dos modelos de mudanças markovianas de regimes, introduzindo conceitos como as probabilidades filtradas e suavizadas, além dos diferentes métodos de obtenção dos estimadores de máxima verossimilhança dos parâmetros. Aplicamos a metodologia sobre as séries dos títulos de 30 anos do tesouro dos EUA e o "spread" entre dois dos principais títulos da dívida externa brasileira, o C-bond e o Par-bond, obtendo resultados mais satisfatórios que a técnica tradicional de séries temporais, regida sob a hipótese de linearidade da série, sejam em termos de "goodness-of-fit", seja em termos de erros quadrátricos médios de previsão.

Finalmente, no capítulo 3, mostramos como o modelo de mudanças markovianas de regimes pode ser também utilizado em modelos condicionalmente heterocedásticos. Nossa preocupação foi a de gerar modelos de previsão de volatilidade melhores que o largamente utilizado GARCH (1,1). O ponto central de análise foi, partindo-se dos modelos SWARCH desenvolvidos por Hamilton & Susmel (1994) e Cai (1994), incorporar, seguindo na forma proposta por Dueker (1997), na especificação da variância condicional, o seu valor no instante anterior. Desta forma construímos os modelos SWGARCH, que foram aplicados sobre a série dos retornos diários de Telebrás PN. O desempenho dos modelos foram comparados utilizando-se quatro diferentes funções-perda para o erro de predição a um passo. Os modelos SWGARCH, neste aspecto, conseguiram reduzir o erro de predição.

De maneira geral, concluímos que a abordagem hamiltoniana oferece melhor explicação ao comportamento de séries temporais financeiras em relação aos modelos tradicionais lineares e mesmo até em relação a alguns modelos não lineares, conforme vimos no capítulo 3.

Ressalta-se que esta habilidade, como vimos, associada ao fato de que os modelos de mudanças de regimes conseguem tratar de maneira mais adequeda características peculiares a séries financeiras, que não são normalmente verificadas em outras áreas de pesquisa.

BIBLIOGRAFIA

Black, F. e Scholes, M. (1973), "*The Pricing of Options and Corporate Liabilities*", Journal of Political Economy, 81, 637-959.

Bollerslev, T. (1986), "*Generalized Autoregressive Conditional Heterokedasticity*", Journal of Econometrics, 31, 307-327.

Box, G. E. P. e Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, Revised Edition, São Francisco: Holden Day.

Cai, J. (1994), "*A Markov Model of Switching-Regime ARCH*", Journal of Business & Economic Statistics, 12, 306-316.

Campbell, J. Y., Lo, A. W. e MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Nova Jersey: Princeton University Press.

Correia, M. M. R. da L. (1997), *Memória Longa, Agrupamento de Valores Extremos e Assimetrias em Séries Financeiras*, Dissertação de Mestrado, FEA-USP, São Paulo.

Dueker, M. J. (1997), *Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility*, Journal of Business & Economic Statistics, 15, 26-34.

Engel, C. e Hamilton, J. D. (1990), "*Long Swings in the Dollar: Are They in the Data and Do Markets Knows It ?*", American Economic Reveiw, 80, 689-713.

Engle, R. F. (1982), "*Autoregressive Conditional Heterokedasticity with Estimates of the Variance of U.K. Inflation*", Econometrica, 50, 987-1008.

Kim, C. J. (1994), "*Dynamic Linear Models With Markov Switching*", Journal of Econometrics, 64, 1-22.

Hamilton, J. D. (1989), "*A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*", Econometrica, 57, 357-384.

Hamilton, J. D. (1990), "*Analysis of Time Series Subject to Changes in Regime*", Journal of Econometrics, 45, 39-70.

Hamilton, J. D. (1993), "*Estimation, Inference and Forecasting of Time Series Subject to Changes in Regime*", in Maddala, G. S., Rao, C. R. e Vinod, H. D., eds., Handbook of Statistics, Vol. 11, Nova York: North-Holland.

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

Hamilton, J. D. e Susmel, R. (1994), "*Autoregressive Conditional Heterokedasticity and Changes in Regime*", Journal of Econometrics, 64, 307-333.

Hull, J. C. (1997), *Options, Futures and Other Derivatives*, 3rd Ed., Nova Jersey: Prentice-Hall.