Universidade Federal do Rio de Janeiro

Investimentos no Mercado de Petróleo: Uma Abordagem Utilizando Modelos de Markov Ocultos

Edmundo Grüne de Souza e Silva

2009

COPPE/UFRJ

INVESTIMENTOS NO MERCADO DE PETRÓLEO: UMA ABORDAGEM UTILIZANDO MODELOS DE MARKOV OCULTOS

Edmundo Grüne de Souza e Silva

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Planejamento Energético, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Planejamento Energético.

Orientador: Luiz Fernando Loureiro Legey

Rio de Janeiro

Março de 2009

INVESTIMENTOS NO MERCADO DE PETRÓLEO: UMA ABORDAGEM UTILIZANDO MODELOS DE MARKOV OCULTOS

Edmundo Grüne de Souza e Silva

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM PLANEJAMENTO ENERGÉTICO.

Aprovada por:

Prof. Luiz Fernando Loureiro Legey, Ph.D.

Prof. Lucio Guido Tapia Carpio, D.Sc.

Dr. Mario Veiga Ferraz Pereira, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2009

Souza e Silva, Edmundo Grüne de

Investimentos no Mercado de Petróleo: Uma Abordagem Utilizando Modelos de Markov Ocultos/ Edmundo Grüne de Souza e Silva. – Rio de Janeiro: UFRJ/COPPE, 2009

XXV, 154 p.:il; 29,7 cm

Orientador: Luiz Fernando Loureiro Legey

Dissertação (mestrado) - UFRJ/COPPE/Programa de Planejamento Energético, 2009

Referência Bibliográficas: p. 150-154

1. Preço do Petróleo. 2. Estratégias de Investimento. 3. Previsão de Séries Temporais. 4. Modelos de Markov Ocultos. 5. Wavelets. I. Legey, Luiz Fernando Loureiro II.Universidade Federal do Rio de Janeiro, COPPE, Programa de Planejamento Energético. III. Titulo.

Agradecimentos

Em primeiríssimo lugar, gostaria de agradecer a três grandes professores e amigos, sem os quais este trabalho jamais teria sido realizado, e cujos ensinamentos levarei por toda vida. São eles: meu orientador, Prof. Legey, que sempre me recebeu com enorme entusiasmo em sua sala. A cada conversa, me sentia mais empolgado para continuar seguindo em frente, desbravando novos caminhos, olhando um pouco mais além, montado sobre ombros de gigantes. Meu pai e co-orientador, Prof. Edmundo, que teve participação fundamental na elaboração deste trabalho, e que, desde pequeno, insistia (muitas vezes exaustivamente) em me dizer que conhecimento é o maior bem que uma pessoa pode ter. E, finalmente, Profa. Eloá, que me mostrou não haver qualquer distinção entre a matemática e a língua. Graças a ela, e afirmo isso com toda certeza que alguém pode ter, hoje sei expressar o que sinto e o que penso; sou capaz de, como Clarice e Cecília, transmitir minhas mais loucas idéias; e tenho a sorte de conseguir me comunicar. A todos eles, devo muito mais do que estas simples palavras aqui escritas.

Agradeço também à minha mãe, Ursula, pelo amor incondicional que tem me dado ao longo de toda minha vida, e pela paciência que teve em me criar, o que, com certeza, não foi uma tarefa simples. Sim, eu sei que, às vezes, sou muito chato!

Aos meus irmãos, Christina, Mario e Carolina, especialmente pelo companheirismo nos dias de surfe, os quais, definitivamente, foram palco das minhas melhores idéias! E, também, é claro, à minha família: minha avó, meus tios e tias, primos e primas, por sempre me proporcionarem maravilhosos momentos quando estamos juntos.

À minha namorada e cúmplice, Cris, por ter transformado a minha vida, e ter me mostrado que não existe o impossível. Com ela descobri que contos de fadas não se restringem aos livros, mas fazem parte da vida, se permitirmos que aconteçam.

Aos meus amigos do LAND; especialmente ao Fabrício, Luiz, Bernardo e GD, por estarem sempre dispostos a me ajudar e tirar minhas dúvidas sobre o mundo Linux; Hugo, Carolzinha, David e Ebenézer pelo companheirismo e horas de estudo durante toda a faculdade; Flávio, Guto, Ana, Sadoc e Fernando pela paciência, lá em 2002 ainda, em me ensinar a programar (ah, e Fernando pelo **mtk** também!); e, é claro, a CALÓL, CALÓL, CALÓL, sempre feliz e disposta a ajudar. Com certeza, ela é o coração deste grupo!

Por fim, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, pelo financiamento deste estudo através da bolsa de mestrado, e ao Programa de Planejamento Energético da COPPE, por ter me aceitado como aluno.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

INVESTIMENTOS NO MERCADO DE PETRÓLEO: UMA ABORDAGEM UTILIZANDO MODELOS DE MARKOV OCULTOS

Edmundo Grüne de Souza e Silva

Março/2009

Orientador: Luiz Fernando Loureiro Legey

Programa: Planejamento Energético

O preço do petróleo é influenciado por uma enorme quantidade de fatores, cujas características, muitas vezes extremamente complexas, tornam sua previsão por fundamentos uma tarefa bastante complicada. Uma forma alternativa de se tentar fazer esta previsão é através de modelos de séries temporais, os quais, observando apenas seu comportamento passado, tentam inferir seu comportamento futuro. Nesta dissertação, investigamos a capacidade de um modelo de séries temporais não linear, conhecido como Modelo de Markov Oculto (HMM), em prever movimentos futuros do preço do petróleo. Tomando-o como base, elaboramos uma metodologia de previsão que consiste em, basicamente, duas etapas. Na primeira, devido a intensa volatilidade intrínseca, suavizamos a série de preços utilizando Wavelets. Em seguida, utilizamos o HMM para prever a distribuição probabilística da variação acumulada pelo preço, ao final de uma janela futura de F dias. A partir desta distribuição, inferimos a direção futura do preço. Nossos resultados sugerem que o Modelo de Markov Oculto pode ser uma interessante ferramenta de auxílio a tomada de decisão de investimentos por participantes do mercado de petróleo.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

INVESTMENT IN THE CRUDE OIL MARKET: AN APPROACH USING HIDDEN MARKOV MODELS

Edmundo Grüne de Souza e Silva

March/2009

Advisor: Luiz Fernando Loureiro Legey

Department: Energy Planning

The price of crude oil is determined by a large quantity of factors, most of which have very complex characteristics. For this reason, its forecasting by fundamentals is a very difficult task. An alternative way to build such a forecast is to use a time series approach, in which its past behavior is observed and used to predict future movements. In this thesis, we investigate the capability of a nonlinear time series model, know as hidden Markov model (HMM), to predict future crude oil price movements. Using a HMM, we develop a forecasting methodology that consists of, basically, two steps. First, due to the high volatility observed, we use Wavelets to smooth the price series, removing its noise. Then, the HMM is used to forecast the probability distribution of the price return at the end of the next F days. Finally, from this distribution, we guess the future direction of prices. Our results indicate that the hidden Markov model might be an interesting auxiliary decision tool for the participants of the crude oil market.

Glossário

ANN	:	Rede neural artificial (*Artificial Neural Network*).
AR	:	Modelo auto-regressivo (*Autoregressive model*).
ARIMA	:	Modelo auto-regressivo integrado de médias móveis (*Autoregressive Integrated Moving Average model*).
ARMA	:	Modelo auto-regressivo e de médias móveis (*Autoregressive Moving Average model*).
CCDF	:	Função de distribuição acumulada complementar (*Complementary Cumulative Distribution Function*).
CWT	:	Transformada contínua de wavelets (*Continuous Wavelet Transform*).
DWT	:	Transformada discreta de wavelets (*Discrete Wavelet Transform*).
EM	:	Algoritmo utilizado na estimação de parâmetros de um HMM (*Expectation-Maximization algorithm*).
FWT	:	Transformada rápida de wavelets (*Fast Wavelet Transform*).
GARCH	:	(*Generalized Autoregressive Conditional Heterocedasticity model*).
HMM	:	Modelo de Markov oculto (*Hidden Markov Model*).
MMV	:	Método de Máxima Verossimilhança.
MSE	:	Erro médio quadrático (*Mean Squared Error*).
OECD	:	Organização para Cooperação e Desenvolvimento Econômico (*Organisation for Economic Co-operation and Development*).
OPEP	:	Organização dos Países Exportadores de Petróleo.
PH	:	Distribuição *Phase-Type*.
RW	:	Modelo de passeio aleatório (*Random Walk*).

WN : Modelo de ruído branco (*White Noise*).

WTI : Óleo leve, americano, com entrega em Cushing, Oklahoma (*West Texas Intermediate*).

Notação

Referente a Modelos de Markov:

S_i	:	Estado i de uma cadeia de Markov.
q_t	:	Estado no qual uma cadeia de Markov se encontra no instante de tempo t.
π	:	Distribuição de probabilidades iniciais dos estado de uma cadeia de Markov.
π_i	:	Probabilidade inicial do estado S_i de uma cadeia de Markov.
A	:	Matriz de probabilidades de transição de estados de uma cadeia de Markov de tempo discreto.
a_{ij}	:	Probabilidade da transição de estados (S_i, S_j) em uma cadeia de Markov de tempo discreto.
V	:	Conjunto de símbolos de um modelo de Markov oculto.
v_k	:	Símbolo individual, pertencente ao conjunto V.
B	:	Matriz de probabilidades de emissão de símbolos de uma cadeia de Markov oculta.
$b_j(k)$	:	Probabilidade de emissão do símbolo v_k, dado que a cadeia de Markov oculta se encontra no estado S_j.
Q	:	Matriz de taxas de transição de estados de uma cadeia de Markov de tempo contínuo.
λ_{ij}	:	Taxa, exponencial, de ocorrência da transição de estados (S_i, S_j) em uma cadeia de Markov de tempo contínuo.

q_{ij} : Elemento da matriz Q, cujo valor é λ_{ij} se $i \neq j$ ou $-\sum_{i \neq j} \lambda_{ij}$ se $i = j$.

N : Número de estados de uma cadeia de Markov.

M : Número de símbolos de um modelo de Markov oculto.

λ : Conjunto de parâmetros de um modelo de Markov oculto.

O : Conjunto de observações.

$\alpha_t(i)$: Probabilidade da seqüência parcial $O_1 O_2 \ldots O_t$ e $q_t = S_i$ dado o modelo λ.

Referente a Wavelets:

$f(\cdot)$: Uma função qualquer.

$\tilde{f}(\cdot)$: Aproximação da função $f(\cdot)$.

ψ : Uma Wavelet.

$\psi_{j,k}$: Wavelet ψ escalonada de 2^j e transladada de k unidades.

ϕ : Função de escala a qual ψ está associada.

$\phi_{j,k}$: Função de escala ϕ escalonada de 2^j e transladada de k unidades.

A_{2^j} : Operador que aproxima uma função $f(\cdot)$ na resolução 2^j.

$A_{2^j} f$: Aproximação de $f(\cdot)$ na resolução 2^j.

$A^d_{2^j} f$: Aproximação discreta de $f(\cdot)$ na resolução 2^j.

ϕ_{2^j} : $2^j \phi(2^j x)$.

ψ_{2^j} : $2^j \psi(2^j x)$.

$\mathbb{V}_{2^j}$: Conjunto de todas as aproximações possíveis, na resolução 2^j, de funções em $\mathbf{L}^2(\mathbb{R})$.

$\mathbb{O}_{2^j}$: Complemento ortogonal de $\mathbb{V}_{2^j}$ em $\mathbb{V}_{2^{j+1}}$, ou seja, $\mathbb{O}_{2^j} \oplus \mathbb{V}_{2^j} = \mathbb{V}_{2^{j+1}}$.

P_{2^j}	:	Operador de projeção ortogonal em $\mathbb{O}_{2^j}$.
$D_{2^j}f$	:	Detalhes de $f(\cdot)$ na resolução 2^j.

Referente a Especificação do Modelo Construído:

N	:	Número de estados da cadeia de Markov oculta.
M	:	Número de símbolos do modelo de Markov oculto.
F	:	Tamanho, em dias, da janela de previsão.
H	:	Tamanho, em dias, do histórico recente utilizado na previsão.
Ψ	:	Intervalo entre duas previsões.
V	:	Tamanho da subdivisão da faixa de retornos.
τ	:	Intervalo entre dois treinamentos do HMM.
T	:	Quantidade de observações utilizadas em um treinamento do HMM.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Assim como ocorre com a maioria das mercadorias comercializadas, o preço do petróleo é, basicamente, determinado pelo balanço entre oferta e demanda. Apesar de ser, aparentemente, simples, este balanço é bastante complexo, pois existem inúmeras variáveis, muitas das quais são imprevisíveis e extremamente voláteis, que podem afetá-lo. Guerras e tensões políticas, descobertas de novas reservas, intensidade dos invernos, desenvolvimento de novas fontes de energia, introdução de novas tecnologias de exploração, desenvolvimento econômico, são apenas alguns exemplos de fatores que pressionam a balança para um lado ou para outro, e, portanto, exercem influência sobre preço do petróleo.

A recente crise econômica é um excelente exemplo do enorme conjunto de fatores que, de uma forma ou outra, afetam o mercado do petróleo. Até o primeiro semestre de 2008, o preço do óleo cru vinha experimentando uma forte valorização, impulsionada pela alta liquidez mundial e pelo acelerado crescimento econômico do mundo, especialmente da China. Somente entre janeiro de 2007 e julho de 2008 , o preço do *West Texas Intermediate* (WTI)[1] elevou-se de 60 para 140 dólares, sofrendo uma incrível valorização de 130%. No entanto, a crise do mercado imobiliário americano, que surgiu no segundo semestre de 2008, foi suficiente para desencadear uma série de eventos que colocaram o mundo em uma grave recessão econômica, e, consequente-

[1]Óleo leve, americano, com entrega em Cushing, Oklahoma

mente, fizeram o preço do WTI despencar dos 140 dólares, para, aproximadamente, 40 dólares, ao final de 2008.

Esta intensa volatilidade de preço é extremamente prejudicial para empresas e países dependentes de petróleo, tanto consumidores quanto produtores. Tomemos como exemplo o caso das companhias aéreas. Um insumo essencial e insubstituível neste ramo é o querosene de aviação. A elevação brusca dos preços do petróleo, como a mencionada acima, aumenta, na mesma velocidade, os custos destas companhias, que, obrigadas a se adaptarem rapidamente a nova realidade, acabam repassando este aumento para os consumidores e, muitas vezes, despedindo funcionários.

Existem diversas maneiras pelas quais um agente pode tentar se proteger destas oscilações no preço. A mais comum é a utilização de instrumentos de *hedge*. A forma mais usual de se fazer um *hedge* é utilizando os mercados futuros ou a termo, como é ilustrado pelo exemplo a seguir. Suponha que um refinador saiba que, no mês seguinte, precisará comprar uma determinada quantidade de óleo cru. Evidentemente, o refinador não sabe quanto o barril do petróleo estará custando nesta data. Com medo de um possível aumento do preço, ele decide, hoje, negociar com seu fornecedor o preço de compra do barril daqui há um mês. Assim, ambos fazem um contrato, conhecido como contrato a termo, no qual comprometem-se a, no mês seguinte, negociar o barril de petróleo por, digamos, 50 dólares. Desta forma, o refinador garante seu preço de compra e não precisa mais se preocupar com as flutuações do preço. Uma excelente discussão sobre instrumentos e estratégias de *hedge* pode ser encontrada em [1].

No exemplo acima, apesar de o refinador ter eliminado seu risco, fixando o preço de compra do barril em 50 dólares, nada garante que ele fez um bom negócio. Isto só será verdade se, no mês da compra, o preço do petróleo no mercado à vista for superior a 50 dólares o barril. Caso não o seja, ele poderá se arrepender de ter feito o *hedge*!

1.1 Objetivo

O exemplo anterior ilustra a importância de se construir um modelo de previsão para o preço do petróleo. No entanto, devido às inúmeras variáveis que o afetam, e da complexidade destas, a construção de um modelo de regressão, que relaciona o preço a um conjunto de outras variáveis, em uma estrutura de casualidades, é uma tarefa bastante difícil. Assim, uma forma alternativa de se tentar fazer esta previsão, é através de modelos de séries temporais, os quais, observando somente o comportamento passado de uma variável, tentam inferir seu comportamento futuro. Nosso trabalho se insere neste contexto.

O objetivo desta dissertação é propor uma metodologia que, observando a dinâmica recente dos preços do petróleo, tenta prever seu comportamento futuro, no curto prazo[2]. Mais especificamente, procuramos, através da metodologia proposta, prever a distribuição da variação acumulada pelo preço do petróleo ao final de uma janela de tempo futura de F dias. Esperamos, deste modo, construir um instrumento que auxilie a tomada de decisão de investimentos, por participantes do mercado de petróleo.

Devido a volatilidade dos preços, utilizamos uma técnica de suavização derivada da aproximação de funções por Wavelets, buscando, assim, eliminar parte do ruído existente na série e ressaltar suas reais tendências. Em seguida, fazemos a modelagem da dinâmica dos preços, e sua previsão, utilizando um modelo de Markov oculto.

As principais contribuições desta dissertação são a apresentação de uma técnica para suavização de séries temporais financeiras (que já é bastante consolidada na área de compressão de imagens e processamento de sinais, mas ainda pouco explorada na área de finanças), e a verificação da serventia dos modelos de Markov ocultos na previsão das mesmas.

[2]Consideramos como curto prazo um intervalo de 20, 30 dias.

1.2 Roteiro

No capítulo 2 discorremos sobre os principais conceitos matemáticos que serão utilizados ao longo desta dissertação. Começamos introduzindo a teoria básica de modelos de Markov e, então, extendemo-la para os modelos de Markov ocultos. Em seguida, apresentamos as Wavelets. No capítulo 3 discutimos os principais modelos que são utilizados na previsão de séries financeiras, e apresentamos alguns resultados obtidos por outros trabalhos. No capítulo 4 a metodologia proposta é detalhadamente apresentada, e seus resultados são exibidos no capítulo 5. Finalmente, o capítulo 6 traz as conclusões e algumas propostas para trabalhos futuros.

Capítulo 2

Conceitos Teóricos

2.1 Modelos de Markov Ocultos

Fenômenos que ocorrem no mundo geralmente produzem algum tipo de saída observável, a qual denominamos de sinal. Devido à dificuldade e/ou custo de se manipular estes sistemas diretamente, é de grande interesse estudar e caracterizar seus sinais através de modelos, para que possamos aprender sobre seu processo gerador, sem que haja necessidade de termos o processo real disponível. Tais modelos nos permitem simular a fonte geradora do sinal e construir diferentes sistemas, como os de previsão, de reconhecimento, de identificação, etc., de maneira eficiente. Existem duas classes de modelos: determinísticos e estocásticos.

Nesta dissertação, utilizamos um modelo estocástico cada vez mais em uso: o modelo de Markov oculto (HMM - do inglês *Hidden Markov Model*)[2]. Devido a sua rica e estabelecida estrutura matemática, o HMM tem sido extensivamente utilizado na área de reconhecimento de voz, e vem se tornando cada vez mais popular em diversas outras áreas [3], como bioinformática, processamento de sinais e redes de computadores. A idéia principal por trás do modelo é que existem diversos fenômenos cujas saídas dependem de fatores que não são diretamente observáveis (estão ocultos) mas podem ser inferidos a partir destas saídas. Utilizando-o, é possível fazer uma distinção estatística destes fatores ocultos, separando-os em diferentes estados de uma cadeia de Markov, o que por sua vez, em muitos casos, permite também uma fácil visualização, interpretação e caracterização dos mesmos.

Começamos este capítulo introduzindo os conceitos dos modelos de Markov [4, 5], e extendemo-los para o modelo de Markov oculto. Em seguida, caracterizamos o HMM e apresentamos seus fundamentos matemáticos.

2.1.1 Modelos de Markov de Tempo Discreto

Considere um sistema que possa ser descrito como estando em um de N estados distintos, $S_1, S_2, \cdots, S_N$; e cujas transições de estado são dadas por um conjunto de probabilidades associadas a cada estado S_i, como ilustrado na Figura 2.1 (onde $N = 3$ por simplicidade). A cada instante de tempo $t_1, t_2, \cdots$, o sistema é observado[1], a fim de se determinar seu estado atual. Denotaremos por q_t o estado no qual o sistema se encontra no tempo t. Em geral, uma descrição probabilística completa de um sistema como este requer que sejam especificados o estado atual q_t, assim como todos os estados anteriores, $q_{t-1}, q_{t-2}, \ldots, q_1$. Entretanto, para o caso de um processo markoviano, essa descrição é reduzida a apenas o estado atual e o imediatamente anterior, ou seja,

$$P[q_t = S_i | q_{t-1} = S_j, q_{t-2} = S_k, \ldots] = P[q_t = S_i | q_{t-1} = S_j] \tag{2.1}$$

[1] Os instantes de observação não precisam ser, necessariamente, uniformemente espaçados.

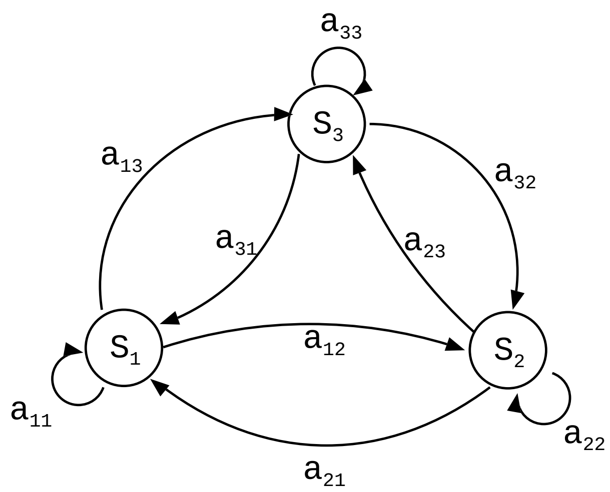


Figura 2.1: Cadeia de Markov de tempo discreto com 3 estados (S_1, S_2, S_3) e suas transições.

Consideramos ainda, apenas os processos cujo lado direito de (2.1) é independente do tempo, o que nos leva a um conjunto de probabilidades de transição de estados da forma

$$a_{ij} = P[q_t = S_i | q_{t-1} = S_j], \qquad 1 \leq i, j \leq N \tag{2.2}$$

cujos coeficientes têm as seguintes propriedades

$$a_{ij} \geq 0 \tag{2.3a}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{2.3b}$$

pois obedecem às restrições estocásticas.

O processo estocástico descrito acima é conhecido como modelo de Markov observável, uma vez que sua saída é o conjunto dos estados nos quais a cadeia se encontrava em cada instante de tempo t, e cada estado corresponde a um evento diretamente observável. Para ilustrar os conceitos apresentados, suponha um modelo de Markov de tempo discreto, simples, de três estados, que descreva o tempo meteorológico de uma região qualquer. Assumimos que, a cada dia, o tempo é observado como estando em um de três estados distintos:

Estado 1: Chuvoso

Estado 2: Nublado

Estado 3: Sol

e que cada estado da cadeia representa um dos três estados acima.

Seja A a matriz de probabilidades de transição entre estados definida como

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Dado que o tempo no primeiro dia ($t = 1$) é sol (estado 3), podemos nos perguntar qual seria a probabilidade (de acordo com o modelo) de que nos próximos sete dias o tempo seja *sol-sol-chuva-chuva-sol-nublado-sol*. Formalmente, definimos uma seqüência $O = S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3$ que corresponde a $t = 1, 2, \ldots, 8$, e o que queremos saber é, dado o modelo, qual a probabilidade de O. Esta probabilidade pode ser expressada como

$$\begin{aligned} P(O|modelo) &= P[S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3|modelo] \\ &= P[S_3] \cdot P[S_3|S_3] \cdot P[S_3|S_3] \cdot P[S_1|S_3] \cdot P[S_1|S_1] \cdot P[S_3|S_1] \cdot P[S_2|S_3] \cdot P[S_3|S_2] \\ &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\ &= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\ &= 1.536 \times 10^{-4} \end{aligned}$$

onde usamos a notação

$$\pi_i = P[q_1 = S_i], \qquad 1 \leq i \leq N \tag{2.4}$$

Distribuição do Tempo de Permanência em um Estado

Outra pergunta que podemos fazer é: dado que o modelo se encontra em um determinado estado, qual a probabilidade de ele permanecer no mesmo, por exatamente d dias? Em termos formais, queremos encontrar a probabilidade da seqüência de observações

$$O = \{\underbrace{S_i, S_i, S_i, \ldots, S_i, S_j}_{d+1}\} \qquad \text{com } S_j \neq S_i$$

dado o modelo. Esta pode ser expressa como

$$P(O|\text{modelo},\, q_1 = S_i) = a_{ii}^{d-1}(1 - a_{ii}) = p_i(d) \tag{2.5}$$

onde $p_i(d)$ é a distribuição de probabilidades discreta da duração d no estado i. É importante ressaltar que, esta distribuição exponencial (ou geométrica, no caso da cadeia de Markov em tempo discreto) é característica da duração de um estado nas cadeias de Markov.

Podemos assim, dado $p_i(d)$, calcular o tempo esperado (a duração, em número de transições) que a cadeia permanece em um mesmo estado, dado que ela já se encontra no mesmo, como

$$E[d] = \overline{d_i} = \sum_{d=1}^{\infty} d \cdot p_i(d) \tag{2.6a}$$

$$= \sum_{d=1}^{\infty} d \cdot a_{ii}^{d-1}(1 - a_{ii}) \tag{2.6b}$$

$$= \frac{1}{1 - a_{ii}} \tag{2.6c}$$

Voltando ao nosso problema de previsão da meteorologia, o número esperado de dias consecutivos de sol é, segundo o modelo, $1/(1 - 0.8) = 5$ dias; para tempo nublado, esse valor é de 2.5 dias, e para chuva, 1.67 dias.

2.1.2 Modelos de Markov de Tempo Contínuo

Nos modelos de Markov de tempo discreto, estudados na seção 2.1.1, o sistema é observado em instantes de tempo discretos, bem definidos. No nosso modelo de previsão meteorológica, por exemplo, observamos o tempo uma vez a cada dia. Existe, no entanto, uma outra classe de modelos de Markov, nos quais o sistema é observado continuamente, a todo instante de tempo. Estes são conhecidos como modelos de Markov de tempo contínuo.

A principal característica desta classe de modelos é que, para cada estado, há um conjunto de um ou mais eventos que, ao ocorrerem, podem fazer a cadeia transicionar de estado. Para cada evento i, o tempo decorrido até sua próxima ocorrência segue uma distribuição exponencial

$$
\begin{aligned}
f(t_i) &= \lambda e^{-\lambda t_i}, \qquad t_i > 0 \\
&= 0, \qquad \text{para quaisquer outros valores.}
\end{aligned}
$$

onde t_i é o intervalo de tempo entre duas ocorrências do evento i, e λ é a taxa de ocorrência do evento i. Para todos os eventos, seus tempos até a próxima ocorrência são independentes entre si e, para um mesmo evento, tempos de ocorrência sucessivos são, também, independentes entre si.

Visando ilustrar estes novos conceitos, vamos voltar ao problema da seção anterior, e criar um modelo de Markov de tempo contínuo para o tempo meteorológico. O novo modelo é ilustrado na Figura 2.2. Observe que a estrutura da cadeia permanece a mesma, cada estado continua representando um dos três estados definidos para o tempo: sol, chuva, nublado. Entretanto, as transições, agora, não são mais descritas por probabilidades, e sim por taxas do tipo λ_{ij}, que definem a distribuição exponencial do intervalo de tempo até a próxima ocorrência do evento associado à transição ij.

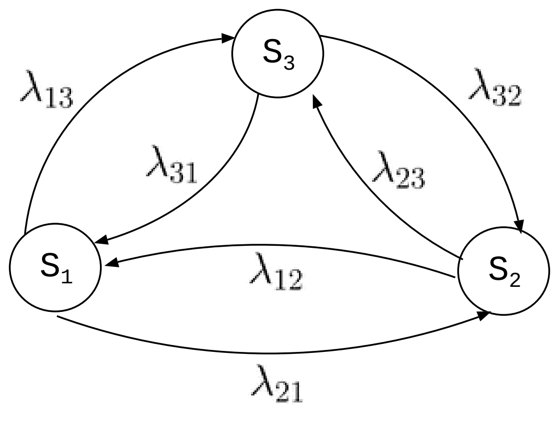


Figura 2.2: Cadeia de Markov de tempo contínuo com 3 estados (S_1, S_2, S_3) e suas transições.

Comparando os dois modelos construídos, o de tempo contínuo e o de tempo discreto, e suas interpretações, fica evidente a diferença do primeiro em relação ao segundo. Enquanto no modelo de tempo discreto o sistema é observado em instantes de tempos definidos, no modelo de tempo contínuo ele é observado continuamente, e suas mudanças de estado, que são definidas pela ocorrência de eventos, podem ocorrer em qualquer instante do tempo. Isto significa que, no caso do nosso exemplo, um modelo de tempo discreto permite que haja, apenas, um estado meteorológico por dia. Já o modelo de tempo contínuo permite que, em um mesmo dia, haja dois ou mais estados meteorológicos distintos, tempo nublado e chuva, por exemplo. A Figura 2.3 ilustra bem essa diferença.

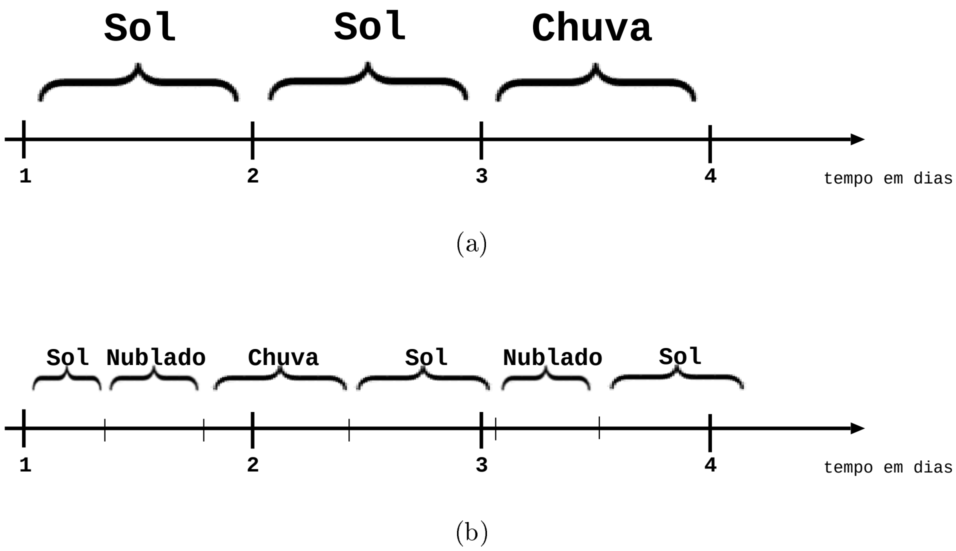


Figura 2.3: Conjunto amostral possível em: (a) um modelo de Markov de tempo discreto; (b) um modelo de Markov de tempo contínuo.

Formalmente, descrevemos a cadeia de Markov de tempo contínuo por dois elementos. O primeiro, o vetor $\pi = P[q_1 = S_i]$, é idêntico àquele definido para cadeias de Markov de tempo discreto, e informa a probabilidade de a cadeia se encontrar no estado i. O segundo, a matriz Q, é semelhante à matriz A definida anteriormente, e descreve as transições da cadeia. Definimos Q, para nosso modelo contínuo criado acima, como

$$Q = \{q_{ij}\} = \begin{bmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31} + \lambda_{32}) \end{bmatrix}$$

Observe que não há transições de um estado para ele mesmo, e que o elemento q_{ii} é definido como $q_{ii} = -\sum_{j \neq i} \lambda_{ij}$, ou seja, $-q_{ii}$ é a taxa de saída do estado i.

Distribuição do Tempo de Permanência em um Estado

Dado que o tempo de ocorrência de cada evento é exponencial, é evidente que o tempo de permanência em um mesmo estado i é, também, exponencialmente distribuído, com taxa $\sum_{j \neq i} \lambda_{ij}$. Assim, o tempo médio de permanência em um estado é

$$E[tp_i] = \overline{tp_i} = \frac{1}{\sum_{j \neq i} \lambda_{ij}} \tag{2.7}$$

onde tp_i é o tempo de permanência no estado i.

2.1.3 Modelos de Markov Ocultos em Tempo Discreto

No modelo markoviano descrito na seção 2.1.1, cada estado corresponde a um evento diretamente observável (chuva, tempo nublado ou sol). Entretanto, existem fenômenos nos quais podemos apenas observar suas saídas (seus sinais), e não o processo que as gerou. Para modelá-los, estendemos o conceito de modelo markoviano para incluir casos nos quais os sinais são uma função probabilística do estado da cadeia. Tais modelos são conhecidos como modelos de Markov ocultos. Em [2], o HMM é definido como um processo estocástico duplamente embutido, no qual o primeiro não é diretamente observável (é oculto), mas pode ser inferido através de um segundo conjunto de processos estocásticos, os quais produzem a seqüência de observação (os sinais).

A fim de clarear estes conceitos, considere o exemplo a seguir: Suponha que, em um quarto, exista uma barreira (uma cortina, por exemplo), atrás da qual há um jogador que dispõe de uma ou mais moedas, que podem ou não estar viciadas. Uma segunda pessoa, a quem chamaremos de observador, está do outro lado da cortina, e não pode ver o que acontece do lado do jogador. O jogador, segundo um critério qualquer, escolhe uma moeda, joga-a para o alto, e informa, ao observador, o resultado (cara ou coroa). Ele então repõe a moeda ao conjunto de moedas, escolhe uma novamente, e realiza o mesmo experimento. Assim, o observador conhece apenas o resultado de cada jogada (esta é a saída observável), mas desconhece qual das moedas (que estão ocultas) gerou esta saída. Uma possível seqüência de observações deste experimento seria:

$$\begin{aligned}\mathbf{O} &= O_1 O_2 O_3 \dots O_T \\ &= HHTTT \dots H\end{aligned}$$

onde H representa Cara e T Coroa.

Dado este cenário, um problema de interesse é o de como construir um modelo markoviano que explique a seqüência de caras e coroas observada. Uma primeira tentativa seria construir um modelo com apenas dois estados, no qual cada estado representa um lado da moeda. Neste caso, estaríamos assumindo que há apenas uma moeda sendo lançada, e, portanto, o modelo construído seria um modelo markoviano observável [2]. Definida a estrutura do modelo, o próximo passo seria decidir o valor para o viés da moeda, ou seja, a probabilidade de cada evento ocorrer. Este modelo pode ser visto na Figura 2.4. Entretanto, poderíamos assumir que em vez de uma, há duas moedas sendo utilizadas. Assim, construiríamos um HMM de dois estados, no qual cada um representaria uma moeda diferente, e seria caracterizado por uma distribuição de probabilidades de eventos, que, neste caso, se resumiria ao viés da moeda. O mecanismo que representaria a escolha das moedas seria, então, representado por uma matriz de probabilidades de transição de estados. A Figura 2.5 ilustra este modelo. Procedendo da mesma maneira, poderíamos assumir que há três moedas sendo jogadas e, nesses caso, construiríamos um HMM de três estados.

[2]Observe que a unidade de tempo, neste experimento, é definida como uma jogada da moeda. Assim, a cadeia de Markov utilizada é a de tempo discreto.

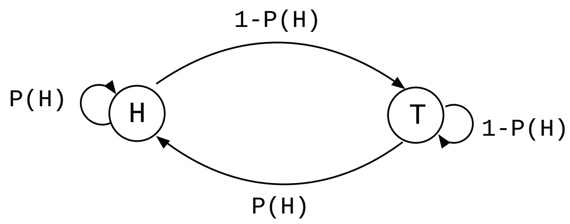


Figura 2.4: Modelo de Markov observável, de tempo discreto, para a suposição de que existe, apenas, uma moeda sendo lançada.

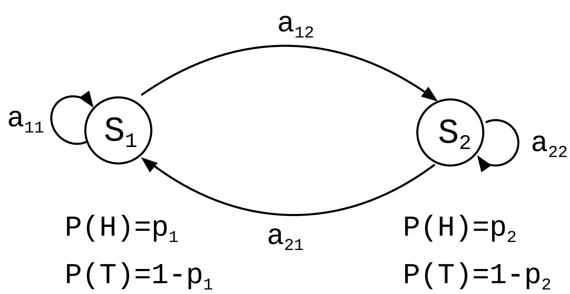


Figura 2.5: Modelo de Markov oculto, de tempo discreto, para a suposição de que existem duas moedas.

Elementos de um HMM de Tempo Discreto

Formalmente, um HMM é caracterizado pelos cinco elementos a seguir:

1. N: o número de estados do modelo. Apesar de estarem ocultos, geralmente são dotados de algum significado físico. Representamos cada estado individual por $S = \{S_1, S_2, \ldots, S_N\}$ e o estado no tempo t por q_t.

2. M: o número de símbolos distintos observáveis por estado. Correspondem às saídas físicas observáveis do sistema modelado. Individualmente, são representados por $V = \{v_1, v_2, \ldots, v_M\}$.

3. $A = \{a_{ij}\}$: a matriz de probabilidades de transição de estados; onde $a_{ij} = P[q_{t+1} = S_j | q_t = S_i],\ 1 \leq i, j \leq N,\ a_{ij} \geq 0,\ \sum_{j=1}^{N} a_{ij} = 1$

4. $B = \{b_j(k)\}$: a distribuição de emissão de símbolos no estado j; onde $b_j(k) = P[v_k | q_t = S_j],\ 1 \leq j \leq N,\ 1 \leq k \leq M,\ b_j(k) \geq 0,\ \sum_{k=1}^{M} b_j(k) = 1.$

5. $\pi = \{\pi_i\}$: o vetor de distribuição inicial de estados; onde $\pi_i = P[q_1 = S_i]$, $1 \leq i \leq N$.

Assim, como pode ser visto acima, uma descrição completa de um HMM necessita que sejam especificados dois parâmetros do modelo (N e M), os símbolos de observação, e mais três medidas de probabilidade (A, B e π). Por conveniência, usaremos, no restante desta dissertação, a notação compacta $\lambda = (A, B, \pi)$ para indicar o conjunto completo de parâmetros de um modelo HMM.

2.1.4 Estimação de Parâmetros de um HMM

Após a especificação do modelo, o problema que surge naturalmente é: como calibrar seus parâmetros de acordo com a seqüência de observações $O = O_1 O_2 O_3 \ldots O_T$ dada? Esta pergunta pode ser dividida em duas etapas. Na primeira, queremos encontrar uma maneira eficiente de avaliar $P(O|\lambda)$, i.e., a probabilidade de a seqüência O ter sido gerada pelo modelo λ. Na segunda, precisamos estimar os parâmetros A, B e π, de modo a maximizar $P(O|\lambda)$, ou seja, buscamos ajustar nosso modelo, da melhor forma possível, aos dados reais. Este processo de ajuste é conhecido como Método de Máxima Verossimilhança (MMV). A seqüência de observações usada para fazer esta adaptação é chamada de **seqüência de treinamento**, uma vez que é usada para **treinar** [3] o HMM.

O MMV busca encontrar um estimador de máxima verossimilhança. Infelizmente, para o caso de um HMM, não se conhece uma expressão analítica para este estimador. Entretanto, existe um método, derivado do algoritmo *Expectation-Maximization*(EM), conhecido como algoritmo de *Baum-Welch* [6], que realiza a estimação de parâmetros de máxima verossimilhança para um HMM de forma iterativa. Ele começa com uma atribuição arbitrária de valores para λ, e produz estimativas sucessivamente melhores, garantindo a convergência para um máximo

[3] Na literatura de HMM's, a palavra treinamento é utilizada como sinônimo de estimação de parâmetros. Entretanto, é importante ressaltar que, diferentemente de um treinamento de redes neurais, por exemplo, o treinamento do HMM não envolve aprendizado com os erros.

local da função de verossimilhança, sempre que um existir [7].

A seguir, apresentamos as soluções matemáticas para ambas as etapas descritas acima. A primeiro (cálculo de $P(O|\lambda)$) pode ser resolvida pelo procedimento *forward-backward* [8, 9], e a segunda (estimação do parâmetros) utilizando o algoritmo *EM*.

Procedimento *Forward-Backward*

Seja a variável *forward* $\alpha_t(i)$ definida como:

$$\alpha_t(i) = P(O_1 O_2 \ldots O_t, q_t = S_i|\lambda) \tag{2.8}$$

isto é, a probabilidade de observarmos a seqüência parcial $O_1O_2 \ldots O_t$, e de a cadeia estar no estado S_i, no tempo t, dado o modelo λ. O valor de $\alpha_t(i)$ pode ser calculado com o seguinte algoritmo:

1. Inicialização:

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \leq i \leq N \tag{2.9}$$

2. Indução:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \qquad \begin{array}{l} 1 \leq t \leq T-1 \\ 1 \leq j \leq N \end{array} \tag{2.10}$$

3. Término:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2.11}$$

A variável *forward* é inicializada com o valor da probabilidade de a cadeia estar no estado i em $t = 1$ e o símbolo O_1 ter sido emitido pelo mesmo. A indução é ilustrada na Figura 2.6. Em $t+1$, o estado S_j pode ter sido atingido partindo-se de qualquer um dos N estados, portanto, a probabilidade de a cadeia estar em S_j e da seqüência $O_1O_2 \ldots O_t$, em $t+1$, é dada pelo somatório da indução. Uma vez no estado S_j, temos que considerar a emissão do símbolo O_{t+1}. Para tal, basta multiplicarmos o valor do somatório por $b_j(O_{t+1})$. Finalmente, se somarmos todas as variáveis terminais $\alpha_T(i)$, encontramos $P(O|\lambda)$.

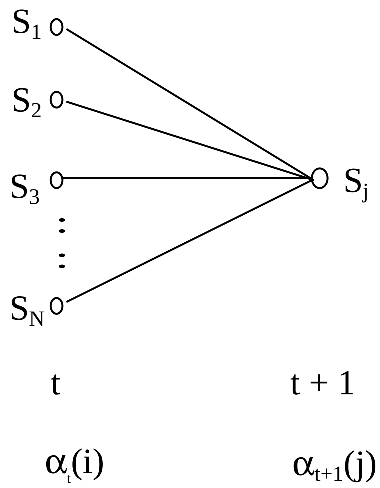


Figura 2.6: Procedimento Forward

Seja a variável *backward* $\beta_t(i)$ definida como:

$$\beta_t(i) = P(O_{t+1}O_{t+2}\dots O_T|q_t = S_i, \lambda) \tag{2.12}$$

isto é, a probabilidade de observarmos a seqüência parcial $O_{t+1}O_{t+2}\dots O_T$, no intervalo de tempo $[t+1, T]$, dado que a cadeia está no estado S_i, no tempo t, e o modelo λ. A variável $\beta_t(i)$ pode ser calculada usando o seguinte algoritmo:

1. Inicialização:

$$\beta_T(i) = 1, \qquad 1 \leq i \leq N \tag{2.13}$$

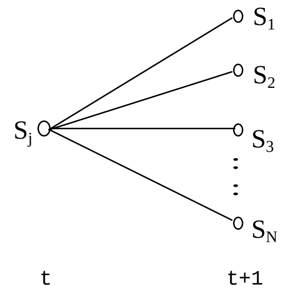
2. Indução:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \qquad t = T-1, T-2, \dots, 1$$

$$1 \leq i \leq N \tag{2.14}$$

Na inicialização, definimos, arbitrariamente, que $\beta_T(i) = 1$. A indução é ilustrada na Figura 2.7. Em $t+1$ a cadeia pode ter transicionado para qualquer um de seus N estados (por isso o termo a_{ij}). Neste, com probabilidade $b_j(O_{t+1})$ é emitido o símbolo

O_{t+1}. Assim, $\beta_t(i)$ pode ser calculado a partir de $\beta_{t+1}(j)$, segundo a indução descrita acima.



$\beta_t(i)$ $\beta_{t+1}(j)$

Figura 2.7: Procedimento Backward

Note-se que podemos calcular $P(O|\lambda)$ usando apenas a variável *forward* do procedimento *forward-backward*. Entretanto, a variável *backward* é utilizada pelo algoritmo EM (descrito a seguir) e, portanto, foi necessário definí-la também.

Algoritmo *Expectation-Maximization*(EM)

O algoritmo EM [7, 10] é um método para estimação de parâmetros de máxima verossimilhança de uma distribuição, a partir de um conjunto de dados incompleto. Existem dois cenários principais nos quais é interessante a aplicação do EM. No primeiro, o conjunto de dados não contém todos os valores, ou seja, há dados que estão faltando, que foram perdidos. Isto pode ocorrer por problemas ou limitações no processo de observação. No segundo, a otimização da função de verossimilhança é intratável analiticamente, mas pode ser aproximada se assumirmos a existência de parâmetros ocultos. Este último cenário é o que se aplica a um HMM.

Como definido anteriormente, seja $O = O_1 O_2 \ldots O_T$ a seqüência de dados observada, a qual chamaremos de seqüência incompleta de dados, e $q = q_1 q_2 \ldots q_T$ a seqüência (oculta) de estados da cadeia que gerou O. Seja $Z = (O, q)$ o conjunto completo de dados. Podemos então definir a seguinte função de verossimilhança

$$L(Z, \lambda) = P(Z|\lambda) = P(O, q|\lambda) = P(q|O, \lambda)P(O|\lambda) \tag{2.15}$$

chamada de verossimilhança de dados completos, onde a última igualdade é obtida pelo teorema de Bayes. É importante observar que esta função é, na realidade, uma variável aleatória, uma vez que q é desconhecido e, presumidamente, governado por alguma distribuição.

O primeiro passo do algoritmo EM, chamado de passo-E, é encontrar a esperança do log da função de verossimilhança de dados completa, $\log[P(O,q|\lambda)]$, com respeito ao conjunto de dados ocultos q, dada a seqüência observada O e os valores atuais (iniciais) dos parâmetros do modelo. Assim, definimos a função auxiliar

$$\begin{aligned} Q(\lambda,\lambda^{i-1}) &= E\Big[\log P(O,q|\lambda)\big|O,\lambda^{i-1}\Big] \\ &= \sum_{q\in\mathcal{Q}} \log P(O,q|\lambda)P(q|O,\lambda^{i-1}) \end{aligned} \tag{2.16}$$

onde λ^{i-1} é o conjunto de parâmetros atuais do modelo usados para calcular a esperança, λ o conjunto de novos parâmetros que serão otimizados a fim de maximizar Q, e $\mathcal{Q}$ é o conjunto de todas as seqüências de estados possíveis, de tamanho T.

Dada a seqüência particular de estados q, podemos representar $P(O,q|\lambda)$ por

$$P(O,q|\lambda) = \pi_{q_1} b_{q_1}(o_1) \prod_{t=2}^{T} a_{q_{t-1}q_t} b_{q_t}(o_t) \tag{2.17}$$

e assim, substituindo (2.17) em (2.16), a função Q pode ser reescrita como

$$\begin{aligned} Q(\lambda,\lambda^{i-1}) = \sum_{q\in\mathcal{Q}} \log \pi_{q_1} P(q|O,\lambda^{i-1}) + \sum_{q\in\mathcal{Q}} \left(\sum_{t=1}^{T-1} \log a_{q_t q_{t+1}} \right) P(q|O,\lambda^{i-1}) \\ + \sum_{q\in\mathcal{Q}} \left(\sum_{t=1}^{T} \log b_{q_t}(o_t) \right) P(q|O,\lambda^{i-1}) \end{aligned} \tag{2.18}$$

O segundo passo do EM, chamado de passo-M, é maximizar a esperança obtida no primeiro passo. Assim, buscamos

$$\lambda^i = \underset{\lambda}{\operatorname{argmax}}\ Q(\lambda,\lambda^{i-1})$$

Como os parâmetros que desejamos otimizar (π, A e B) estão separados em três termos independentes na soma vista na equação (2.18), podemos maximizar individualmente cada um. Maximizando cada termo, e levando em conta as restrições estocásticas, chegamos às seguinte fórmulas de reestimação de parâmetros [7, 10]:

$$\pi_i = P(q_1 = S_i | O, \lambda^{i-1}) \tag{2.19a}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(q_t = S_i, q_{t+1} = S_j | O, \lambda^{i-1})}{\sum_{t=1}^{T-1} P(q_t = S_i | O, \lambda^{i-1})} \tag{2.19b}$$

$$b_{ij} = \frac{\sum_{t=1}^{T} P(q_t = S_i | O, \lambda^{i-1}) \mathbb{I}\{o_t = v_j\}}{\sum_{t=1}^{T} P(q_t = S_i | O, \lambda^{i-1})} \tag{2.19c}$$

onde utilizamos a notação $\mathbb{I}\{c\}$ para representar a *função indicadora* de uma condição c, que vale 1 quando a condição é satisfeita, e 0, caso contrário.

As probabilidades encontradas nas equações (2.19a), (2.19b) e (2.19c) podem ser calculadas a partir das variáveis *forward* e *backward*. Seja

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \tag{2.20}$$

isto é, a probabilidade de se estar no estado S_i no instante t, dada a seqüência de observações O e o modelo λ. Podemos expressar a equação (2.20) em termos das variáveis *forward* e *backward*:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{2.21}$$

onde $\alpha_t(i)$ contribui com a seqüência parcial de observação $O_1 O_2 \ldots O_t$ e o estado S_i no tempo t, enquanto $\beta_t(i)$ contribui com o restante da seqüência $O_{t+1} O_{t+2} \ldots O_T$, dado o estado S_i no tempo t. Se somarmos $\gamma_t(i)$ para todo t, encontramos um valor que pode ser interpretado como o número esperado (no tempo) de visitas ao estado S_i, ou, equivalentemente, o número esperado de transições a partir do estado S_i (se excluirmos o tempo T da soma).

Seja também a variável $\xi_t(i,j)$, definida como a probabilidade de transição do estado i, no instante de tempo t, para o estado j, no instante de tempo $t+1$, dado o modelo λ e a seqüência de observações O:

$$\begin{aligned} \xi_t(i,j) &= P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\displaystyle\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \tag{2.22}$$

Similarmente a $\gamma_t(i)$, o somatório de $\xi_t(i,j)$ em t pode ser interpretado como o número esperado de transições do estado S_i para o estado S_j.

Assim, expressões em (2.19) podem ser reescritas em termos de $\gamma_t(i)$ e $\xi_t(i,j)$, e interpretadas, intuitivamente, como:

$$\begin{aligned}
\pi_i &= \gamma_1(i) \\
&= \#\text{ esperado de vezes no estado } S_i \text{ em } t = 1. && (2.23a) \\
a_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\
&= \frac{\#\text{ esperado de transições de } S_i \text{ para } S_j}{\#\text{ esperado de transições a partir de } S_i} && (2.23b) \\
b_{ij} &= \frac{\sum_{t=1}^{T} \gamma_t(j)\mathbb{I}\{o_t = v_k\}}{\sum_{t=1}^{T} \gamma_t(j)} \\
&= \frac{\#\text{ esperado de vezes no estado } S_j \text{ emitindo o símbolo } v_k}{\#\text{ esperado de vezes no estado } S_j} && (2.23c)
\end{aligned}$$

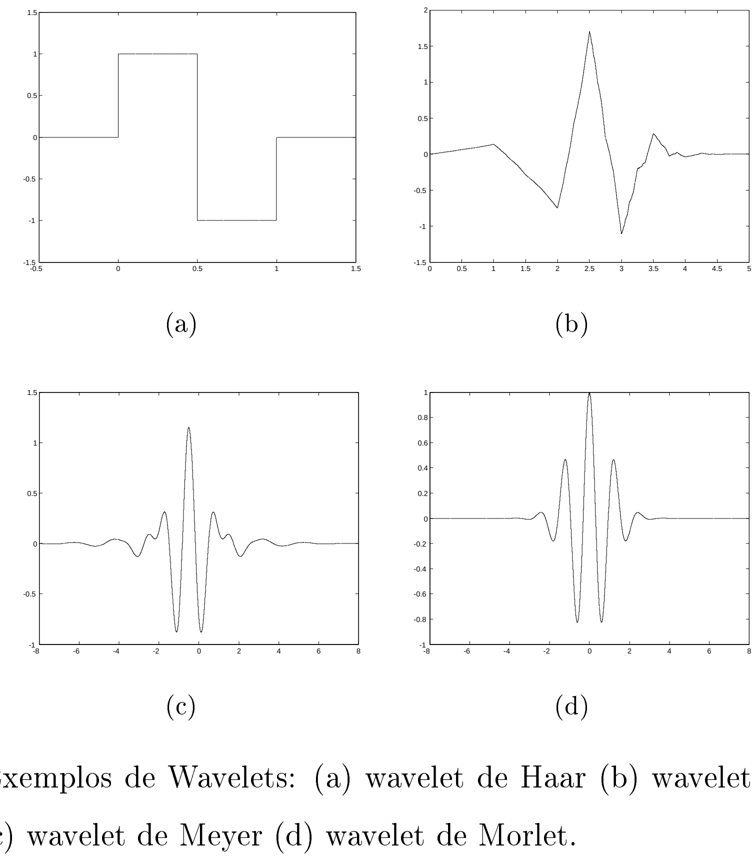
Estes dois passos (passo-E e passo-M) são repetidos até que $P(O|\lambda^i) - P(O|\lambda^{i-1}) \leq \epsilon$. Foi provado por *Baum* e seus co-autores [2] que, a cada iteração do algoritmo, ou (1) os parâmetros iniciais λ definem um ponto crítico da função de verossimilhança e, no caso, $\lambda^i = \lambda^{i-1}$; ou (2) o novo modelo estimado λ^i é mais provável que o inicial λ^{i-1} ($P(O|\lambda^i) > P(O|\lambda^{i-1})$). Desse modo, fica garantido que o algoritmo converge para um máximo local da função de verossimilhança. Uma excelente referência, de fácil leitura, para a compreensão do algoritmo EM pode ser encontrada em [10].

Influência dos Valores Iniciais dos Parâmetros

Em teoria, as equações de reestimação (2.19) devem fornecer valores dos parâmetros do HMM que correspondam à um máximo local da função de verossimilhança. Conforme visto acima, o algoritmo EM necessita que, para cada parâmetro a ser estimado, tenha sido dado um valor inicial. Assim, uma pergunta que surge é como estimar estes valores iniciais, de modo que o máximo local atingido pelo algoritmo corresponda ao máximo global. Infelizmente, não há resposta simples para esta questão. [2] indica que, para os parâmetros π e A, uma escolha uniforme ou aleatória de valores iniciais é adequada para a maioria dos casos. Entretanto, para os parâmetros de B, uma boa escolha dos valores iniciais, baseada no conhecimento do processo o qual se está estudando, pode ter grande influência.

2.2 Wavelets

Wavelets são funções matemáticas ondulatórias, de duração limitada, cuja integral no tempo é zero, conforme ilustrado na Figura 2.8. Assim como as funções seno e cosseno na análise de Fourier, wavelets são utilizadas como funções base para a representação de outras funções pertencentes a $\mathbf{L}^2(\mathbb{R})$[4]. As wavelets surgiram no



(a) (b) (c) (d)

Figura 2.8: Exemplos de Wavelets: (a) wavelet de Haar (b) wavelet de Daubechie de ordem 3 (c) wavelet de Meyer (d) wavelet de Morlet.

início do século XX, tendo sua primeira menção sido feita num apêndice da tese de A. Haar em 1909. Entretanto, a maior parte do trabalho foi feito nos anos 30. Hoje,

[4] $\mathbf{L}^2(\mathbb{R})$ é o espaço vetorial das funções unidimensionais, mensuráveis e quadraticamente integráveis. Uma função é dita quadraticamente integrável se a integral do quadrado de seu valor absoluto é finito, ou seja, se $\int_{-\infty}^{\infty} | f(x) |^2 \, \mathrm{d}x$ é finita.

a teoria matemática por trás das wavelets já está bastante consolidada [11], e suas aplicações podem ser vistas nas áreas de compressão de dados, astronomia, acústica, engenharia nuclear, processamento de sinais e imagens, música, ótica, previsão de terremotos, solução de equações diferenciais, entre outras.

Suponha uma função qualquer $f(x)$. A análise de Fourier consiste em construir uma aproximação $\tilde{f}(x)$, a partir da soma de senos e cossenos com diferentes freqüências

$$\tilde{f}(x) = a_0 + \sum_{k=1}^{\infty}(a_i \sin(kx) + b_i \cos(kx)) \tag{2.24}$$

onde os coeficientes a_0, a_i e b_i são obtidos através da Transformada de Fourier, definida como

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\mathrm{d}t. \tag{2.25}$$

Os coeficientes obtidos de $F(\omega)$ (conhecidos como coeficientes de Fourier) quando multiplicados por suas respectivas senóides de freqüência ω, formam os componentes senoidais nos quais a função original é decomposta. Graficamente, a transformada pode ser vista na Figura 2.9 e seu resultado é ilustrado na Figura 2.11(a).

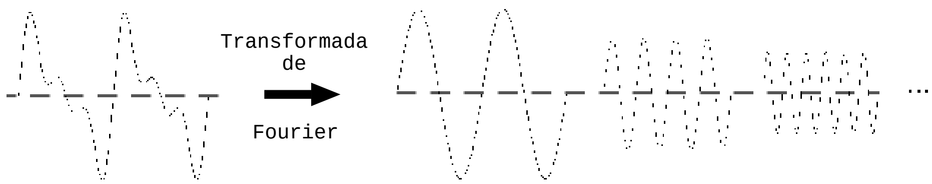


Figura 2.9: Sinal original e suas componentes senoidais

A análise por wavelets é feita de forma semelhante, no entanto, ao invés das funções seno e cosseno, utilizamos versões escalonadas (através de dilatações e contrações) e transladas da wavelet original $\psi(x)$, que é conhecida como wavelet mãe (do inglês *mother wavelet*). Assim, a aproximação $\tilde{f}(x)$ é obtida pela combinação linear

$$\tilde{f}(x) = \sum_{\substack{\forall\,\text{translação} \\ \forall\,\text{escala}}} C(\text{ escala,translação })\psi(\text{ escala,translação,x }) \tag{2.26}$$

onde a transformada de wavelet

$$C(\text{ escala,translação }) = \int_{-\infty}^{\infty} f(x)\psi(\text{ escala,translação,x })\mathrm{d}x \tag{2.27}$$

fornece os coeficientes de wavelet $C(\text{ escala,translação })$ que, multiplicados pela wavelet mãe, apropriadamente escalonada e transladada, formam os componentes nos quais a função original é decomposta. Este processo pode ser visto, graficamente, na Figura 2.10 e seu resultado é ilustrado na Figura 2.11(b).

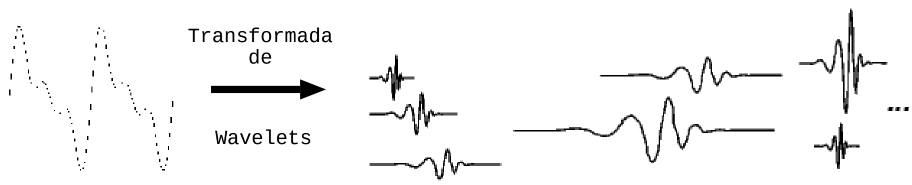
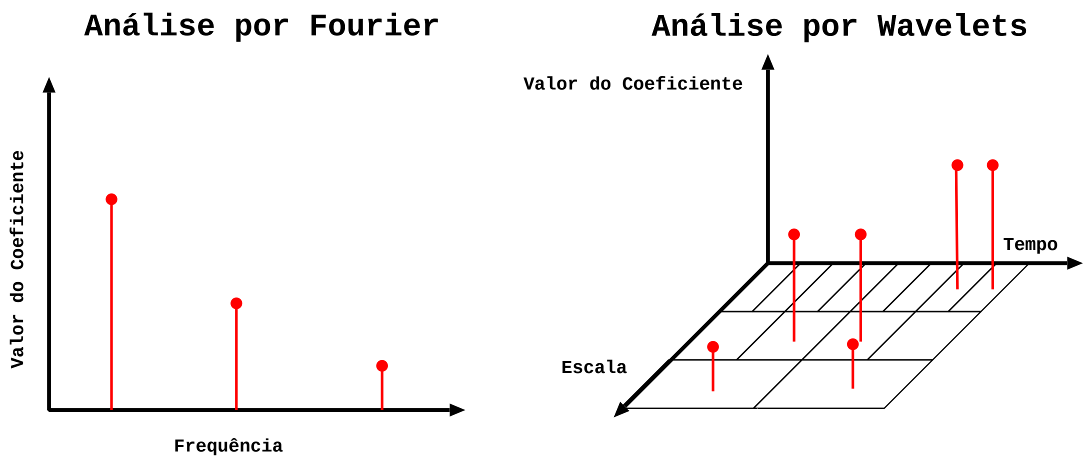


Figura 2.10: Sinal original e suas componentes de wavelets

As principais diferenças entre a análise por wavelets e a análise de Fourier são conseqüência da característica local das wavelets. Note que, enquanto as funções seno e cosseno são infinitas, i.e., se estendem sobre todo o domínio de $f(x)$, as wavelets têm duração limitada, são finitas. Esta importante característica traz à análise por wavelets uma série de vantagens sobre a análise de Fourier. Ela nos permite fazer um estudo local do sinal, i.e, de partes específicas do mesmo, sem que seus resultados afetem todo o domínio da função, como ocorre com a análise de Fourier. Suponha que nossa função $f(x)$ seja suave, definida no domínio $x \in (-\infty, \infty)$ e que, em um pequeno intervalo dela, $x \in [0, 1]$, por exemplo, haja um ruído de alta freqüência. Na análise de Fourier, a representação deste ruído terá reflexos em todo o domínio de $\tilde{f}(x)$, uma vez que os senos e cossenos de alta freqüência, utilizados pela análise para representar o ruído, se estendem por todo o domínio. Já na análise por wavelets isto não ocorre. Nesta, o ruído também será representado por wavelets de alta freqüência, no entanto, devido a característica local destas funções, sua utilização terá reflexos apenas no intervalo dentro do qual o ruído está contido, ou, no máximo, em uma vizinhança curta do mesmo. Este fato torna as wavelets mais vantajosas de serem utilizadas (do que a análise de Fourier) para representar funções que apresentam alguma descontinuidade ou picos abruptos.

A análise por wavelets também fornece mais informações sobre o sinal do que a análise de Fourier. O estudo de uma função pelo método de Fourier nos mostra as diferentes freqüências que compõe o sinal, como pode ser visto na Figura 2.11(a). No entanto, ele não nos permite identificar nem quando, nem onde, cada uma destas freqüências ocorre. Já através de um estudo com wavelets, conseguimos saber ambas estas informações. Como uma wavelet é descrita por uma escala e uma posição no tempo, observando as wavelets que compõe a representação da função, podemos identificar exatamente quando, no tempo, as diferentes freqüências do sinal ocorrem, como pode ser visto na Figura 2.11(b).



(a) (b)

Figura 2.11: (a) Resultado da análise de Fourier (b) Resultado da análise por Wavelets.

Além de conter mais informações, em muitos casos, as representações feitas com wavelets são mais compactas do que aquelas feitas pela análise de Fourier [12]. Isto significa que são necessárias significativamente menos wavelets do que senos e cossenos para atingir uma aproximação comparável. O tempo de cálculo da transformada é outra vantagem das wavelets sobre Fourier. A da primeira tem complexidade computacional $O(n)$ enquanto a da segunda tem complexidade $O(n \cdot \log_2 n)$.

Estas e outras vantagens fazem das wavelets um excelente método para análise de dados. Dentre suas principais aplicações, podemos citar sua utilização na remoção de ruídos de séries temporais e na compactação de dados. Apenas como exemplo de sua importância, podemos citar que, no início da década de 90, o FBI [5] padronizou o uso de wavelets na compactação de imagens de impressões digitais [11, 12].

A seguir, introduzimos as transformadas contínua e discreta de wavelets, e ilustramos o algoritmo conhecido como transformada rápida de wavelets, utilizado para calcular a transformada discreta. Procuramos deixar bem claros os conceitos, apresentamos a base matemática por detrás dos métodos, e utilizamos exemplos sempre que necessário.

2.2.1 A Transformada Contínua de Wavelets

A equação (2.27) define a Transformada Contínua de Wavelets (CWT, do inglês *Continuous Wavelet Transform*). A CWT é a soma, sobre todo o tempo, do sinal multiplicado por versões escalonadas e transladadas da wavelet mãe. Este processo produz os coeficientes de wavelets que são funções de escala e posição. Ele é bem simples:

1. Selecione uma wavelet e compare-a a uma seção do início do sinal original.

2. Calcule um número C, que representa o quão correlacionados estão a wavelet e a seção em questão do sinal. Quanto maior for C, maior a similaridade entre ambas. Note que o resultado vai depender da forma da wavelet escolhida.

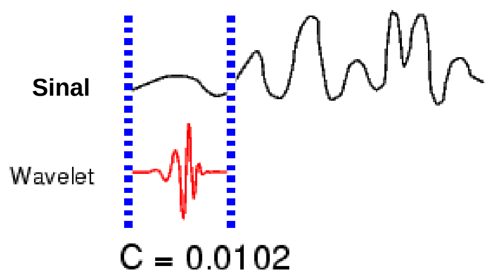


Figura 2.12: CWT passo 2

3. Translade a wavelet para a direita, e repita os passos 1 e 2 até que tenha

[5]Federal Bureau of Investigation

coberto todo o sinal.

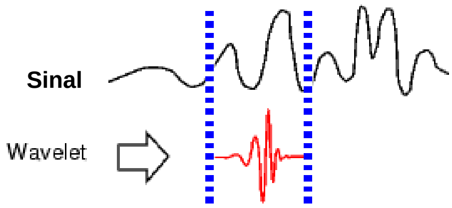


Figura 2.13: CWT passo 3

4. Dilate a wavelet, e repita os passos 1 a 3.

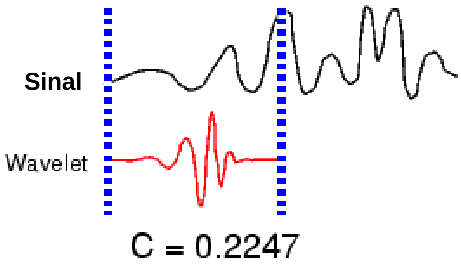


Figura 2.14: CWT passo 4

5. Repita os passos de 1 a 4 para todas as escalas (dilatações) possíveis.

Como resultado, teremos os coeficientes de wavelets produzidos por diferentes escalas e em diferentes seções do sinal.

O problema com esta transformada é que, por ser contínua, ou seja, por analisar o sinal em todas as escalas e posições possíveis, leva um tempo muito grande para ser calculada, e gera um conjunto enorme de dados. Precisamos então de um método que nos permita trabalhar apenas com um subconjunto destas escalas e posições. Isto nos leva a Transformada Discreta de Wavelets (DWT do inglês *Discrete Wavelet Transform*).

2.2.2 A Transformada Discreta de Wavelets

A Transformada Discreta de Wavelets recebe esse nome pois usa um subconjunto finito de escalonamentos e translações da wavelet mãe para construir a aproximação. Normalmente, opta-se por fazer o escalonamento das wavelets em intervalos inteiros múltiplos de potências de dois, pois tal escolha torna a análise mais eficiente e tão precisa quanto a transformada contínua. [13].

Ela utiliza, além da wavelet mãe ψ, uma função de escala (do inglês *scaling function*) ϕ, a qual ψ está associada[6]. Assim, a DWT decompõe a função original $f(x)$ em versões escalonadas e transladadas de ψ e ϕ.

A seguir, mostraremos, através de um exemplo, como a transformada discreta pode ser calculada de forma direta. Acreditamos que este exemplo ajudará a fixar os conceitos. Em seguida, introduzimos o algoritmo desenvolvido em 1989 por Stephane G. Mallat [14], conhecido como transformada rápida de wavelets (do inglês *Fast Wavelet Transform*), que é utilizado para calcular a DWT.

Aproximando Funções com Wavelets

Escolhemos para trabalhar, a mais simples e antiga de todas as wavelets, a wavelet de Haar $\psi(x)$. Ela é uma função degrau, definida da seguinte forma:

$$\psi(x) = \begin{cases} 1 & se \quad 0 \leq x < 0.5 \\ -1 & se \quad 0.5 \leq x < 1 \\ 0 & \text{caso contrário} \end{cases} \tag{2.28}$$

Sua representação gráfica pode ser vista na Figura 2.8(a).

Para fazer a decomposição, iremos utilizar versões escalonadas e trasladadas de $\psi(x)$, as quais representaremos por:

$$\psi_{j,k}(x) = \delta \cdot \psi(2^j x - k) \tag{2.29}$$

[6] A origem de ϕ, sua interpretação, e sua relação com ψ, serão explicadas mais adiante.

onde δ é uma constante, o termo k é responsável pela translação da wavelet, e o termo j, por sua escala.

[14] afirma que Yves Meyer mostrou em [15] que existe uma wavelet $\psi(x)$ tal que (2.29) define uma base ortogonal em $\mathbf{L}^2(\mathbb{R})$. Isto significa que qualquer elemento em $\mathbf{L}^2(\mathbb{R})$ pode ser representado como uma combinação linear (possivelmente infinita) destas funções base. A ortogonalidade de $\psi_{j,k}$ é simples de se verificar. Por inspeção, podemos ver que

$$\int \psi_{j,k} \cdot \psi_{j',k'} = 0$$

sempre que $j = j'$ e $k = k'$ não forem satisfeitos simultaneamente. Se $j \neq j'$, então os valores não nulos de uma wavelet estarão contidos na região constante da outra, o que tornará a integral igual a zero. Se $j = j'$ e, portanto, $k \neq k'$, então o produto será nulo. Assim, verificamos que as funções $\psi_{j,k}$ são ortogonais. O valor da constante δ que torna a base ortonormal[7] é $\delta = 2^{j/2}$.

Para a wavelet de Haar, $\phi(x)$ é definida como

$$\phi(x) = \begin{cases} 1 & se \quad 0 \leq x < 1 \\ 0 & \text{caso contrário} \end{cases} \tag{2.30}$$

Com ϕ e ψ definidas, estamos prontos para iniciar a expansão.

Considere $S = (s_0, s_1, \ldots, s_{2^n-1})$ nossa amostra, de tamanho 2^n, de um sinal qualquer. Este vetor pode ser associado à uma função f, gerada a partir de escalonamentos e translações de $\phi(x)$, no domínio $[0, 2^n)$

$$f(x) = \sum_{k=0}^{2^n-1} s_k \cdot \phi(x-k) \tag{2.31}$$

A decomposição em wavelets de $f(x)$ será da forma

$$\tilde{f}(x) = a_{-n,0}\phi_{-n,0}(x) + \sum_{j=-n}^{-1} \sum_{k=0}^{2^{j+n}-1} d_{j,k}\psi_{j,k}(x) \tag{2.32}$$

[7]Um conjunto $f_1, f_2, \ldots, f_n$ será ortonormal se $< f_j, f_k >= 1$ para $j = k$ e $< f_j, f_k >= 0$ para $j \neq k$; onde $<,>$ representa o produto interno, que, para funções pertencentes a $\mathbf{L}^2(\mathbb{R})$, é definido como $< f_j(x), f_k(x) >= \int_{-\infty}^{\infty} f_j(x) f_k(x) \mathrm{d}x$ [14].

onde $\phi_{-n,0}(x) = \phi(2^{-n}x - 0)$, $\tilde{f}(x)$ é uma aproximação de $f(x)$, e $a_{-n,0}$ e $d_{j,k}$ são os coeficientes associados à $\phi_{-n,0}(x)$ e $\psi_{j,k}(x)$, respectivamente.

Vamos fixar nossa amostra S e calcular a decomposição explicitamente. Seja $S = (1, 0, -3, 2, 1, 0, 1, 2)$. A função f correspondente pode ser vista na Figura 2.15.

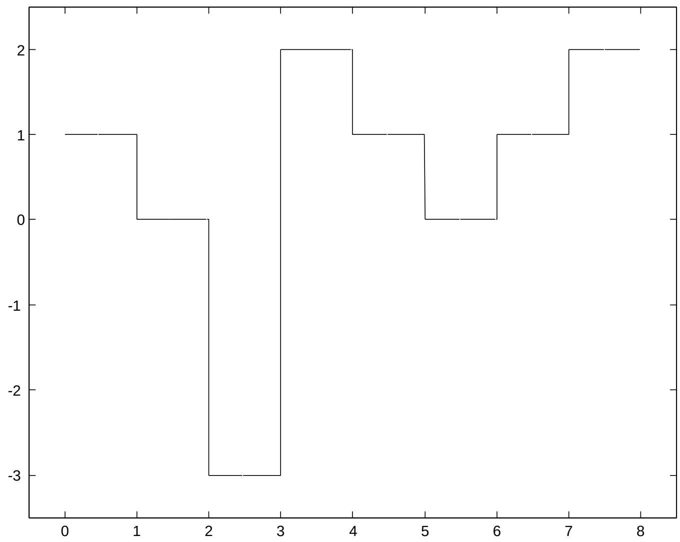


Figura 2.15: Função da amostra S em $[0, 8)$

Os coeficientes de (2.32) podem ser calculados diretamente, através da equação matricial a seguir. Observe que a matriz operadora é formada pela multiplicação da constante $\delta = 2^{j/2}$ pela wavelet correspondente, nos níveis de escala $j = -3, -2$ e -1. É importante ressaltar também, que cada equação descrita neste sistema é responsável pela representação de uma parte do sinal. Assim, a primeira equação descreve a função no intervalo $x \in [0, 1)$, a segunda no intervalo $x \in [1, 2)$ e assim sucessivamente, e é por isso que aparecem os elementos 0 matriz operadora. Estes são conseqüência do valor da wavelet utilizada no intervalo considerado. Por exemplo: a primeira equação descreve o sinal em $x \in [0, 1)$. Neste intervalo, as wavelets $\psi_{-2,1}(x), \psi_{-1,1}(x), \psi_{-1,2}(x), \psi_{-1,3}(x)$ têm valor zero. Portanto, os elementos da primeira linha da matriz operadora que correspondem a cada uma destas wavelets serão iguais a 0.

$$\begin{bmatrix} 1 \\ 0 \\ -3 \\ 2 \\ 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} a_{-3,0} \\ d_{-3,0} \\ d_{-2,0} \\ d_{-2,1} \\ d_{-1,0} \\ d_{-1,1} \\ d_{-1,2} \\ d_{-1,3} \end{bmatrix}$$

A solução é

$$\begin{bmatrix} a_{-3,0} \\ d_{-3,0} \\ d_{-2,0} \\ d_{-2,1} \\ d_{-1,0} \\ d_{-1,1} \\ d_{-1,2} \\ d_{-1,3} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ -\sqrt{2} \\ 1 \\ -1 \\ \frac{1}{\sqrt{2}} \\ -\frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Logo,

$$\begin{aligned} \tilde{f}(x) =& \sqrt{2}\phi_{-3,0}(x) - \sqrt{2}\psi_{-3,0}(x) + \psi_{-2,0}(x) - \psi_{-2,1}(x) + \frac{1}{\sqrt{2}}\psi_{-1,0}(x) \\ & - \frac{5}{\sqrt{2}}\psi_{-1,1}(x) + \frac{1}{\sqrt{2}}\psi_{-1,2}(x) - \frac{1}{\sqrt{2}}\psi_{-1,3}(x) \end{aligned} \tag{2.33}$$

A solução é facilmente verificável. Por exemplo, se $x \in [0,1)$

$$\tilde{f}(x) = \sqrt{2} \cdot \frac{1}{2\sqrt{2}} - \sqrt{2} \cdot \frac{1}{2\sqrt{2}} + 1 \cdot \frac{1}{2} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = 1 = f(x)$$

O problema de se calcular os coeficientes por esta forma direta é que, a medida que S vai crescendo, o custo computacional para resolver o sistema de equações se torna muito alto. Assim, foi desenvolvido em 1989, por Stephane Mallat [14] um algoritmo que calcula a DWT em n operações, dado um vetor de tamanho n, ou seja, sua complexidade é $O(n)$. Tal algoritmo é conhecido como Transformada Rápida de Wavelets (FWT, do inglês *Fast Wavelet Transform*) e é apresentado a seguir.

2.2.3 A Transformada Rápida de Wavelets

A Transformada Rápida de Wavelets foi desenvolvida por Stephane G. Mallat [14] e tem como base a análise multiresolucional. Dada uma função $f(x)$, uma decomposição multiresolucional produz sucessivas aproximações de $f(x)$ em diferentes escalas (ou resoluções). Seja $(r_j)_{j\in\mathbb{Z}}$ uma seqüência crescente de resoluções. Os detalhes de uma função, aproximada na resolução r_j, são definidos como a diferença entre sua aproximação em r_j e sua aproximação em r_{j-1}, onde r_{j-1} é uma resolução menor que r_j. É possível mostrar que estes detalhes podem ser representados por uma combinação linear de wavelets, de modo que, se formos sucessivamente produzindo aproximações de $f(x)$ em r_j e calculando sua diferença para a aproximação na resolução r_{j-1}, podemos obter uma decomposição da função original em wavelets.

A seguir, introduzimos os principais conceitos da análise multiresolucional e mostramos como, a partir desta, se obtêm a decomposição em wavelets de uma função de uma dimensão. Utilizamos a mesma notação de [14], e procuramos seguir os mesmos passos que ele apresenta, com o objetivo de facilitar ao leitor que deseja se aprofundar no assunto.

Análise Multiresolucional

É de conhecimento geral que, quanto maior a resolução[8] de uma foto, melhor é a qualidade de sua imagem. Isto ocorre pois o aumento da resolução implica no aumento da quantidade de amostras que são colhidas por unidade de área da imagem, o que permite que mais e mais detalhes possam ser representados. A Figura 2.16 ilustra um exemplo. Na primeira foto (a mais a esquerda), dispomos de 1 amostra para compor toda a área da imagem. Como se pode observar, com uma amostra apenas, não é possível fazermos qualquer distinção entre os elementos que

[8] O termo "resolução" pode ser entendido como o número de amostras por unidade de área que são utilizados na representação de uma imagem. Normalmente, a resolução também é descrita pelo número de amostras por linha vezes o número de amostras por coluna de uma imagem - 1024x768, por exemplo.

compõe a imagem e, portanto, não conseguimos representá-la. A medida que vamos aumentando o número de amostras colhidas, conseguimos diferenciar mais e mais detalhes da imagem, até que, finalmente, temos uma representação nítida da letra **R**.

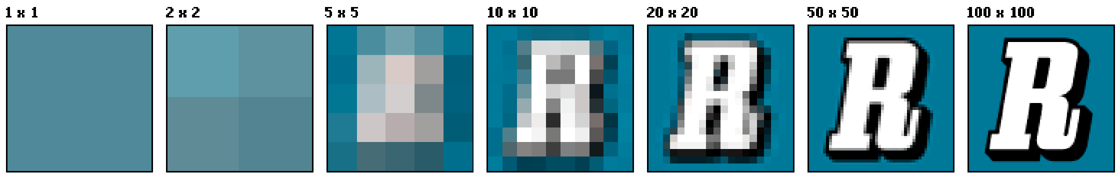


Figura 2.16: Representação de uma imagem sob diferentes resoluções.

Idealmente, gostaríamos de representar a imagem de forma contínua (ou com um número infinito de amostras), no entanto, isso não é possível, e assim, temos de discretizá-la.

O objetivo da análise multiresolucional é justamente estudar as informações de uma função sob diferentes resoluções. Ela procura separar o que é essencial do que é acessório. A grande vantagem deste tipo de análise é que ela nos permite ter a mesma interpretação da função independentemente de sua escala. Suponha, por exemplo, que estamos estudando a imagem de uma casa. Ao trabalhamos com uma resolução baixa, apenas suas características principais (seus objetos maiores, como a casa em si) são distinguíveis. A medida que a resolução cresce, mais e mais detalhes vão aparecendo (como as janelas, ou a maçaneta da porta de entrada). No entanto, sendo a resolução alta ou baixa, sabemos que estamos olhando para uma casa.

Aproximação Multiresolucional

Seja $r_j = 2^j$ uma resolução, e A_{2^j} um operador que aproxima uma função $f(x)$ na resolução 2^j. Assumimos que $f(x) \in \mathbf{L}^2(\mathbb{R})$. A_{2^j} pode ser caracterizado pelas seguintes propriedades:

1. A_{2^j} é linear. Se $A_{2^j} f(x)$ for a aproximação de $f(x)$ na resolução 2^j, então

$A_{2^j}f(x)$ não é modificado se o aproximarmos na mesma resolução 2^j novamente . Isto significa que $A_{2^j} \circ A_{2^j} = A_{2^j}$. Podemos entender A_{2^j} como um operador de projeção no espaço vetorial $\mathbb{V}_{2^j} \subset \mathbf{L}^2(\mathbb{R})$, que por sua vez, pode ser interpretado como o conjunto de todas as aproximações possíveis, na resolução 2^j, de funções em $\mathbf{L}^2(\mathbb{R})$.

2. Dentre todas as aproximações de $f(x)$ na resolução 2^j, $A_{2^j}f(x)$ é aquela que mais se assemelha a $f(x)$, i.e,

$$\forall g(x) \in \mathbb{V}_{2^j}, \qquad \| \, g(x) - f(x) \, \| \geq \| \, A_{2^j}f(x) - f(x) \, \| \, . \tag{2.34}$$

   Portanto, o operador A_{2^j} é uma projeção ortogonal no espaço vetorial $\mathbb{V}_{2^j}$.

3. A aproximação de uma função na resolução 2^{j+1} contém todas as informações necessárias para se aproximar a mesma função na resolução 2^j. Este princípio (conhecido como princípio da casualidade) é equivalente a

$$\forall j \in \mathbb{Z}, \qquad \mathbb{V}_{2^j} \subset \mathbb{V}_{2^{j+1}}. \tag{2.35}$$

4. A operação de aproximação é similar para todas as resoluções. Isto significa que os espaços de funções de aproximação devem ser derivados uns dos outros, escalonando-se cada função de aproximação pela taxa de seu valor de resolução:

$$\forall j \in \mathbb{Z}, \qquad f(x) \in \mathbb{V}_{2^j} \Leftrightarrow f(2x) \in \mathbb{V}_{2^{j+1}}. \tag{2.36}$$

5. A aproximação $A_{2^j}f(x)$ de $f(x)$ pode ser caracterizada por 2^j amostras por unidade de comprimento, e se $f(x)$ for transladada de uma quantidade proporcional a 2^{-j}, então $A_{2^j}f(x)$ também será transladada da mesma quantidade.

6. Ao computarmos uma aproximação de $f(x)$ na resolução 2^j, uma quantidade de informação sobre $f(x)$ é perdida. Entretanto, a medida que $j \to +\infty$, o sinal aproximado converge para o sinal original. Inversamente, quando $j \to -\infty$, e a resolução decresce a zero, aproximação contém menos e menos informações e converge para zero.

Denominamos qualquer conjunto de espaços vetoriais $(\mathbb{V}_{2^j})_{j\in\mathbb{Z}}$, definidos acima, uma aproximação multiresolucional de $\mathbf{L}^2(\mathbb{R})$.

Vimos que operador A_{2^j} é uma projeção ortogonal no espaço vetorial $\mathbb{V}_{2^j}$. Portanto, para caracterizá-lo numericamente, precisamos encontrar uma base ortonormal de $\mathbb{V}_{2^j}$. O teorema a seguir, tirado de [14], mostra que uma base deste tipo existe, e pode ser definida pela dilatação e translação de uma única função $\phi(x)$.

Teorema 1 *Seja* $(\mathbb{V}_{2^j})_{j\in\mathbb{Z}}$ *uma aproximação multiresolucional de* $\mathbf{L}^2(\mathbb{R})$. *Existe uma única função* $\phi(x) \in \mathbf{L}^2(\mathbb{R})$, *chamada de função de escala , tal que, se definirmos* $\phi_{2^j}(x) = 2^j\phi(2^jx)$ *para* $j \in \mathbb{Z}$ *(dilatação de* $\phi(x)$ *por* 2^j*), então*

$$\left(\sqrt{2^{-j}}\phi_{2^j}(x - 2^{-j}n)\right)_{n\in\mathbb{Z}}$$

forma uma base ortonormal de $\mathbb{V}_{2^j}$.

Podemos agora encontrar a projeção ortogonal de $f(x)$ em $\mathbb{V}_{2^j}$. Esta será dada pela combinação linear das bases definidas no Teorema 1:

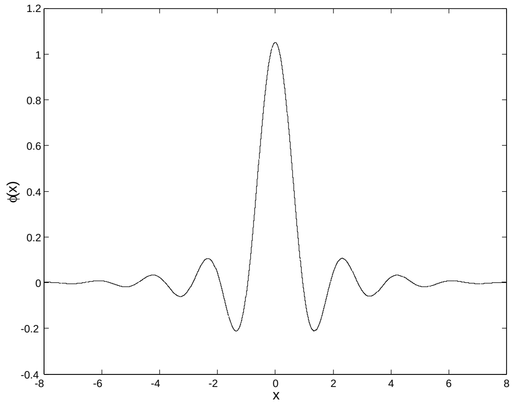
$$\forall f(x) \in \mathbf{L}^2(\mathbb{R}), \quad A_{2^j}f(x) = 2^{-j}\sum_{n=-\infty}^{+\infty}\left\langle f(u), \phi_{2^j}(u - 2^{-j}n)\right\rangle \phi_{2^j}(x - 2^{-j}n) \quad (2.37)$$

Observe que $A_{2^j}f(x)$ é formada pela soma ponderada de funções ϕ_{2^j}, cujos respectivos coeficientes são dados pelos produtos internos

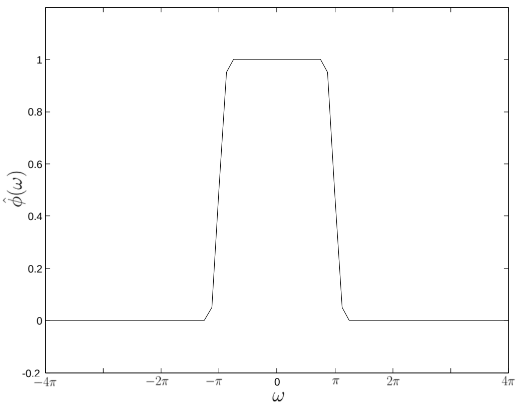
$$A^d_{2^j}f = \left(\left\langle f(u), \phi_{2^j}(u - 2^{-j}n)\right\rangle\right)_{n\in\mathbb{Z}} \quad (2.38)$$

Denominaremos de $A^d_{2^j}f$ a aproximação discreta de $f(x)$ na resolução 2^j.

A Figura 2.17, tirada de [14], ilustra um exemplo de uma função de escala $\phi(x)$ e sua transformada de Fourier. Olhando para 2.17(b), fica evidente que a função $\phi(x)$ é um filtro passa-baixo (do inglês *low-pass filter*), uma vez que é composta apenas por baixas freqüências. Portanto, a aplicação de A_{2^j}, na função $f(x)$, nada mais é que a aplicação de um filtro passa-baixo, que retira suas componentes de alta freqüência, seguido por uma amostragem uniforme à taxa 2^j.



(a)



(b)

Figura 2.17: (a) Exemplo de uma função de escala $\phi(x)$ e (b) sua transformada de fourier $\hat{\phi}(\omega)$. Observe que $\phi(x)$ é um filtro passa-baixo.

Nesta subseção, vimos como calcular a aproximação de $f(x)$ na resolução 2^j. A seguir, mostraremos como esta aproximação pode ser calculada a partir da aproximação em 2^{j+1}, e assim, construiremos um algoritmo piramidal para o cálculo das aproximações.

Transformada Multiresolucional

Na prática, não há como medirmos um sinal (uma função) de forma contínua; podemos apenas amostrá-lo de tempos em tempos. Portanto, os sinais dos fenômenos com os quais trabalhamos já se encontram aproximados em uma determinada resolução. Assumiremos, no restante desta seção, que nosso sinal original encontra-se aproximado na resolução 1. Intuitivamente, isto significa que nosso intervalo de amostragem corresponde à unidade de tempo. Assim, podemos dizer que nosso sinal original amostrado é uma aproximação de $f(x)$ na resolução 1, e é dado por $A_1^d f$.

Pelo princípio da casualidade, podemos calcular todas as aproximações $A_{2^j}^d f$, para $j < 0$, a partir de $A_1^d f$. Seja $(\mathbb{V}_{2^j})_{j\in\mathbb{Z}}$ uma aproximação multiresolucional cuja função de escala correspondente é $\phi(x)$. Pelo Teorema 1, a família de funções $\left(\sqrt{2^{-j-1}}\phi_{2^{j+1}}(x-2^{-j-1}k)\right)_{k\in\mathbb{Z}}$ forma uma base ortonormal para $(\mathbb{V}_{2^{j+1}})_{j\in\mathbb{Z}}$. Sabemos que, para todo $n \in \mathbb{Z}$, a função $\phi_{2^j}(x - 2^{-j}n)$ pertence a $\mathbb{V}_{2^j}$, que por sua vez, está incluído em $\mathbb{V}_{2^{j+1}}$. Logo, podemos expandí-la nas bases $\mathbb{V}_{2^{j+1}}$. Fazendo tal expansão, e mais algum algebrismo (que pode ser visto em [14]), podemos mostrar que

$$\big\langle f(u), \phi_{2^j}(u - 2^{-j}n)\big\rangle = \sum_{k=-\infty}^{+\infty} \big\langle \phi_{2^{-1}}(u), \phi(u-(k-2n))\big\rangle \cdot \big\langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1}k)\big\rangle \tag{2.39}$$

Seja H um filtro discreto cuja resposta é dada por

$$\forall n \in \mathbb{Z}, \quad h(n) = \big\langle \phi_{2^{-1}}(u), \phi(u-n)\big\rangle \tag{2.40}$$

e seja $\tilde{H}$ seu filtro espelho (do inglês *mirror filter*), cuja resposta é $\tilde{h}(n) = h(-n)$. Inserindo (2.40) em (2.39) obtemos

$$\big\langle f(u), \phi_{2^j}(u - 2^{-j}n)\big\rangle = \sum_{k=-\infty}^{+\infty} \tilde{h}(2n-k)\big\langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1}k)\big\rangle \tag{2.41}$$

que nos mostra que podemos obter $A^d_{2^j}f$ de $A^d_{2^{j+1}}f$, passando o último pelo filtro $\tilde{H}$ e mantendo uma de cada duas saídas[9]. É importante ressaltarmos que cada aproximação $A^d_{2^j}f$, para $j < 0$, terá 2^jN amostras, onde N é a quantidade de amostras do sinal original. Isto ocorre pois, quando passamos da resolução 2^{j+1} para a resolução 2^j, as funções ϕ_{2^j} dobram de tamanho, e, portanto, passamos a utilizar apenas uma função para representar um trecho do sinal que antes era aproximado por duas. Um esquema deste processo é ilustrado na Figura 2.18.

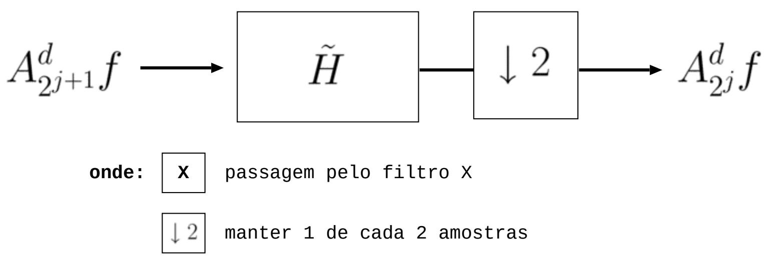


Figura 2.18: Esquema de obtenção da aproximação $A^d_{2^j}f$ a partir de $A^d_{2^{j+1}}f$.

Assim, mostramos como todas as aproximações $A^d_{2^j}f$, para $j < 0$, podem ser obtidas de $A^d_1 f$. A Figura 2.19 ilustra um exemplo de aproximações de um sinal

[9]Na nova resolução, o intervalo de amostragem dobrou. Este processo é conhecido como *down-sampling*

qualquer produzidas em três resoluções diferentes. Mais detalhes sobre as propriedades do filtro H podem ser vistas em [14].

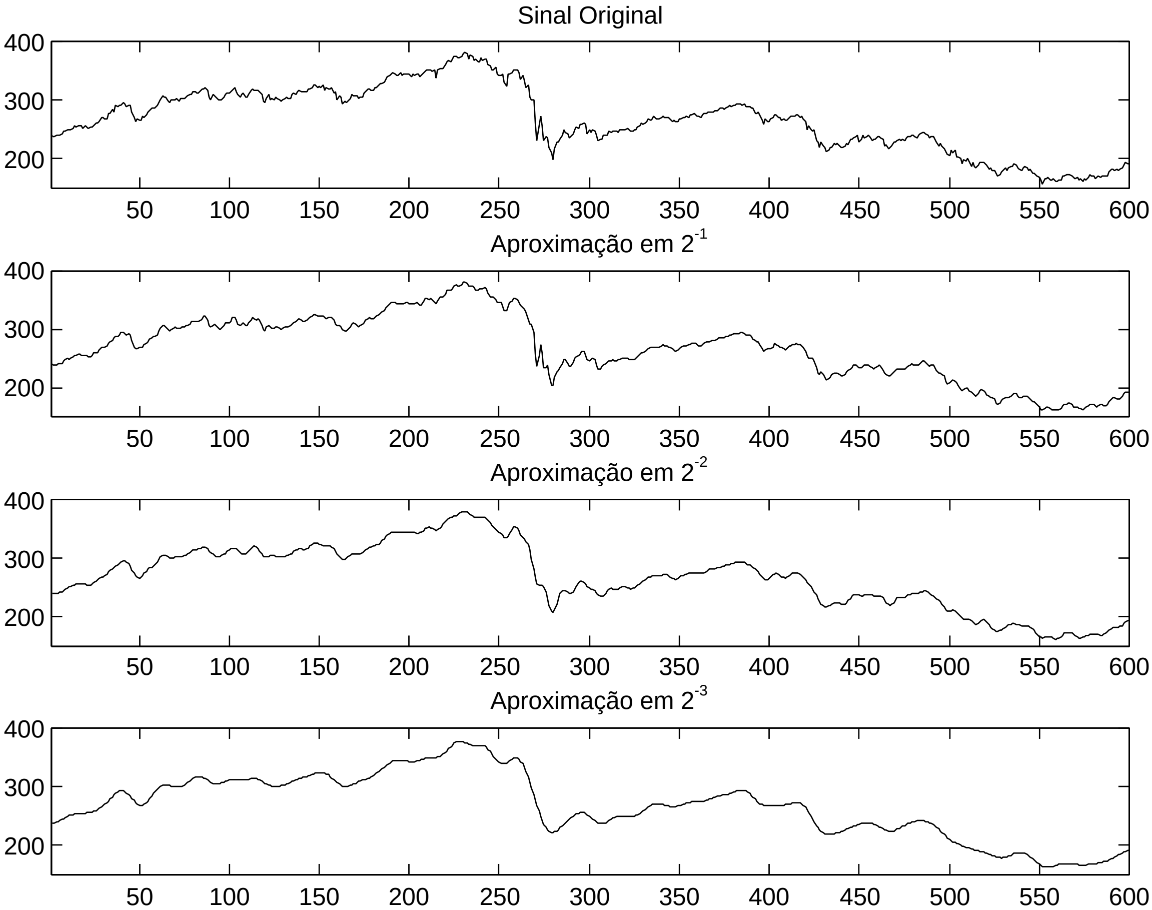


Figura 2.19: Exemplo de um sinal aproximado em três resoluções distintas.

A Representação por Wavelets

Conforme dissemos no início desta seção, a decomposição em wavelets será construída a partir da informação que se perde a cada aproximação do sinal nos níveis de resolução 2^j para $j \in \mathbb{Z}^-$. A informação perdida na resolução 2^j é definida como a diferença entre as aproximações nas resoluções 2^{j+1} e 2^j, e é denominada de **sinal de detalhe** (do inglês *detail signal*). Assim como uma aproximação na resolução 2^j, nada mais é do que uma projeção em $\mathbb{V}_{2^j}$, é possível mostrar que, o sinal de detalhe, na resolução 2^j, também é uma projeção ortogonal do sinal original, só que em $\mathbb{O}_{2^j}$, o complemento ortogonal de $\mathbb{V}_{2^j}$ em $\mathbb{V}_{2^{j+1}}$, isto é

$$\mathbb{O}_{2^j} \text{ é ortogonal a } \mathbb{V}_{2^j}$$

$$\mathbb{O}_{2^j} \oplus \mathbb{V}_{2^j} = \mathbb{V}_{2^{j+1}}$$

Do mesmo modo que fizemos para $\mathbb{V}_{2^j}$, precisamos encontrar uma base ortonormal para $\mathbb{O}_{2^j}$, a fim de computar a projeção ortogonal de $f(x)$ neste espaço. O teorema a seguir é muito semelhante ao Teorema 1, e mostra que uma base deste tipo existe, e pode ser formada por escalonamentos e translações de uma função $\psi(x)$. Sua prova pode ser encontrada em [14].

Teorema 2 *Seja $(\mathbb{V}_{2^j})_{j\in\mathbb{Z}}$ um espaço vetorial multiresolucional, $\phi(x)$ sua função de escala, e H seu filtro correspondente. Seja $\psi(x)$ uma função cuja transformada de Fourier é dada por*

$$\hat{\psi}(\omega) = G\Big(\frac{\omega}{2}\Big)\hat{\phi}\Big(\frac{\omega}{2}\Big)$$
$$\textit{com } G(\omega) = e^{-i\omega}\overline{H(\omega+\pi)} \tag{2.42}$$

e $\psi_{2^j}(x) = 2^j\phi(2^jx)$ a dilatação de $\psi(x)$ por 2^j. Então $\left(\sqrt{2^{-j}}\psi_{2^j}(x-2^{-j}n)\right)_{n\in\mathbb{Z}}$ é uma base ortonormal de $\mathbb{O}_{2^j}$ e

$$\left(\sqrt{2^{-j}}\psi_{2^j}(x-2^{-j}n)\right)_{(n,j)\in\mathbb{Z}^2}$$

é uma base ortonormal de $\mathbf{L^2}(\mathbb{R})$, e $\psi(x)$ é denominada uma **wavelet ortogonal** *(do inglês orthogonal wavelet).*

O Teorema 2 mostra, também, como podemos construir uma wavelet: partimos de um filtro H (cuja transformada de Fourier deve satisfazer a determinadas condições descritas em [14]), encontramos sua função de escala $\phi(x)$ correspondente e, usando a equação (2.42), determinamos a wavelet $\psi(x)$.

Seja P_{2^j} o operador de projeção ortogonal em $\mathbb{O}_{2^j}$. A decomposição de $f(x)$ em wavelets na resolução 2^j, que nos fornece o sinal de detalhe na mesma, é, portanto, dada por

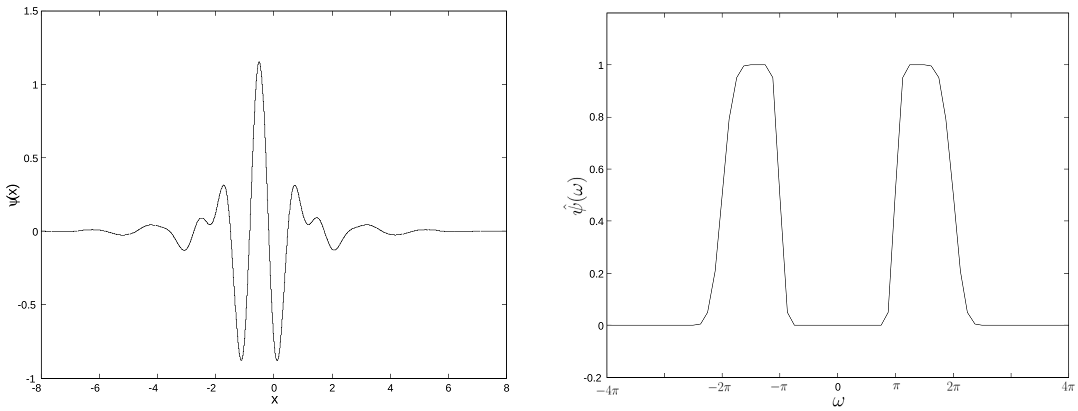
$$P_{2^j}f(x) = 2^{-j}\sum_{n=-\infty}^{+\infty}\left\langle f(u), \psi_{2^j}(u-2^{-j}n)\right\rangle\, \phi_{2^j}(x-2^{-j}n) \tag{2.43}$$

Analogamente à $A_{2^j}f$, $P_{2^j}f$ é caracterizado pelos produtos internos $\langle f(u), \psi_{2^j}(u - 2^{-j}n)\rangle$. Assim, podemos definir

$$D_{2^j}f = \Big(\langle f(u), \psi_{2^j}(u - 2^{-j}n)\rangle\Big)_{n\in\mathbb{Z}} \tag{2.44}$$

como o **sinal de detalhe discreto** (do inglês *discrete detail signal*) de $f(x)$ na resolução 2^j. Ele contém a diferença de informação entre $A^d_{2^{j+1}}f$ e $A^d_{2^j}f$.

A Figura 2.20(a) mostra a wavelet $\psi(x)$ associada à função de escala $\phi(x)$ da Figura 2.17. Em 2.20(b) podemos observar que a wavelet é um filtro passa-faixa (do inglês *band-pass filter*), cujas bandas de freqüência são, aproximadamente, iguais à $[-2\pi, -\pi] \cup [\pi, 2\pi]$. Assim, o sinal de detalhe $D_{2^j}f$ descreve $f(x)$ nas bandas de freqüência $[-2^{-j+1}\pi, -2^{-j}\pi] \cup [2^{-j}\pi, 2^{-j+1}\pi]$. Mais detalhes sobre como, dada uma wavelet, conseguimos representar todas as freqüências de uma função, podem ser encontrados em [16]. Podemos agora montar uma descrição completa de $A^d_1 f$, nossa



(a) (b)

Figura 2.20: (a) Exemplo de uma wavelet $\psi(x)$ e (b) sua transformada de fourier $\hat{\psi}(\omega)$. Observe que $\psi(x)$ é um filtro passa-faixa.

amostra original do sinal $f(x)$. Para $J > 0$, esta representação tem a forma

$$\left(A^d_{2^{-J}}f, (D_{2^j}f)_{-J\leq j\leq -1}\right) \tag{2.45}$$

Este conjunto de sinais discretos é denominado uma **representação ortogonal por wavelets** (do inglês *orthogonal wavelet representation*). Note que esta representação consiste em um sinal de referência, dado pela aproximação $A_{2^{-J}}f$ na resolução 2^{-J} mais os detalhes nas resoluções 2^j para $-J \leq j \leq -1$.

Transformada de Wavelets

Da mesma forma que fizemos para $A^d_{2^j}f$, podemos demonstrar que $D_{2^j}f$ pode ser calculado a partir de $A^d_{2^{j+1}}f$. Como $\psi_{2^j}(x-2^{-j}n)$ pertence à $\mathbb{O}_{2^j} \subset \mathbb{V}_{2^{j+1}}$, podemos expandí-la nas bases de $\mathbb{V}_{2^{j+1}}$. Fazendo tal expansão e mais alguns algebrismos, chegamos a

$$\big\langle f(u), \psi_{2^j}(u-2^{-j}n)\big\rangle = \sum_{k=-\infty}^{+\infty} \big\langle \psi_{2^{-1}}(u), \phi(u-(k-2n))\big\rangle \cdot \big\langle f(u), \phi_{2^{j+1}}(u-2^{-j-1}k)\big\rangle \tag{2.46}$$

Seja G um filtro discreto cuja resposta é dada por

$$g(n) = \big\langle \psi_{2^{-1}}(u), \phi(u-n)\big\rangle \tag{2.47}$$

e $\tilde{G}$ o filtro simétrico a G, com resposta $\tilde{g}(n) = g(-n)$. Inserindo (2.47) em (2.46) temos

$$\big\langle f(u), \psi_{2^j}(u-2^{-j}n)\big\rangle = \sum_{k=-\infty}^{+\infty} \tilde{g}(2n-k)\big\langle f(u), \phi_{2^{j+1}}(u-2^{-j-1}k)\big\rangle \tag{2.48}$$

que nos mostra que podemos obter o sinal de detalhe $D_{2^j}f$ passando $A^d_{2^{j+1}}f$ pelo filtro $\tilde{G}$ e mantendo uma de cada duas saídas. Assim, a representação em wavelets de um sinal discreto pode ser computada através da decomposição sucessiva de $A^d_{2^{j+1}}f$ em $A^d_{2^j}f$ e $D^d_{2^j}f$, para $-J \leq j \leq -1$. Um esquema deste algoritmo é ilustrado na Figura 2.21.

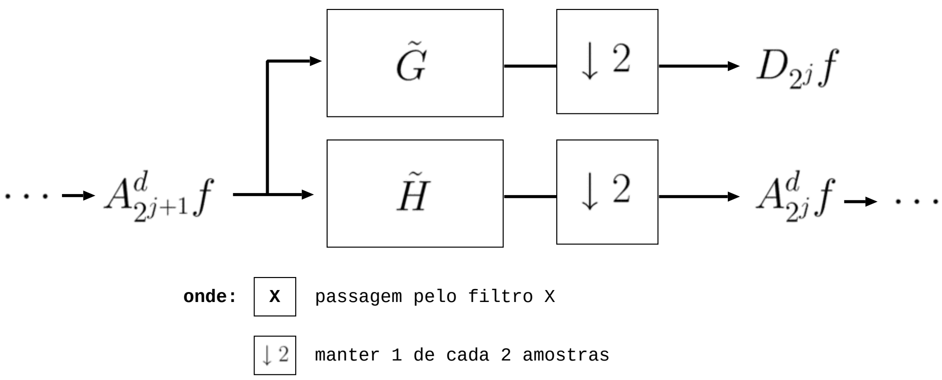


Figura 2.21: Esquema da transformada discreta de wavelets.

A equação (2.42) do Teorema 2 nos mostra que os filtros G e H estão relacionados. Mais especificamente

$$g(n) = (-1)^{1-n} h(1-n) \tag{2.49}$$

G é o filtro espelho (do inglês *mirror filter*) de H, e do tipo passa-alto (do inglês *high-pass filter*). Na área de processamento de sinais, G e H são conhecidos como *quadrature mirror filters*. A equação (2.48) pode ser interpretada como uma filtragem passa-alto do sinal $A^d_{2^{j+1}}f$.

Se o sinal original é composto por N amostras, então cada um dos sinais discretos $A^d_{2^j}f$ e $D_{2^j}f$ têm $2^j N$ amostras[10]. Portanto, a representação por wavelets (2.45) contém a mesma quantidade de amostras, N, de $A^d_1 f$, nossa amostragem original de $f(x)$. A Figura 2.22 ilustra os detalhes obtidos das aproximações do sinal utilizado na Figura 2.19. Observe que, nos pontos onde $A_{2^{j+1}}f$ e $A_{2^j}f$ são significativamente diferentes, os detalhes apresentam uma grande amplitude.

[10]Não esqueça que $j < 0$!

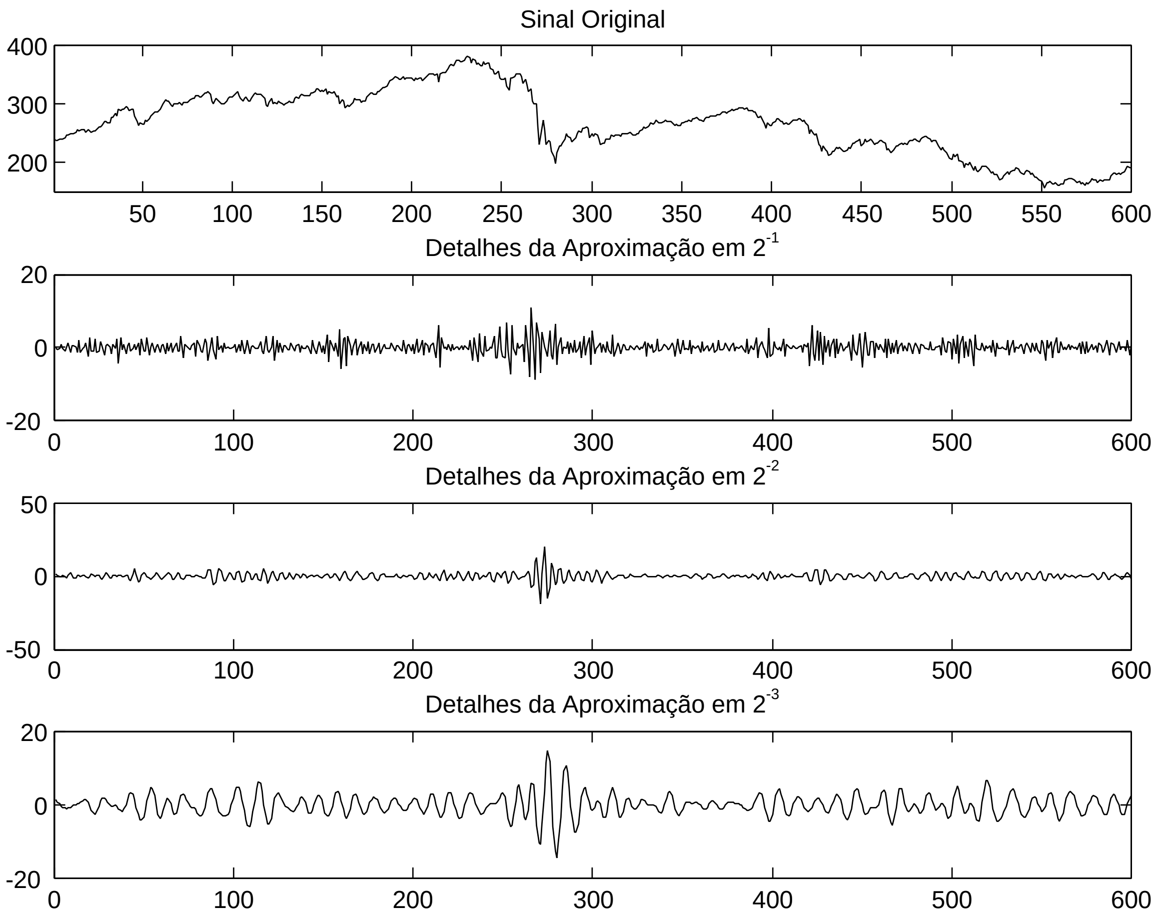


Figura 2.22: Exemplo de detalhes de um sinal aproximado em três resoluções distintas.

Um Exemplo

A fim de clarear os conceitos e o algoritmo descrito anteriormente, ilustramos, a seguir, o processo da transformada discreta de wavelets através de um exemplo. Escolhemos trabalhar com o mesmo sinal utilizado no início desta seção, e, novamente, com a wavelet de Haar. Suponha que $S = (1, 0, -3, 2, 1, 0, 1, 2)$ seja nossa amostra, discreta, do sinal $f(x)$ a ser estudado. Sabemos que a aproximação de S na resolução 2^j é composta de $2^j N$ amostras. Portanto, a resolução mínima com a qual podemos trabalhar é $2^{-3} = 1/8$, na qual nossa aproximação será composta de, apenas, uma amostra (uma função ϕ).

O algoritmo para calcular a DWT consiste em, iterativamente, ir passando a aproximação $A_{2^{j+1}}$ pelos filtros $\tilde{H}$ e $\tilde{G}$, até que tenhamos uma aproximação na resolução mínima 2^{-J}, que, no caso, ocorre para $J = 3$. A Figura 2.23 ilustra o processo completo da transformada.

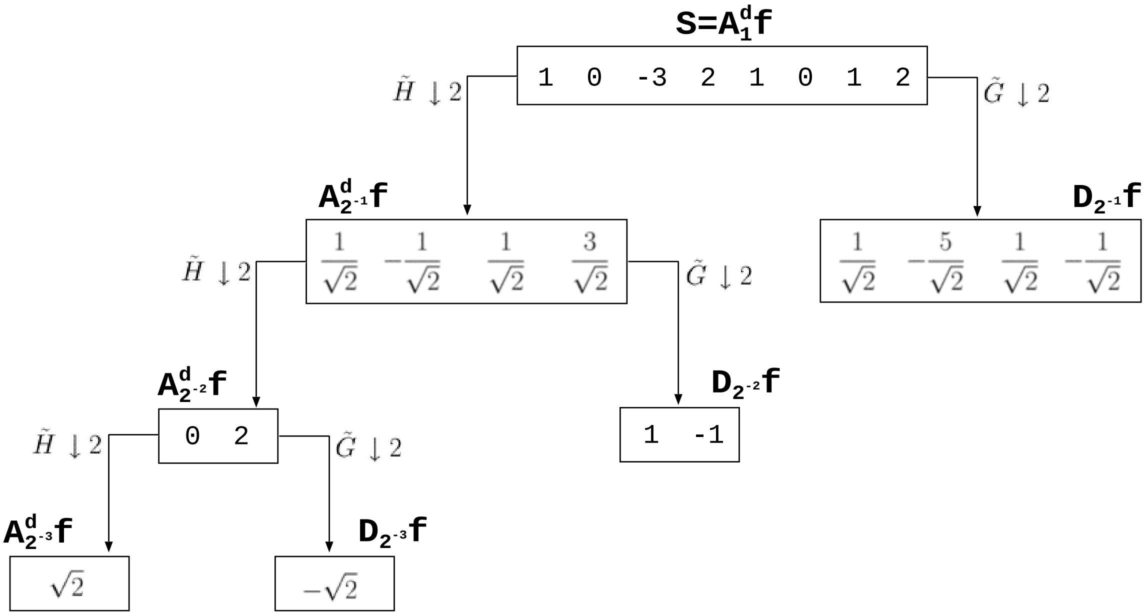


Figura 2.23: Transformada discreta de S utilizando a wavelet de Haar.

Obtemos, desta forma, como resultado da transformada, os coeficientes

$$\left(\sqrt{2}, (-\sqrt{2}), (1, -1), (\frac{1}{\sqrt{2}}, -\frac{5}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})\right) \tag{2.50}$$

idênticos aos obtidos em (2.33) quando os calculamos diretamente.

2.2.4 A Transformada Rápida Inversa de Wavelets

Assim como construímos uma representação em wavelets de um sinal discreto S, podemos, através de um algoritmo piramidal, reconstruir o sinal original S dada sua decomposição em wavelets. Vimos que $\mathbb{O}_{2^j}$ é o complemento ortogonal de $\mathbb{V}_{2^j}$ em $\mathbb{V}_{2^{j+1}}$, portanto, $\left(\sqrt{2^{-j}}\phi_{2^j}(x - 2^{-j}n), \sqrt{2^{-j}}\psi_{2^j}(x - 2^{-j}n)\right)_{n \in \mathbb{Z}}$ é uma base ortonormal de $\mathbb{V}_{2^{j+1}}$. Se decompusermos $\phi_{2^{j+1}}(x - 2^{-j-1}n)$ nesta base, utilizarmos os filtros H

e G, e fizermos algum algebrismo (que pode ser visto em [14]), chegamos a

$$\left\langle f(u), \phi_{2^{j+1}}(u - 2^{-j-1}n) \right\rangle = 2 \sum_{k=-\infty}^{+\infty} h(n-2k) \left\langle f(u), \phi_{2^j}(u - 2^{-j}k) \right\rangle + 2 \sum_{k=-\infty}^{+\infty} g(n-2k) \left\langle f(u), \psi_{2^j}(u - 2^{-j}k) \right\rangle \quad (2.51)$$

A equação (2.51) nos mostra que $A^d_{2^{j+1}}f$ pode ser reconstruída se colocarmos zeros entre cada amostra de $A^d_{2^j}f$ e $D_{2^j}f$, passarmos os sinais resultantes pelos filtros H e G respectivamente, e somarmos os resultados. A Figura 2.24 ilustra este esquema.

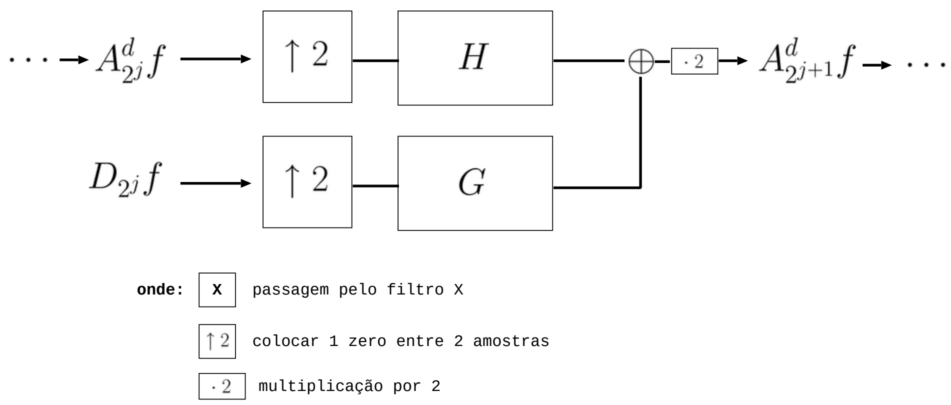


Figura 2.24: Esquema da transformada discreta inversa de wavelets.

A Figura 2.25 ilustra o processo de reconstrução de nosso sinal $S = (1, 0, -3, 2, 1, 0, 1, 2)$, utilizado nos exemplos anteriores, a partir de sua decomposição calculada com a transformada discreta de wavelets, na qual trabalhamos com a wavelet de Haar.

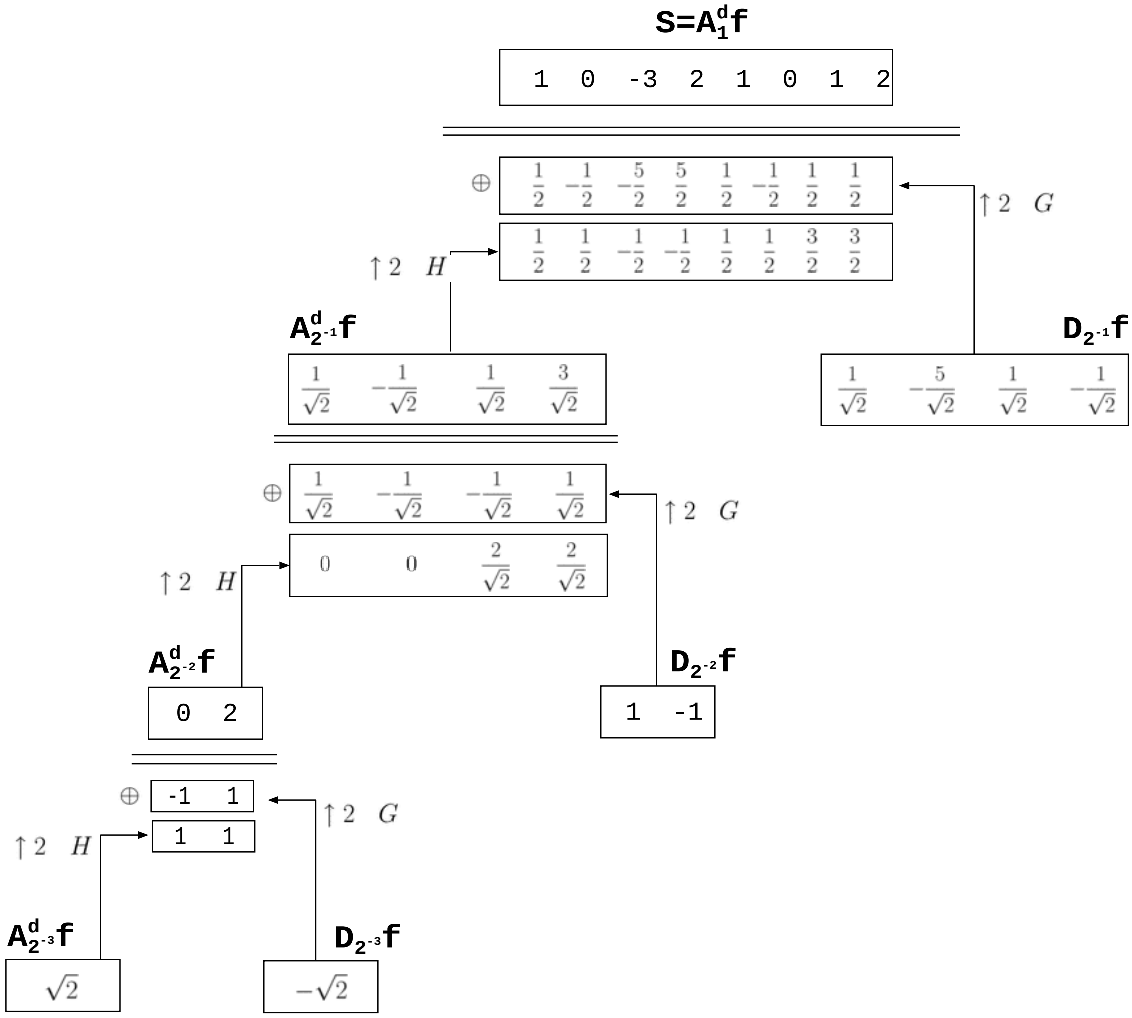


Figura 2.25: Transformada discreta inversa de S utilizando a wavelet de Haar.

2.2.5 Remoção de Ruídos

Na seção anterior, vimos que a DWT decompõe uma função através de sucessivas suavizações, pela aplicação do filtro $\tilde{H}$, e que, a cada suavização, os detalhes são retirados pelo filtro $\tilde{G}$. Ao final, a função é representada pela soma de diversas wavelets, cujos coeficientes estão diretamente associados à intensidade dos detalhes. Se estes detalhes forem pequenos, podemos omití-los, sem que isso altere significativamente as características principais do sinal. Para tal, elaboramos uma política (conhecida como *thresholding*) que muda para zero cada coeficiente de wavelet cujo valor seja menor que um determinado *threshod* λ. Assim, esperamos "limpar"a função original, retirando detalhes não importantes, que são considerados ruído.

Como exemplo, vamos usar novamente nossa amostra $S = (1, 0, -3, 2, 1, 0, 1, 2)$. A DWT de S nos dá, como resultado, o vetor de coeficientes $(\sqrt{2}, -\sqrt{2}, 1, -1, \frac{1}{\sqrt{2}}, -\frac{5}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. Se decidirmos que todos os coeficientes menores que 0.9 representam ruído, e os substituirmos por zero, então o vetor resultante será $(\sqrt{2}, -\sqrt{2}, 1, -1, 0, -\frac{5}{\sqrt{2}}, 0, 0)$. A Figura 2.26 ilustra a aproximação $\tilde{f}(x)$ após a remoção do ruído.

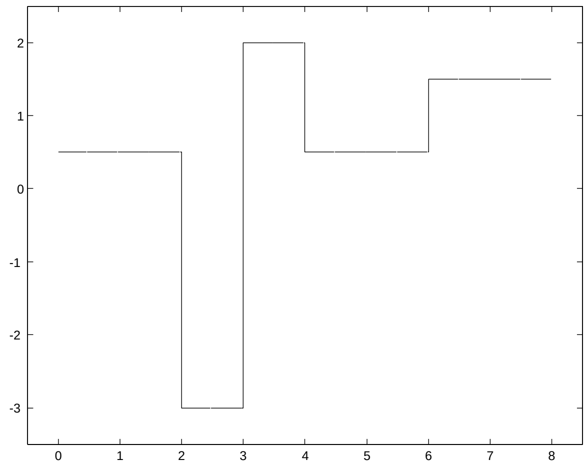


Figura 2.26: Função $\tilde{f}(x)$ após a remoção de ruídos. Compare-a com a Figura 2.15.

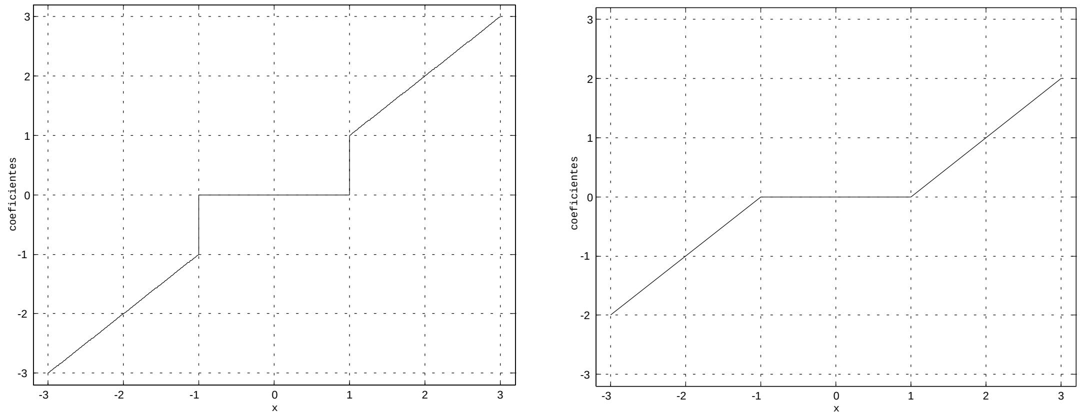
O processo de *thresholding* pode ser dividido em duas partes. Na primeira, é escolhida a função de *threshod* T. As duas opções são: *hard* e *soft thresholding*, cujas políticas são dadas por

$$T^{hard}(d, \lambda) = \begin{cases} d \quad se \quad |d| > \lambda \\ 0 \quad \text{senão} \end{cases}$$

$$T^{soft}(d, \lambda) = \begin{cases} d - \left(\frac{d}{|d|}\lambda\right) \quad se \quad |d| > \lambda \\ \qquad 0 \quad \text{senão} \end{cases}$$

e suas diferenças podem ser vistas na Figura 2.27

O próximo passo é a escolha do *treshold* λ. Existem diversos métodos para se fazer essa escolha, dentre os quais podemos citar o *Universal Treshold* [12] e SURE (*Stein's Unbiased Estimator of Risk*) [12]. O primeiro usa um *threshold* fixo $\lambda^U = \sqrt{2 \log n}\hat{\sigma}$ e sua idéia central é remover todos os coeficientes menores que o



(a) (b)

Figura 2.27: Políticas de *thresholding*: (a) hard thresholding e (b) soft thresholding.

máximo valor esperado de uma amostra de uma variável aleatória i.i.d de tamanho n. O segundo estima o risco de um determinado *threshold* λ, minimiza esse risco, e então encontra o valor de λ.

Capítulo 3

Modelos de Previsão de Séries Financeiras

Existem, basicamente, dois tipos de modelos que são utilizados na previsão de variáveis econômicas. Os primeiros, chamados de modelos de regressão, relacionam a variável a ser estudada a um conjunto de outras variáveis, as quais, acredita-se, influenciam o comportamento da primeira. Por exemplo, podemos tentar prever uma taxa de juros de curto prazo usando um modelo de regressão que relaciona-a ao PIB, aos preços, e à oferta monetária, caso acreditemos que o comportamento destas três variáveis explique, de forma satisfatória, o comportamento da taxa de juros. Existem diversos problemas neste tipo de abordagem, dentre os quais podemos citar a possível falta de dados sobre todas as variáveis explanatórias. Isto ocorre principalmente com variáveis que são influenciadas por um enorme conjunto de fatores sobre os quais pouco se conhece, como o clima meteorológico, mudanças de gosto, ou movimentos de manada desencadeados por fatores psicológicos. Ainda há casos em que, para obtermos uma previsão a partir de uma equação de regressão, é necessário fazer uma previsão das próprias variáveis explanatórias, o que pode ser mais difícil do que prever a variável principal.

As dificuldades encontradas nos modelos de regressão levaram ao desenvolvimento de um segundo tipo de modelo, o chamado modelo de séries temporais. Este tenta inferir o comportamento de uma variável observando apenas seu passado. Ele assume que existem padrões de comportamento que se repetem ao longo do tempo, e que estes, uma vez identificados, podem ser utilizados para se construir uma previsão.

O capítulo a seguir descreve, de forma sucinta, alguns dos modelos de previsão de séries financeiras mais comuns encontrados na bibliografia. Dedicamos uma seção a cada modelo, na qual fazemos uma breve descrição sobre suas principais características e, em seguida, citamos alguns trabalhos que o utilizam na previsão do preço do petróleo, e seus resultados.

3.1 Regressão sobre Variáveis Fundamentalistas

O principal fundamento responsável pela flutuação do preço de um ativo ou de uma *commodity* é o balanço entre sua demanda e sua oferta. No caso do petróleo, este balanço é ditado basicamente por duas atividades: exploração/produção e refino/comercialização. A primeira é responsável por garantir a oferta, e tem como seu principal personagem a organização dos países exportadores de petróleo, conhecida como OPEP. A segunda é responsável pela demanda. As refinarias compram óleo cru e o refinam, produzindo os derivados de petróleo que são então comercializados às indústrias e à população. Apesar de aparentemente simples, esse balanço é bastante complexo, pois depende de uma série de fatores voláteis e, muitas vezes, imprevisíveis. Guerras e tensões políticas podem rapidamente fazer desaparecer parte da oferta, como ocorreu durante o primeiro choque do petróleo em 1973, ocasião na qual a OPEP decidiu suspender suas exportações para os países que apoiavam Israel na guerra de Yom Kippur. O crescimento econômico (ou recessão) de países provoca alterações na demanda; invernos mais quentes ou mais frios mudam a demanda por determinados derivados; furacões podem destruir plataformas produtoras cortando temporariamente a oferta; inflação; desvalorização do dólar, etc.

A quantidade (e imprevisibilidade) de fatores que influenciam o preço do petróleo torna sua previsão por regressão de fundamentos uma tarefa difícil. Entretanto, há uma variável que, teoricamente, resume de forma satisfatória o balanço entre a demanda e a oferta: o nível dos estoques de petróleo. Devido aos riscos existentes do lado da oferta, e da volatilidade de preço, uma grande parte da indústria consumidora de petróleo mantêm estoques do mesmo. Os estoques, além de oferecerem proteção contra eventuais aumentos de preço, também permitem maior flexibilidade e velocidade de resposta em caso de aumento repentino da demanda. Assim, em teoria, o nível destes estoques funciona como um medidor do equilíbrio entre a demanda e a oferta. Um estudo mais profundo da relação entre estoques e preços pode ser visto em [17].

Partindo do pressuposto acima, [18] constrói um modelo de previsão mensal do preço do petróleo fazendo uma regressão deste sobre o nível relativo de estoques dos países da Organização para Cooperação e Desenvolvimento Econômico, a OECD (do inglês *Organisation for Economic Co-operation and Development*). O nível relativo de estoques é definido como a diferença entre o nível real e um nível normal, determinado historicamente por uma regressão. A equação de regressão proposta por [18] é

$$\begin{aligned} \mathrm{WTI}_t = \alpha_0 + \alpha_1 \mathrm{WTI}_{t-1} + \sum_{i=0}^{k} \beta_i \mathrm{RIN}_{t-i} + \sum_{i=0}^{k} (\lambda_i \mathrm{LIN}_{t-i} + \lambda\lambda_i \mathrm{LIN2}_{t-i}) + \\ \sum_{i=0}^{k} (\delta_i \mathrm{HIN}_{t-i} + \delta\delta_i \mathrm{HIN2}_{t-i}) + \sum_{i=0}^{6} \phi_i \mathrm{DSEP01}_i + \gamma \mathrm{LAPR99} + \varepsilon_t \end{aligned} \tag{3.1}$$

onde $\alpha_0, \alpha_1, \beta_i, \lambda_i, \lambda\lambda_i, \delta_i, \delta\delta_i, \phi_i, \gamma$ são os parâmetros a serem estimados; WTI_t é o preço do WTI no tempo t; RIN é o estoque relativo; LIN, LIN2, HIN, HIN2 são variáveis não-lineares referentes aos níveis baixo e alto dos estoques; DSEP01 e LAPR99 variáveis *dummy* que capturam os efeitos de 11 de setembro de 2001 e abril de 1999, respectivamente, e ε_t um erro. Este modelo é uma evolução de outros mais simples, propostos pelos mesmos autores em [19, 20]

O modelo acima apresenta resultados preditivos satisfatórios. Entretanto, ele tem o inconveniente de necessitar do nível dos estoques no tempo t, para fazer a previsão do preço no tempo t, de modo que a regressão é dependente da previsão de uma de suas variáveis explanatórias[1].

Recentemente, [22] argumentou que, a partir do ano de 2004, esta relação entre preço e níveis de estoque parece ter se enfraquecido, o que é observado pelo surgimento de um prêmio nos preços reais sobre os preços previstos pelo modelo (3.1). O autor, então, testa novas variáveis na tentativa de explicar este descolamento do preço. Seus resultados indicam que o aumento da especulação parece ser a principal causa do surgimento deste prêmio.

3.2 Modelos Lineares de Séries Temporais

A seguir, descrevemos, de forma breve, alguns dos modelos de séries temporais lineares mais conhecidos. Eles se aplicam tanto aos processos estacionários[2] quanto aos não-estacionários homogêneos[3]. Uma discussão mais abrangente e detalhada de cada modelo pode ser encontrada em [23, 24].

3.2.1 Passeio Aleatório

O modelo de série temporal mais simples é o passeio aleatório (RW, do inglês *random walk*). Na sua forma mais comum, cada mudança sucessiva em y_t é extraída, independentemente, de uma distribuição de probabilidade com média 0 e variança σ^2

$$y_t = y_{t-1} + \varepsilon_t \tag{3.2}$$

onde ε_t tem média 0 e variança σ^2. Este modelo é normalmente usado como referência de comparação no teste de outros modelos.

[1] A *US Energy Information Administration* (EIA) [21] fornece, gratuitamente, previsões dos estoque de petróleo da OECD.

[2] Um processo é dito estacionário quando as características do processo estocástico subjacente não mudam com o tempo. Uma definição mais formal pode ser vista em [23, 24].

[3] Um processo é dito não-estacionário homogêneo quando pode ser diferenciado uma ou mais vezes para produzir um processo estacionário.

Se a hipótese de eficiência de mercado é correta, então espera-se que o preço de uma mercadoria siga um passeio aleatório, de modo que nenhum investidor pode obter lucros excepcionais seguindo alguma regra para tomar suas decisões de compra e venda. De fato, tal hipótese é, normalmente, assumida para os mercados financeiros à vista, como o mercado de ações. Entretanto, [24] afirma que o mercado *spot* do petróleo não segue um passeio aleatório, porém faz a ressalva de que isto não significa que um investidor possa obter retornos excepcionais na comercialização desta mercadoria.

[25] utiliza um modelo RW como referência no teste de seu modelo de Redes Neurais para a previsão do preço futuro do petróleo. Seus resultados indicam que o mercado de petróleo não é eficiente, e que há como explorar esta ineficiência usando modelos matemáticos não-lineares, como as redes neurais.

3.2.2 Modelo de Média Móvel

No processo de médias móveis de ordem q, cada observação de $y(t)$ é gerada por uma média ponderada de q perturbações aleatórias ε_t. Denotamos este processo como MA(q) e o definimos como

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \tag{3.3}$$

onde $\{\varepsilon_t\}$ é uma seqüência de variáveis aleatórias, não correlacionadas, com média 0 e variança σ^2, e $\{\theta_t\}$ seus respectivos coeficientes. Seqüências deste tipo são conhecidas como ruído branco (WN do inglês *white noise*) e, para representá-las, usamos a notação $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$.

3.2.3 Modelos Auto-Regressivos

No processo auto-regressivo de ordem p, a observação corrente y_t é gerada por uma média ponderada de p observações anteriores mais uma perturbação aleatória

ε_t. Denotamos este processo como AR(p) e o definimos como

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \delta + \varepsilon_t \tag{3.4}$$

onde $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$ e δ está relacionado com a média μ do processo da seguinte forma:

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \tag{3.5}$$

3.2.4 Modelos Auto-Regressivos e de Médias Móveis

Muitos dos processos aleatórios estacionários não podem ser modelados apenas como médias móveis ou apenas como auto-regressivos, pois possuem qualidades de ambos os tipos de processo. Assim, surge uma extensão lógica que combina-os, formando o modelo auto-regressivo e de médias móveis de ordem (p, q), onde p é a ordem da auto-regressividade e q a ordem da média móvel. Denotamos este processo como ARMA(p, q) e o definimos como

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.6}$$

onde δ é definido em (3.5) e $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$

3.2.5 Modelos ARIMA

Os modelos ARMA assumem que a série temporal modelada é estacionária. Entretanto, na prática, muitas das séries com as quais trabalhamos não são estacionárias, mas podem ser transformadas em séries estacionárias quando as diferenciamos uma ou mais vezes. Dizemos que y_t é homogênea estacionária de ordem d se

$$w_t = \Delta^d y_t \tag{3.7}$$

é uma série estacionária. Δ é o operador de diferenciação, definido como

$$\Delta y_t = y_t - y_{t-1} \qquad \Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \tag{3.8}$$

e assim por diante. Dada a série w_t, podemos voltar a y_t pela soma de w_t num total de d vezes, da forma $y_t = \sum^d w_t$.

Uma vez que a série y_t tiver sido diferenciada, produzindo w_t, podemos modelar w_t como um processo ARMA. Se $w_t = \Delta^d y_t$ e w_t é um processo ARMA(p, q), então dizemos que y_t é um processo auto-regressivo integrado e de médias móveis de ordem (p, d, q) ou simplesmente ARIMA(p, d, q).

Em [26] é utilizado um modelo ARIMA$(1, 1, 0)$ para prever o preço do petróleo ao final de um horizonte de tempo de um mês. Seus resultados desencorajam a utilização do ARIMA pois, por ser um modelo linear, não consegue capturar a dinâmica não-linear do processo, que é bastante presente. Os resultados são avaliados em termos do erro médio quadrático e de um índice de acerto de direção, chamado de D_{stat}. Este último indica a taxa de acerto da previsão da direção do preço. Num período de teste que vai de Janeiro de 2000 a Dezembro de 2003, o ARIMA obtém uma D_{stat} de 54%, inferior a taxa de acerto de outros modelos não-lineares testado no mesmo artigo. Um resultado semelhante é verificado para o erro médio quadrático.

3.2.6 Modelos GARCH

Ao construirmos um modelo de regressão, em alguns casos, há motivos para se acreditar que a variância do erro muda ao longo do tempo, e é dependente dos erros passados. Em tais processos, fica evidente que ocorre uma aglomeração, no tempo, de erros grandes e pequenos. Por exemplo, ao modelarmos um ativo financeiro, como o preço de uma ação ou de uma *commodity* como o petróleo, fica muito claro que há períodos que apresentam alta volatilidade, e outros nos quais a volatilidade é baixa. Nestes casos, dizemos que há um tipo de heterocedasticidade presente, em que a variância do erro da regressão depende da volatilidade dos erros no passado recente.

O modelo *generalized autoregressive conditional heterocedasticity*, conhecido como GARCH, foi introduzido para melhorar a regressão de variáveis que possuem característica heterocedástica. Ele relaciona a variância do erro ao tamanho da volatilidade observada em períodos recentes e sua variância. Um modelo GARCH(p, q) é definido como

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \lambda_1 \sigma_{t-1}^2 + \cdots + \lambda_q \sigma_{t-q}^2 \tag{3.9}$$

onde σ_t^2 é a variância do erro no tempo t, e ε_t o erro no tempo t. A estimação dos parâmetros do GARCH é feita juntamente com a estimação dos parâmetros da regressão.

A utilização de modelos GARCH é muito comum na modelagem de ativos financeiros que usam dados diários ou semanais, os quais quase sempre apresentam o tipo de heterocedasticidade descrito acima [24]. Em [27] é utilizado um modelo GARCH para construir intervalos de confiança sobre a previsão feita pelo contratos futuros de um mês do preço do óleo tipo Brent. O trabalho afirma que os contratos futuros não conseguem prever corretamente nem a direção do preço, acertando, em média, apenas 50% das vezes. Entretanto, seu desempenho como preditor pode ser melhorada construindo-se intervalos de confiança da previsão, o que é feito utilizando o GARCH. Um aumento do intervalo de confiança indica que a capacidade de previsão dos contratos futuros deve cair.

3.3 Modelos não Lineares de Séries Temporais

3.3.1 Redes Neurais e Sistemas Especiais

Uma rede neural artificial (ANN, do inglês *artificial neural network*) é um modelo matemático não-linear, baseado nas redes neurais biológicas. Ela é representada por um grupo de nós (ou neurônios), interconectados, pelos quais flui informação, de maneira semelhante aos neurônios no cérebro humano. As ANN foram introduzidas em 1943 e são, hoje em dia, bastante utilizadas na área de reconhecimento de padrões.

Uma ANN é composta de camadas, cada qual contendo um certo número de nós. As principais camadas são as de entrada, por onde entra informação na rede, e saída, por onde saem as respostas da rede. A camada de entrada pode afetar diretamente a camada de saída, entretanto, o mais comum é que haja uma ou mais camadas intermediárias, conhecidas como camadas ocultas. A Figura 3.1 ilustra esta estrutura.

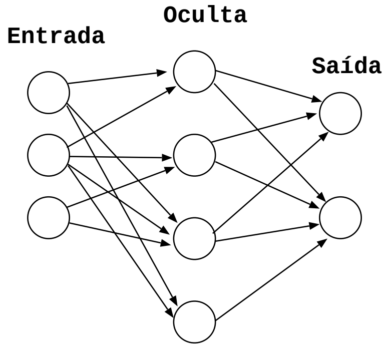


Figura 3.1: Estrutura em camadas de uma ANN

Seja $X = x_i$ o vetor das variáveis de entrada de uma ANN. Suponha que a ANN tenha uma camada oculta de nós h_j e apenas uma variável de saída y. O j-ésimo nó da camada oculta é definido como

$$h_j = f_i\left(\alpha_{0j} + \sum w_{ij}x_i\right) \tag{3.10}$$

onde x_i corresponde a i-ésima variável de entrada. A constante $\alpha_{0,j}$ e os pesos w_{ij} precisam ser estimados. A função f_i é normalmente conhecida como função de ativação, e determina (a partir de um *threshold*) se a saída do nó atual deve ser propagada ao nó seguinte ou não. A camada de saída pode, então, ser definida como

$$y = \alpha_{00} + \sum_{i=1}^{I} \alpha_{i0}x_i + \sum_{j=1}^{J} w_{j0}h_j \tag{3.11}$$

onde I = número de variáveis de entrada e J = número de nós na camada oculta. Note que, nesta rede, os valores de entrada afetam diretamente os valores de saída. Seus parâmetros, normalmente, são estimados utilizando um algoritmo de treinamento iterativo. A cada passo, o algoritmo vai ajustando seus valores de forma a reduzir alguma medida de erro (como o erro médio quadrático) das saídas dadas pela rede em relação as saídas reais. Este processo é chamado de aprendizado da rede. Uma discussão mais profunda e detalhada sobre redes neurais pode ser vista em [28].

Uma busca rápida na bibliografia mostra que as ANN têm sido uma ferramenta bastante utilizada na previsão de séries financeiras devido a sua capacidade de reconhecimento de padrões. [25] utiliza uma ANN para emitir sinais diários de compra e venda no mercado *spot* de petróleo. Seus resultados indicam que é possível obter ganhos extraordinários utilizando a ANN. Em [29] uma ANN é utilizada em conjunto com outras duas ferramentas de inteligência artificial, uma *ruled based expert system* (RES) e uma *web-based text minig* (WTM), para a previsão mensal do preço do petróleo. Enquanto a ANN é utilizada para fazer a previsão dos preços, as outras duas ferramentas são utilizadas na busca (na Internet) de fatos significativos que por ventura possam ter ocorrido no passado recente. Com o resultado da busca, a previsão é então ajustada. A lista de fatos é pré-selecionada, e pode ser vista em [29]. Ela inclui, entre outros, guerras, embargos e crises econômicas. Os resultados encorajam a utilização de um sistema de previsão como este.

Em [26], um tipo especial de algoritmo de rede neural chamado de *support vector machine* (SVM) é utilizado também na previsão do preço do petróleo. Ele compara seus resultados aos de uma ANN normal e ao modelo ARIMA, e conclui a favor do SVM, ressaltando, no entanto, que a ANN também tem significativa capacidade preditiva.

3.3.2 Modelos de Markov Ocultos

A teoria por trás dos modelos de Markov ocultos está descrita no capítulo 2. Em [30] os autores utilizam uma cadeia de Markov oculta com observações contínuas para a previsão de séries financeiras como, o índice S&P 500 e da taxas de câmbio. Em cada estado oculto, eles utilizam um modelo de emissão de símbolos, como um AR ou uma rede neural. Assim, ao invés da matriz de emissão B, há um novo conjunto de parâmetros θ_B que é composto pelos parâmetros do modelo de emissão utilizado. O modelo fornece como saída uma distribuição da variação percentual do ativo, e usa sua média como valor previsto. O fato de se ter a distribuição também permite que seja feito uma análise de risco das previsões. Seus resultados são comparados aos de uma rede neural *feed-forward*, à um modelo auto-regressivo e a uma estratégia de *buy-and-hold*, e indicam que a HMM tem um desempenho superior aos três no período testado.

O trabalho de [31] se baseia em [30] e faz uma modificação no algoritmo de EM para que as observações mais recentes tenham maior peso no treinamento da HMM. Seus resultados indicam que a HMM consegue gerar retornos acima do mercado, e tem desempenho superior a rede neural.

Capítulo 4

Metodologia de Previsão

Este capítulo apresenta, em detalhes, a metodologia utilizada para prever a distribuição do retorno[1] acumulado pelo preço do petróleo, ao final de uma janela de tempo futura de F dias. Ela é composta de cinco etapas, e utiliza as wavelets como ferramenta de suavização, e um modelo de Markov oculto como ferramenta de previsão. Seu fluxograma é ilustrado na Figura 4.1.

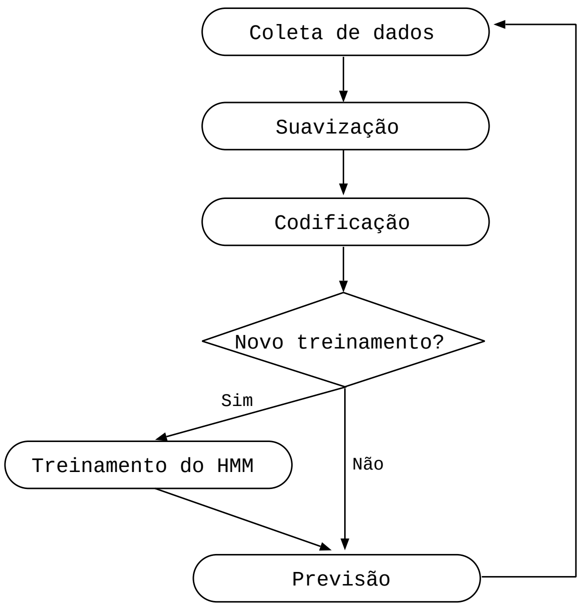


Figura 4.1: Fluxograma do processo de previsão.

[1] Aqui, e no restante desta dissertação, deve-se entender como retorno de um ativo sua variação percentual de preço em um período de tempo definido.

O processo tem início com a coleta dos dados. Como estamos trabalhando com um modelo de séries temporais, a única variável de interesse é o preço do petróleo; logo, esta primeira etapa se resume a obter sua série de preços. Em seguida, suavizamo-la, utilizando o método de remoção de ruído por wavelets. Na terceira etapa, obtemos a série de retornos a partir da série suavizada de preços e codificamo-os, associando a cada retorno um símbolo do conjunto de símbolos do HMM. Caso haja necessidade, os parâmetros do modelo são reestimados. Por último, é feita a previsão. A seguir, descrevemos individualmente cada uma destas etapas.

4.1 Coleta de Dados

Existem, aproximadamente, 161 tipos de petróleo negociados no mundo [21]. O preço de cada um depende, principalmente, de suas características, como a quantidade de enxofre presente e sua gravidade API. Escolhemos trabalhar com o *West Texas Intermidiate* (WTI), pois ele é a principal referência de óleo cru da América, e uma das principais do mundo. Utilizamos nesta dissertação os preços *spot* diários[2], nominais, referentes à um barril do WTI. Seu histórico pode ser adquirido na página da *Energy Information Administration* (EIA) [21], que é o centro oficial de estatísticas energéticas do governo dos Estados Unidos.

Além da série histórica dos preços, a página da EIA contém diversas outras informações importantes relacionadas a área energética, incluindo previsões, estudos, análises e notícias. Recomendamo-la ao leitor interessado como uma excelente fonte de dados.

4.2 Suavização da Série de Preços

As séries de preços de ativos financeiros e *commodities* tendem a ser bastante voláteis, e a série do WTI não é nenhuma exceção. No ano 2006, por exemplo, seu

[2]definidos como o preço de fechamento do dia.

preço no dia 27 de abril era de 70.76 dólares, muito próximo ao preço de 17 de abril, que era de 70.30 dólares. Entretanto, durante este pequeno intervalo de tempo, de apenas dez dias, o preço oscilou entre uma mínima de 67.43 e uma máxima de 73.73 dólares.

A Figura 4.2 mostra a autocorrelação da série de retornos diários do WTI nos 200 primeiros *time lags*, calculada com a ferramenta de modelagem Tangram2 [32]. As linhas horizontais, obtidas pelo resultado de Barlett[3], indicam o limite acima do qual o valor de um coeficiente é estatisticamente significativo (diferente de 0) com nível de significância de 95%. Apesar de os valores encontrados serem um indício de que retornos passados podem conter alguma informação a respeito dos retornos futuros, eles são muito baixos, de modo que estas informações são, aparentemente, pouco relevantes.

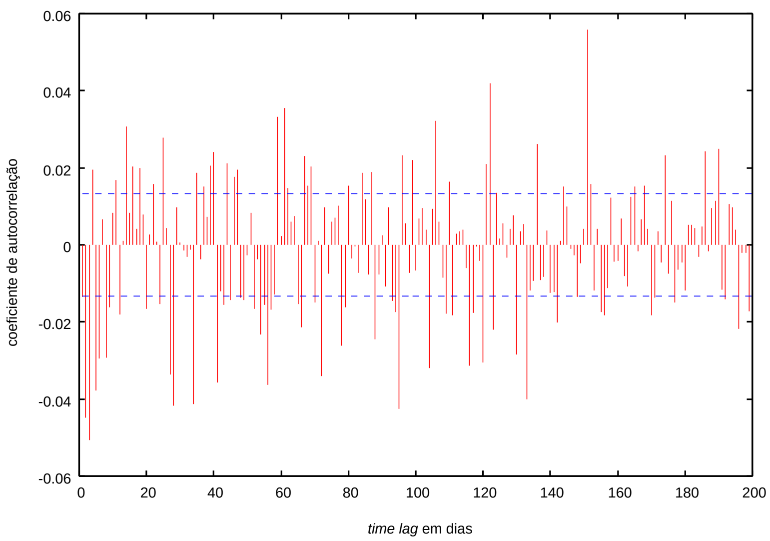


Figura 4.2: Autocorrelação dos retornos diários do WTI.

Conforme mencionado no capítulo 1, não estamos interessados em prever esta volatilidade diária, de curtíssimo prazo, mas em identificar e prever as tendências dos preços em intervalos maiores, de, por exemplo, 20 dias. Assim, procuramos suavizar

[3]Se uma série temporal é gerada por um ruído branco, os coeficientes da função de correlação amostral (para $k > 0$, onde k é o *lag* da correlação) seguem, aproximadamente, uma distribuição normal com média 0 e desvio padrão $\frac{1}{\sqrt{T}}$; onde T é o número de observações da série.

a série, a fim de eliminar esta volatilidade (considerada como ruído) e ressaltar a informação que consideramos relevante.

Existem diversas técnicas para se suavizar séries temporais. Talvez a mais comum e simples delas seja a média móvel. A suavização por média móvel consiste em representar cada ponto da série original pela média aritmética dos n últimos pontos. Assim, temos que uma série suavizada $\tilde{y_t}$ é dada por

$$\tilde{y_t} = \frac{1}{n}(y_t + y_{t-1} + \cdots + y_{t-n+1}) \tag{4.1}$$

Há algumas variações desta técnica, como a média móvel exponencial (que dá mais peso aos valores mais recentes) e a média móvel exponencial dupla. Uma discussão mais detalhada de cada um destes métodos pode ser vista em [24].

O principal problema da média móvel é seu deslocamento no tempo em relação a série original. Se o número de dias utilizados no cálculo da média for pequeno, ela acompanha mais rapidamente as mudanças de tendência, no entanto, ao mesmo tempo, se torna mais sensível aos ruídos, e a suavização é de pior qualidade. Se aumentarmos o número de dias, a eliminação de ruídos é maior, entretanto, a média móvel demora mais para refletir as mudanças de tendência. Neste caso, ela também não consegue capturar satisfatoriamente movimentos abruptos, intensos, de curto intervalo de tempo que, porventura, podem ser importantes. A Figura 4.3 ilustra estes problemas.

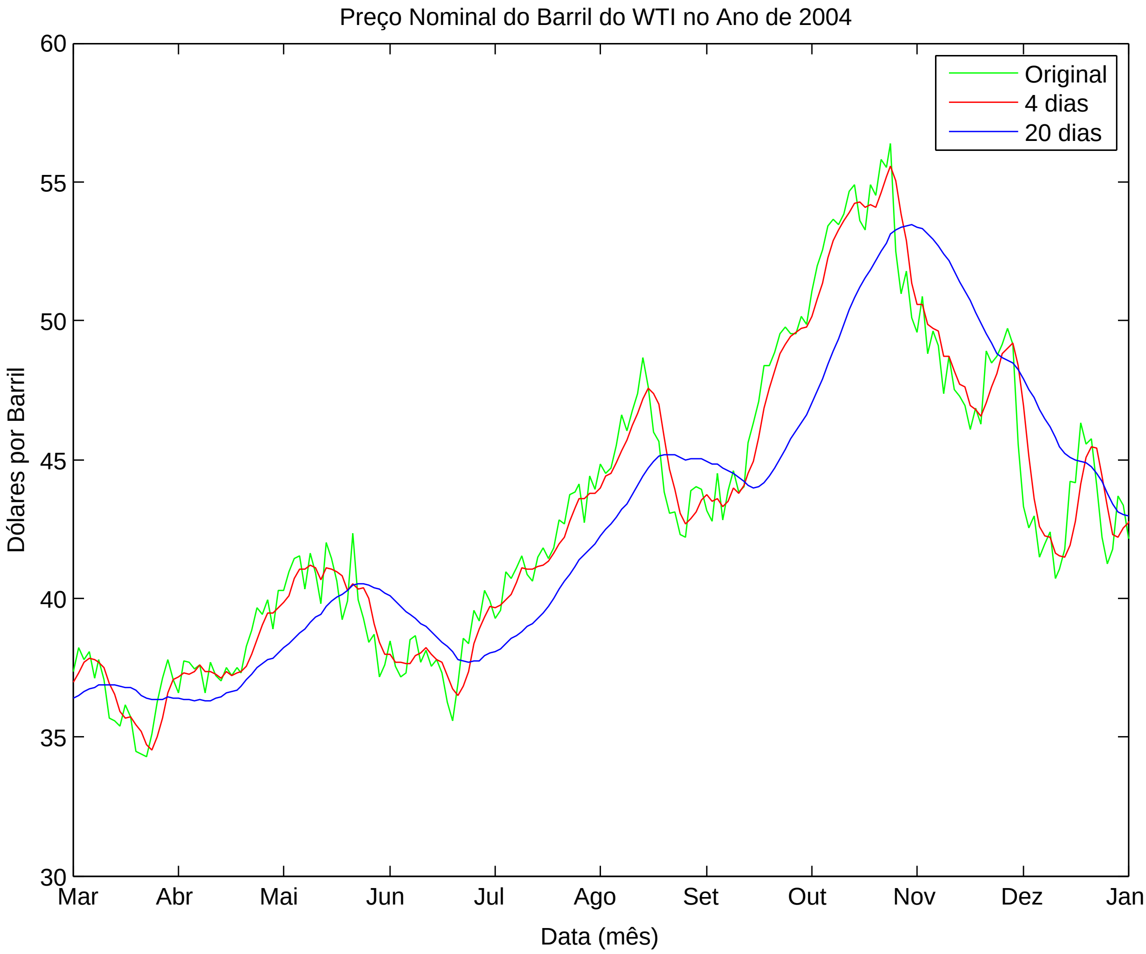
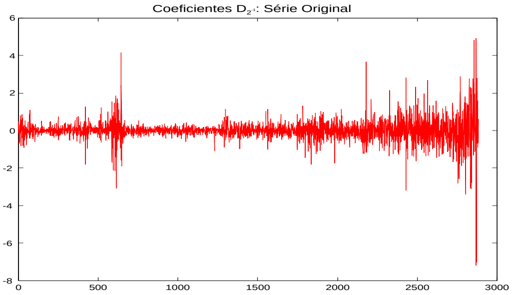


Figura 4.3: Série de preços nominais do WTI e sua média móvel de 4 e 20 dias.

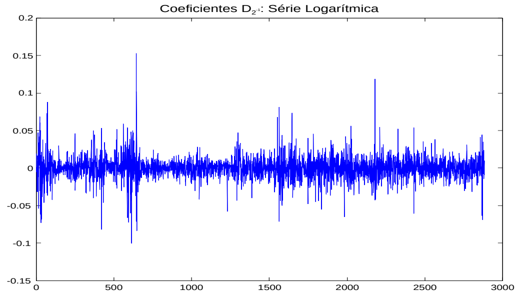
Nesta dissertação, decidimos fazer a suavização da série de preços do WTI utilizando as wavelets. Como já foi visto no capítulo 2, a característica local destas garante que a série suavizada não fique deslocada em relação a série original, como acontece com as médias móveis. Além disto, a remoção de ruídos pelas wavelets consegue capturar movimentos significativos de curta duração melhor que as médias móveis.

O primeiro passo na suavização é obter a série de logaritmos dos preços. Isto é necessário pelo seguinte: sabemos do capítulo 2, que a remoção de ruídos utilizando as wavelets é feita através da eliminação dos coeficientes de wavelet menores que um determinado *threshold* λ, e que eles representam a diferença entre duas aproximações consecutivas do sinal durante a transformada discreta. Agora observe que estes coeficientes tendem a ser maiores quanto mais alto for o preço, uma vez que um desvio de 1 dólar sobre um preço de 100 dólares é o mesmo que um desvio de 0.1 dólares sobre um preço de 10 dólares. Assim, se suavizarmos a série original de preços, ocorrerá que os coeficientes que representam níveis baixos de preços serão eliminados em maior quantidade do que os que representam níveis mais altos, de modo que teremos como resultado uma série cuja suavização é maior nas épocas onde o petróleo custava menos (em termos nominais), e menor nas épocas onde custava mais.

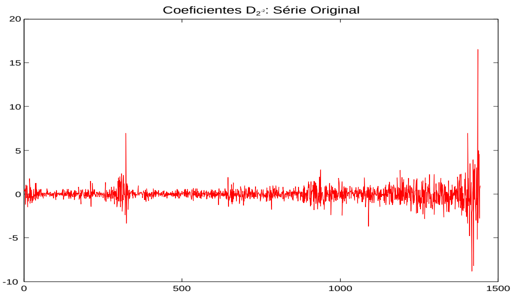
Entretanto, se suavizarmos a série logarítmica dos preços, conseguimos eliminar este problema. A Figura 4.4 mostra os coeficientes de wavelet $D_{2^{-1}}$ e $D_{2^{-2}}$, obtidos na decomposição das séries de preços original e logarítmica do WTI, utilizando a wavelet de Daubechies de ordem 3. Para relembrar como tais coeficientes são obtidos, indicamos rever a Figura 2.23. Note que os coeficientes a direita dos gráficos 4.4(a) e 4.4(c), que representam os preços mais altos da série, são maiores que os da esquerda, que representam preços menores. Observe como isto não acontece com os coeficientes da série de log dos preços, ilustrados nas figuras 4.4(b) e 4.4(d)



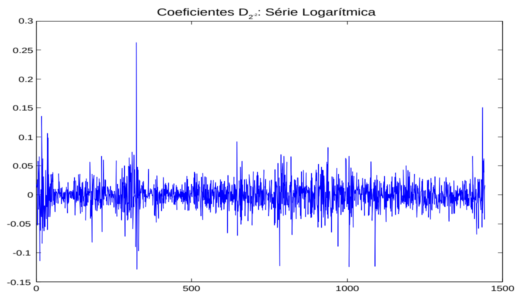
(a)



(b)



(c)



(d)

Figura 4.4: Coeficientes de detalhes obtidos na suavização do WTI: (a) série de preços original, resolução 2^{-1}; (b) série de preços logarítmica, resolução 2^{-1}; (c) série de preços original, resolução 2^{-2}; (d) série de preços logarítmica, resolução 2^{-2}.

A suavização da série do log dos preços é feita por intermédio da função *wden()* do programa Matlab [13]. Após a suavização, obtemos a série de preços original suavizada utilizando função inversa do log. O processo pode ser descrito da seguinte forma:

$$X \to \log(X) \to \mathrm{wden}(\log(X)) = \log(X_s) \to \log^{-1}(\log(X_s)) \to X_s \tag{4.2}$$

onde X é a série de preços original e X_s é a série de preços suavizada. A Figura 4.5 ilustra ambas estas séries. Escolhemos para esta suavização a wavelet de Daubechies de ordem 3 e resolução de decomposição 2^{-4}. A justificativa para esta escolha será dada mais adiante, na subseção 5.1.1

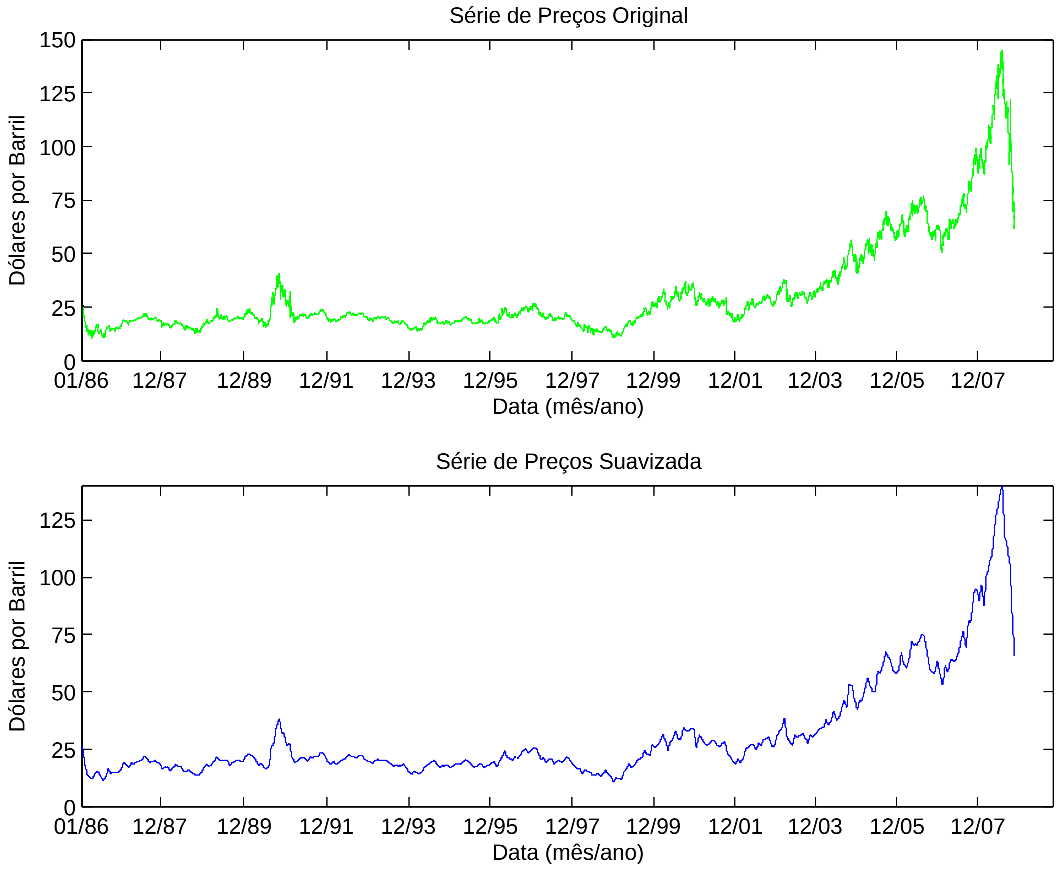


Figura 4.5: Série de preços original e suavizada do WTI, obtida com a wavelet Daubechies de ordem 3 e resolução de decomposição 2^{-4}.

A Figura 4.6 compara ambas as técnicas de suavização discutidas nesta seção - wavelets e média móvel - à série original de preços. O gráfico de cima apresenta a série completa, de 1986 a 2008, e o segundo, de 2003 a 2008. Observando-os, ficam claras as vantagens das wavelets sobre as médias móveis, discutidas anteriormente.

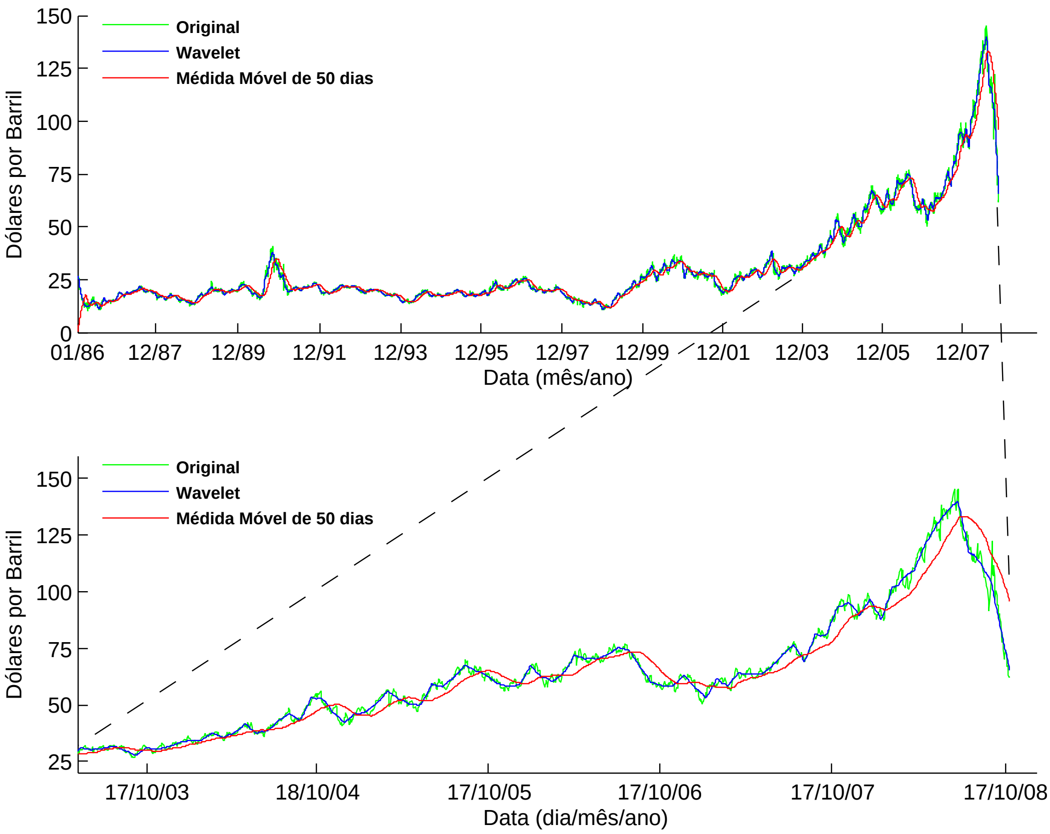


Figura 4.6: Série de preços original do WTI e sua suavização por wavelet e por média móvel de 50 dias, em dois momentos: $[1986, 2008]$ e $[2003, 2008]$.

Conforme dissemos, o objetivo da suavização é remover a volatilidade de curtíssimo prazo presente na série de preços, a fim de ressaltar a dinâmica de seus movimentos mais duradouros. A Figura 4.7 mostra a autocorrelação dos retornos da série suavizada por wavelets, apresentada na Figura 4.6. Observando-a, fica evidente o benefício da suavização.

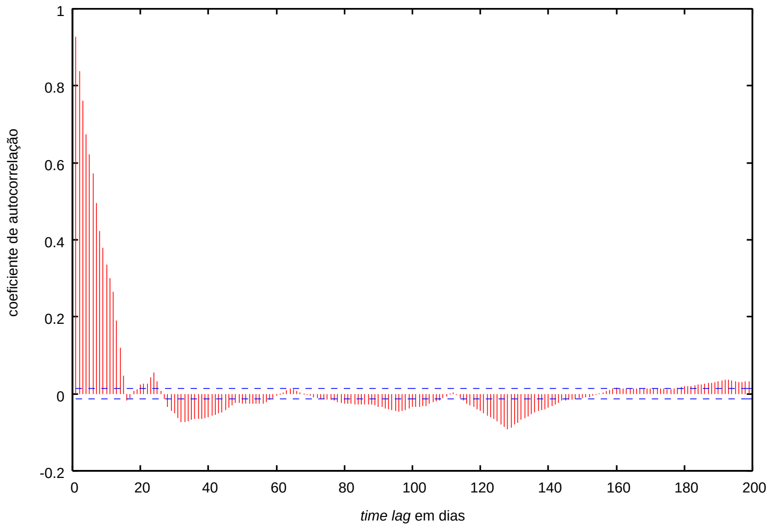


Figura 4.7: Autocorrelação dos retornos diários da série de preços do WTI, suavizada por wavelets.

4.3 Codificação dos Dados

Utilizamos, nesta dissertação, um modelo de Markov oculto de tempo discreto para prever a distribuição do retorno acumulado pelo preço do petróleo em uma janela de tempo futura. Como foi visto no capítulo 2, o HMM trabalha com um conjunto discreto e finito de símbolos, de tamanho M, os quais correspondem às saídas observáveis do processo que se deseja estudar. Evidentemente, existe um *trade-off* entre o número de símbolos e a complexidade do modelo. Quanto maior for M, maior é a capacidade de representação do modelo, no entanto, como mais parâmetros precisam ser estimados, maior é o número de máximos locais da função de verossimilhança. Assim, em processos cujo conjunto de saídas observáveis é muito grande, pode ser interessante processá-las de alguma maneira, de modo a reduzir o conjunto de símbolos utilizado pelo HMM.

Estamos estudando o processo de formação do preço do petróleo, e suas saídas nada mais são que os preços de fechamento diários do mercado. Uma escolha natural de símbolos seria representar cada preço por um símbolo idêntico ao seu valor. Por exemplo, 41.83 dólares seria representado pelo símbolo 41.83. No entanto, este conjunto, apesar de discreto[4], apresenta alguns graves problemas que o tornam inviável de ser utilizado no HMM. Em primeiro lugar, como não existe um limite superior para o preço, não há como tornar este conjunto finito. Em segundo lugar, ele contém elementos demais, o que torna o HMM desnecessariamente complexo, e inviável de ser trabalhado sob a forma discreta. Como solução para este último problema, poderíamos arredondar o preço, a fim de trabalharmos apenas com sua parte inteira, mas ainda assim, restaría-nos um conjunto extenso de símbolos[5]. Por estas razões, fica clara a necessidade de elaborarmos algum código que mapeie as saídas (os preços) à um conjunto reduzido de símbolos.

4.3.1 Definindo o Conjunto de Símbolos do HMM

A série de preços, além de fornecer, evidentemente, o preço, contém outra importante informação: o retorno diário. Definimos este retorno como

$$r_t = \left[\left(\frac{p_t - p_{t-1}}{p_{t-1}}\right) - 1\right] \cdot 100\% \tag{4.3}$$

onde r_t é o retorno percentual no dia t, e p_t o preço neste mesmo dia. O valor do retorno diário dos principais ativos financeiros está, na maioria das vezes, contido em um intervalo bem definido, $[-10\%, +10\%]$ por exemplo. Apesar de poder variar no intervalo $[-100\%, +\infty\%)$, é muito raro encontrarmos casos onde a variação do ativo é de -90% ou $+84\%$ em um único dia. Isso é verdade especialmente no caso de uma *commodity* muito negociada, de bastante liquidez, como o WTI. Durante o período que vai de 1986 até o final de 2008, sua maior variação percentual absoluta de preço, num único dia, foi de 33.40%, em 17 de janeiro de 1991. Portanto, é razoável assumirmos que seus retornos diários estão contidos em um conjunto finito. Aliado a isto, temos que retornos próximos (1.11% e 1.25%, por exemplo) são, para efeitos práticos, muito semelhantes, e assim podem ser considerados iguais. Esta qualidade é de grande interesse, pois nos permite representar um conjunto de saídas

[4] a variação mínima de preço permitida durante sua negociação é de 0.01 dólares.

[5] 80 símbolos só para representar os preços entre 20 e 100 dólares!

com um mesmo símbolo. Assim, dadas estas características do retorno, decidimos utilizá-lo como a saída observável de nosso processo, ao invés do preço.

Feita a escolha das saídas, o método de obtenção do conjunto de símbolos do modelo HMM é descrito a seguir, e está ilustrado na Figura 4.8. O primeiro passo é selecionar um retorno L, que define o valor limite do retorno com o qual vamos trabalhar. Em seguida, dividimos os retornos em faixas de valores de tamanho V, e associamos a cada uma um símbolo diferente, de modo que todos os retornos que pertençam a uma mesma faixa serão mapeado para um mesmo símbolo. A primeira faixa contém retornos negativos maiores que $-L\%$; a última, os retornos positivos maiores que $+L\%$. As intermediárias dividem o intervalo $[-L\%, +L\%]$ em $2L/V$ subintervalos. O último passo consiste em determinar o valor de cada símbolo. Uma rápida observação do histograma de cada faixa, nos mostra que seus valores são, de forma aproximada, uniformemente distribuídos (exceto para as faixas limites). Portanto, decidimos que cada símbolo terá o valor da média do intervalo ao qual está associado. No caso das faixas limites $[-100\%, -L\%]$ e $[+L\%, +\infty\%]$, o valor absoluto do símbolo passa a ser $L + \delta$ onde $\delta = V/2$.

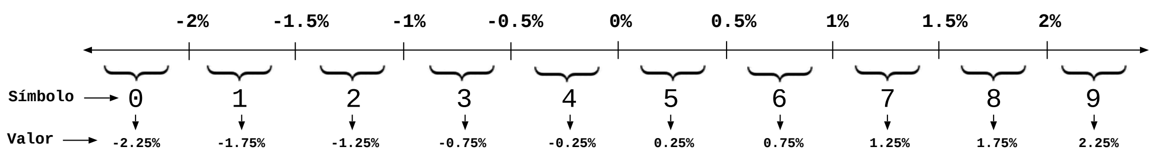


Figura 4.8: Construção do conjunto de símbolos do HMM, considerando $L = 2\%$ e $V = 0.5\%$.

4.3.2 Algoritmo de Codificação

Definido o conjunto de símbolos, a codificação do conjunto de dados é trivial. Basta encontrarmos a faixa a qual cada retorno pertence, e substituí-lo pelo símbolo associado àquela faixa. É importante ressaltar que devemos calcular os retornos a partir da série suavizada de preços, e não a partir da série original.

A fim de clarear o processo descritos acima, ilustramo-no com o exemplo a seguir.

Suponha a seguinte série de preços, já suavizada, $O = (25.56, 26.00, 26.53, 25.85, 25.87, 26.03, 25.65, 25.08, 24.97, 25.18)$. Sua série de retornos, calculada com a equação (4.3), é $O_r = (1.72\%, 2.04\%, -2.56\%, 0.08\%, 0.62\%, -1.46\%, -2.22\%, -0.44\%, 0.84\%)$. Suponha também que, segundo algum critério, escolhemos $L = 2\%$ e $V = 0.5\%$. Assim, teremos criado o seguinte conjunto intervalos

$$\Big\{[-100\%, -2\%], (-2\%, -1.5\%], (-1.5\%, -1\%], (-1\%, -0.5\%], (-0.5\%, 0\%), [0\%, 0.5\%), [0.5\%, 1\%), [1\%, 1.5\%), [1.5\%, 2\%), [2\%, \infty\%)\Big\}$$

ao qual associamos os símbolos $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ respectivamente.

Definidos os símbolos, o próximo passo é codificar nosso conjunto de dados O_r. Seu primeiro elemento, 1.72%, pertence a penúltima faixa definida acima. Como a ela está associado o símbolo 8, substituímos 1.72% por 8 no conjunto de dados. Seguindo o mesmo raciocínio, substituímos 2.04%, o segundo elemento, por 9; -2.56% por 0 e assim em diante, como ilustrado na Figura 4.9. Ao final da codificação, teremos o conjunto de dados $\tilde{O}_r = (8, 9, 0, 5, 6, 2, 0, 4, 6)$.

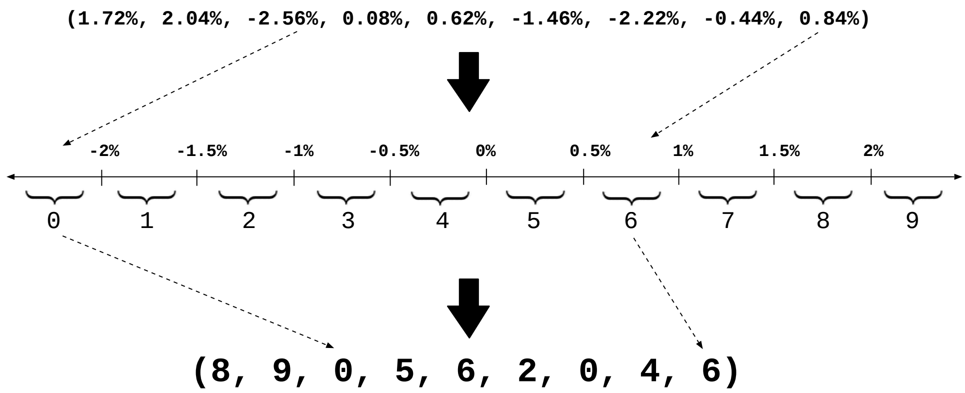


Figura 4.9: Codificação do vetor de retornos no vetor de símbolos do HMM.

4.4 Treinamento do HMM

Uma rápida inspeção visual da série de preços do WTI, ou de qualquer ativo financeiro, mostra que estas não são séries estacionárias. Suas estatísticas, como a média e variância, claramente variam ao longo do tempo. Devido a este dinamismo dos preços, precisamos, de tempos em tempos, reestimar os parâmetros do nosso modelo, a fim de capturar as novas dinâmicas correntes.

A Figura 4.10 ilustra o mecanismo de treinamento adaptativo que utilizamos, baseado em [33, 34]. Os parâmetros do modelo são periodicamente reestimados a cada τ dias, utilizando o algoritmo EM descrito na seção 2.1.4. Em cada treino, apenas as amostras dos últimos T dias são utilizadas no procedimento de estimação.

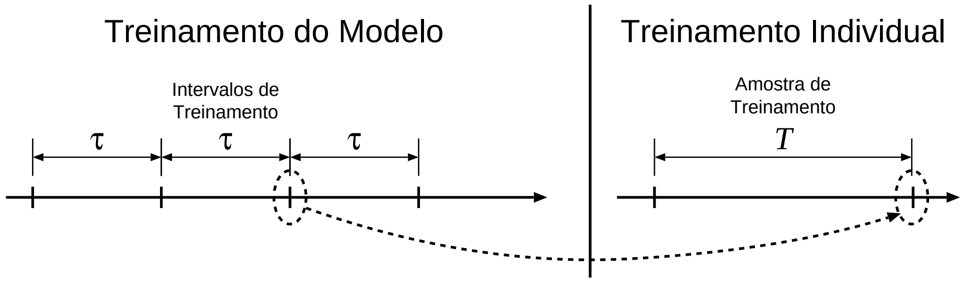


Figura 4.10: Esquema de treinamento adaptativo do HMM.

Os valores de τ e T têm impacto na qualidade da previsão. Se T for muito pequeno, podemos estar omitindo do processo de treinamento informações relevantes da dinâmica corrente dos preços. O contrário, T muito grande, levará o HMM a considerar dinâmicas que já não mais se encontram presentes.

4.5 Método de Previsão

Nesta etapa, temos como objetivo calcular a distribuição do retorno acumulado pelo preço do barril de petróleo, ao final de uma janela futura de F dias, dados o modelo construído e o histórico recente dos preços. Este último é necessário pois, como o processo que estamos modelando não é estacionário, precisamos inferir o estado (da cadeia de Markov oculta) no qual ele se encontra no momento da previsão. Assim, iniciamos a previsão calculando a distribuição dos estados da cadeia no tempo t (o instante atual). Em seguida, utilizamos o paradigma de recompensas de cadeias de Markov [35] (que será descrito mais adiante) para obter a distribuição do retorno em $t + F$, onde F é o tamanho da janela de previsão.

A Figura 4.11 ilustra o esquema geral da metodologia de previsão. O intervalo de treinamento é dividido em intervalos de previsão de tamanho ψ, conforme ilustrado na camada de *previsão da medida*. Cada previsão individual é condicionada nas amostras dos últimos H dias mais recentes. A seguir, descrevemos detalhadamente cada passo de nossa metodologia de previsão, que foi baseada em [33, 34].

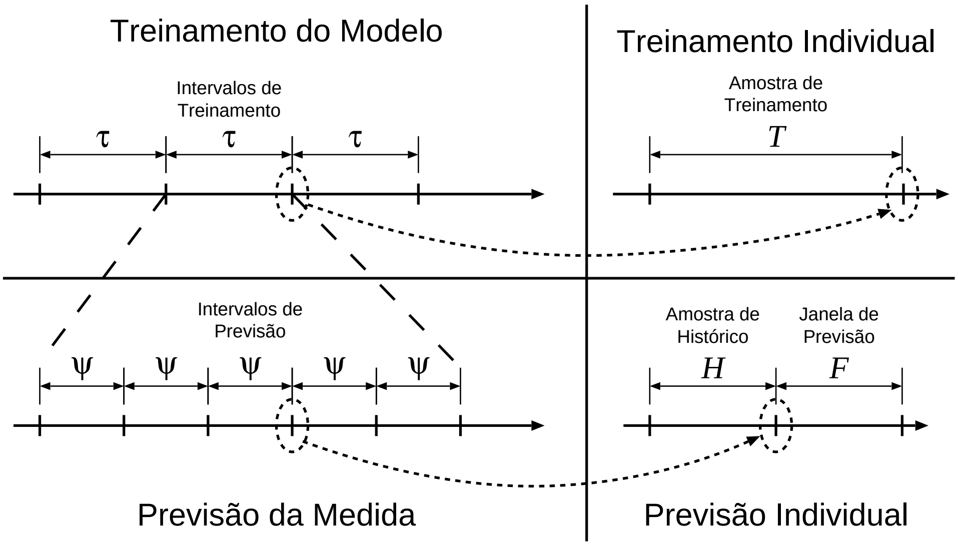


Figura 4.11: Esquema de previsão adaptativo do HMM.

Seja $\pi_{t,h}(i)$ a probabilidade de a cadeia se encontrar no estado oculto S_i, no tempo t, dadas as h observações mais recentes. Ela pode ser facilmente calculada através da variável *forward* $\alpha_t(i)$, definida em (2.8), porém medida apenas no conjunto de observações $[O_{t-h}, O_{t-1}]$:

$$\begin{aligned}
\pi_{t,h}(i) &= \frac{P(q_t = S_i, O_{t-h:t-1})}{P(O_{t-h:t-1})} \\
&= \frac{\sum_{\forall q_{t-1}} P(q_{t-1}, O_{t-h:t-1}, q_t = S_i,)}{P(O_{t-h:t-1})} \\
&= \frac{\sum_{\forall q_{t-1}} P(q_{t-1}, O_{t-h:t-1}) P(q_t = S_i | q_{t-1}, O_{t-h:t-1})}{P(O_{t-h:t-1})} \\
&= \frac{\sum_{\forall q_{t-1}} P(q_{t-1}, O_{t-h:t-1}) P(q_t = S_i | q_{t-1})}{P(O_{t-h:t-1})} \\
&= \frac{\sum_{\forall q_{t-1}} \alpha_{t-1}(q_{t-1}) a_{q_{t-1}S_i}}{P(O_{t-h:t-1})} \qquad (4.4)
\end{aligned}$$

onde $O_{t-h:t-1} = O_{t-h}O_{t-h+1}\dots O_{t-1}$ e $a_{q_{t-1}S_i}$ é a probabilidade da transição de estados (q_{t-1}, S_i). Se denotarmos por α_t o vetor cujo i-ésimo elemento é $\alpha_t(i)$, então (4.4) pode ser reescrito como:

$$\pi_{t,h} = \frac{\alpha_{t-1}A}{P(O_{t-h:t-1})} \tag{4.5}$$

onde A é a matriz de probabilidades de transição de estados.

Dado $\pi_{t,h}$, podemos facilmente determinar a probabilidade, segundo nosso modelo, de o preço do petróleo ter uma alta de $k\%$ no dia seguinte. Para isto, basta calcularmos a probabilidade de a cadeia emitir o símbolo v_k após uma transição, o que pode ser feito da seguinte forma:

$$P[r_{t+1} = k] = \sum_{i=1}^{N}\sum_{j=1}^{N} \pi(i) a_{ij} b_j(v_k) \tag{4.6}$$

onde r_{t+1} é o retorno em $t+1$; v_k é o símbolo que corresponde a uma alta de $k\%$; N é o número de estados da cadeia; $b_j(v_k)$ é a probabilidade de o símbolo v_k ser emitido pelo estado S_j; a_{ij} a probabilidade da transição (S_i, S_j); e $\pi(i)$ a probabilidade do estado S_i em t. Se utilizarmos a equação (4.6) para calcular a probabilidade de todos os símbolos do HMM, ao final, teremos obtido a distribuição do retorno do preço em $t + 1$.

Seguindo o mesmo raciocínio, podemos calcular a distribuição do retorno em $t + \Delta t$ dias. Basta calcularmos todos os retornos possíveis de serem obtidos ao final destes Δt dias e suas respectivas probabilidades. Por exemplo, se $\Delta t = 2$ e o HMM tiver 2 símbolos, teremos 4 possíveis retornos, cujas probabilidades são dadas por

$$P[R_{t+2} = k] = P[r_{t+1} = l, r_{t+2} = m] = P[V_{t+1} = v_l] \cdot P[V_{t+2} = v_m] \tag{4.7}$$

onde R_t é o retorno total acumulado em t dias (cujo valor é $k = l \cdot m$); r_t é o retorno obtido somente no dia t; e V_t é o símbolo emitido em t. A probabilidade da cadeia emitir o símbolo v_j, após t transições, é

$$P[V_t = v_j] = \pi A^t B[j] \qquad 0 \leq j \leq (M-1) \tag{4.8}$$

onde M é o tamanho do conjunto de símbolos, A é a matriz de probabilidades de transição da cadeia de Markov oculta, B é sua matriz de emissão e $B[j]$ é a j-ésima coluna desta matriz.

O problema desta abordagem é que o número de operações feitas para calcular todos os retornos possíveis em $t + \Delta t$ cresce combinatorialmente com Δt. Como vimos, para $\Delta t = 2$ e $M = 2$ temos que realizar 2^2 operações. Agora, suponha que $M = 10$. Se $\Delta t = 3$, precisamos de 10^3 operações. Se quisermos fazer a previsão para daqui há um mês (20 dias úteis), necessitamos de 10^{20} operações. Dado um computador cujo processador trabalhe a 2 GHz, somente para calcular cada retorno possível de ser produzido pela cadeia, após 20 transições, levaríamos $5 \cdot 10^{10}$ segundos $= 1.3 \cdot 10^7$ dias!

Uma das vantagens em se utilizar cadeias de Markov, é que existem algoritmos eficientes que possibilitam o cálculo de diversas medidas de interesse, dada uma cadeia. Um, em específico, nos permite obter a distribuição do acúmulo de determinado valor, após um número finito de transições. Para isto, utiliza o paradigma de recompensas de impulso de uma cadeia de Markov [35].

Apesar de podermos definir recompensas de impulso tanto para cadeias de Markov, quanto para cadeias de Markov ocultas, o algoritmo mencionado acima só pode ser utilizado nas primeiras. Não existe, até o momento, uma generalização que permita que ele seja aplicado também em cadeias de Markov ocultas. No entanto, felizmente, existe uma forma simples de transformarmos uma cadeia de Markov oculta em uma cadeia de Markov equivalente com recompensas. Sendo assim, para computar as previsões, decidimos transformar nossa cadeia de Markov oculta em uma cadeia de Markov com recompensas equivalente, para então, utilizarmos o algoritmo mencionado.

A seguir definimos o que são recompensas de impulso de uma cadeia de Markov. Em seguida, mostramos como é possível expandir uma cadeia de Markov oculta, transformando-a em uma cadeia de Markov com recompensas. Por fim, apresentamos o algoritmo que nos possibilitará obter a distribuição do retorno acumulado pelo preço do WTI ao final de uma janela de F dias.

4.5.1 Recompensas de Impulso em Cadeias de Markov

Os modelos markovianos com recompensas de impulso têm associados, a cada transição (S_i, S_j) da cadeia, um valor $\rho_k \in \{\rho_1, \rho_2, \ldots, \rho_K\}$, denominado recompensa de impulso, que é ganho pelo processo, como recompensa, cada vez que esta transição é feita. A Figura 4.12 apresenta um modelo de Markov com recompensas de impulso.

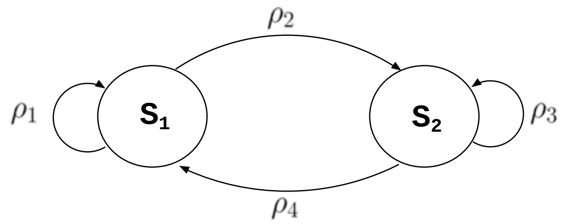


Figura 4.12: Cadeia de Markov discreta de 2 estados com suas recompensas de impulso ρ_1, ρ_2, ρ_3 e ρ_4.

A variável ACI(t) mede a recompensa total acumulada pelo modelo no intervalo $(0, t)$, e é, normalmente, definida como

$$ACI(t) = \sum_{n=0}^{N(t)} \rho_{\hat{c}(\sigma_n)} \tag{4.9}$$

onde $N(t)$ é o número de transições que ocorreram no intervalo $(0, t)$, σ_n é a n-ésima transição e $\hat{c}_{(i,j)}$ é o índice da recompensa associada à transição (S_i, S_j).

Tradicionalmente, a recompensa de impulso é utilizada para contagem. A fim de melhor ilustrar estes conceitos, considere o exemplo a seguir. Suponha que a cadeia da Figura 4.12 represente um jogo de cara ou coroa, no qual apenas uma moeda, não viciada, é utilizada. A cada jogada, o jogador ganha 1 real se o resultado for cara, e perde 1 real se o resultado for coroa. Formalmente, temos que $S_1 =$ Cara, $S_2 =$ Coroa, $\rho_1 = \rho_4 = 1$ e $\rho_2 = \rho_3 = -1$. A variável ACI(t) fornece a quantidade de dinheiro ganha pelo jogador após t jogadas. A Figura 4.13 exemplifica uma possível evolução de ACI(t) no tempo.

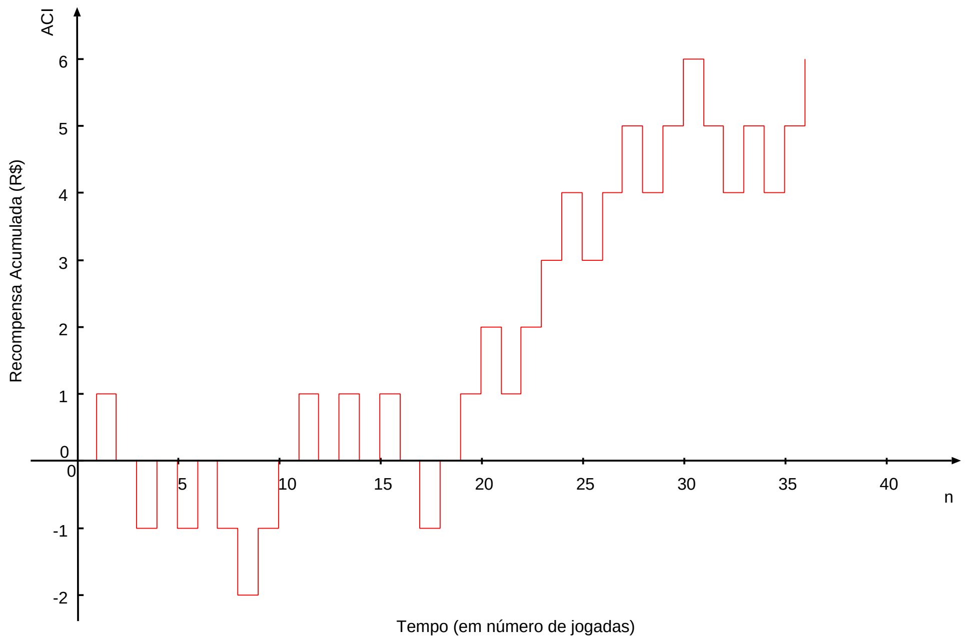


Figura 4.13: Evolução da recompensa acumulada no tempo.

4.5.2 Método de Expansão de um Modelo de Markov Oculto

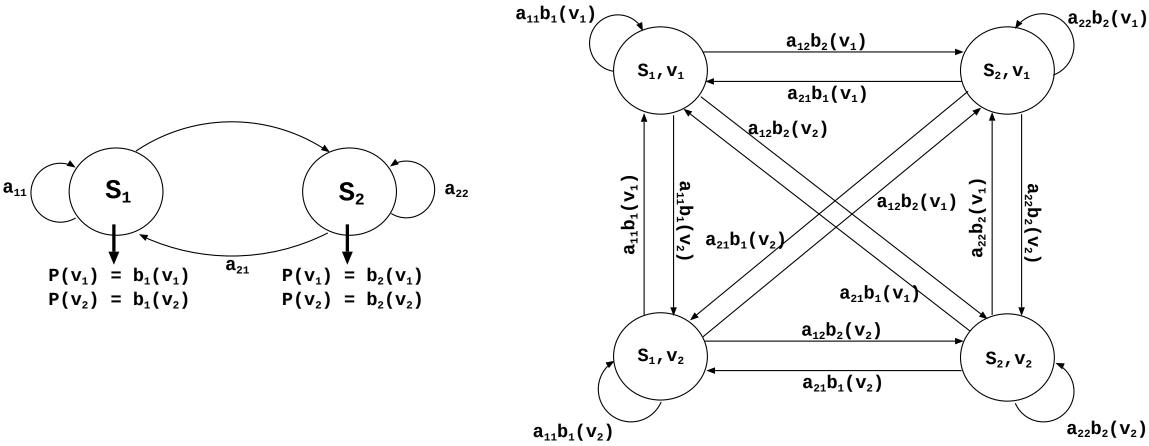
Seja $\mathcal{X}$ a cadeia de Markov oculta de N estados; $A = \{a_{ij}\}, 1 \leq i, j \leq N$ sua matriz de probabilidades de transição de estados; $\mathcal{Y}$ o processo de observações de tamanho M; $V = \{v_1, v_2, \ldots, v_M\}$ o conjunto de símbolos; e $B = \{b_j(k)\}, 1 \leq j \leq N$ e $1 \leq k \leq M$ sua matriz de emissão de símbolos. É possível construir uma cadeia de Markov discreta $\mathcal{W}$, equivalente a $\mathcal{X}$, com $N \times M$ estados, cuja distribuição de estados inicial $\hat{\pi}$ e matriz de probabilidades de transição $\hat{A}$ são dadas por

$$\hat{\pi}(i,k) = \frac{\pi(i)}{M} \tag{4.10a}$$

$$\hat{a}_{i,(j,k)} = a_{ij} b_j(k) \tag{4.10b}$$

e onde a cada transição $(i,(j,k))$ é associada uma recompensa $\rho_{\hat{c}_{(i,(j,k))}} = v_k$, com $v_k \in V$. A prova da equivalência entre as cadeias $\mathcal{X}$ e $\mathcal{W}$, em relação a probabilidade de $\mathcal{X}$ emitir a seqüência de símbolos $V_1, V_2, \ldots, V_Z$ de tamanho Z e $\mathcal{W}$ receber a seqüência de recompensas $V_1, V_2, \ldots, V_Z$ de tamanho Z pode ser vista em [36]. A principal característica do processo de Markov $\mathcal{W}$ é que seu espaço de estados é uma expansão do espaço de estados do processo oculto.

A Figura 4.14 ilustra o processo de expansão. Em 4.14(a) é apresentada a cadeia de Markov oculta a qual desejamos expandir e, em 4.14(b), a cadeia de Markov resultante da expansão, na qual foram utilizadas as equações (4.10a) e (4.10b).



(a) Modelo de markov oculto **(b) Modelo de markov estendido**

Figura 4.14: Processo de expansão de um HMM. Em (a) está ilustrada a cadeia de Markov oculta e, em (b), sua cadeia de Markov equivalente.

4.5.3 Cálculo da Distribuição da Recompensa Acumulada

A seguir, apresentamos um algoritmo desenvolvido por [37] que permite calcular a distribuição da recompensa acumulada por uma cadeia de Markov, em um instante de tempo t, de forma eficiente. Porém, antes de descrevê-lo, precisamos introduzir três novas variáveis.

Introdução de Limites à Recompensa Acumulada

Em muitos dos processos que estudamos, há algum limite nos valores quantitativos observados. Por exemplo, no jogo de cara ou coroa apresentado na seção 4.5.1, um jogador pode ficar com, no mínimo, $-D$ reais, onde D é a quantidade de dinheiro apostada por ele[6] e com no máximo B reais, onde B é a quantidade de

[6]Estamos considerando que não existe a possibilidade de ser contrair algum empréstimo, ou algo do gênero.

dinheiro da banca de apostas. Portanto, é necessário introduzirmos estes limites à variável ACI(t), para que esta fique condizente com a realidade.

Definimos então, as variáveis $l.i.$ e $l.s.$ que representam, respectivamente, os valores dos limites inferior e superior de ACI(t). Para ilustrar, suponha que, em nosso jogo de cara ou coroa, um jogador tenha apostado 1 real, e que a banca de apostas estabeleça em 4 reais o prêmio máximo que pode ser pago a um jogador. A Figura 4.15 mostra em 4.15(a) o caminho amostral de ACI(t) sem os limite, e em 4.15(b) o novo caminho, após a introdução dos limites $l.i. = -1$ e $l.s. = 4$.

Introdução de Níveis de Recompensa

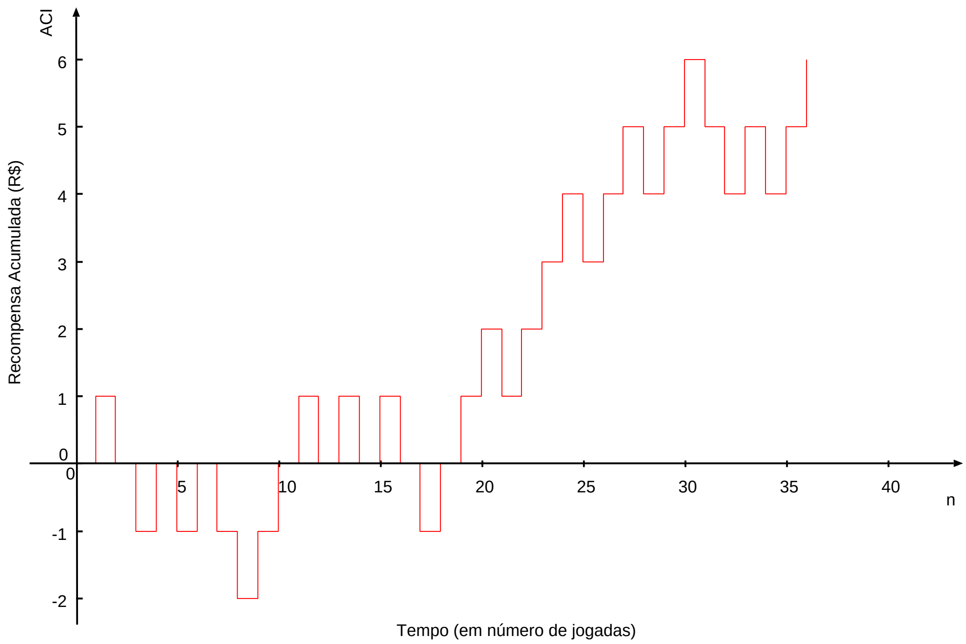
Além de limites inferiores e superiores, para evitarmos de trabalhar com todos os valores de recompensa no intervalo $[l.i., l.s.]$, definimos também a variável *gran*[7] que estabelece o valor mínimo de recompensa que pode ser adicionado ou subtraído à recompensa acumulada, cada vez que o processo faz uma transição. Desta forma, estamos definindo níveis de recompensa r', que serão múltiplos da granularidade escolhida. No nosso jogo de cara e coroa, o natural seria escolhermos $gran = 1$, uma vez que o valor mínimo de recompensa que um jogador pode ganhar ou perder é 1.

Escolhidas as variáveis $l.i.$, $l.s.$, e *gran*, estamos prontos para calcular a distribuição da recompensa acumulada.

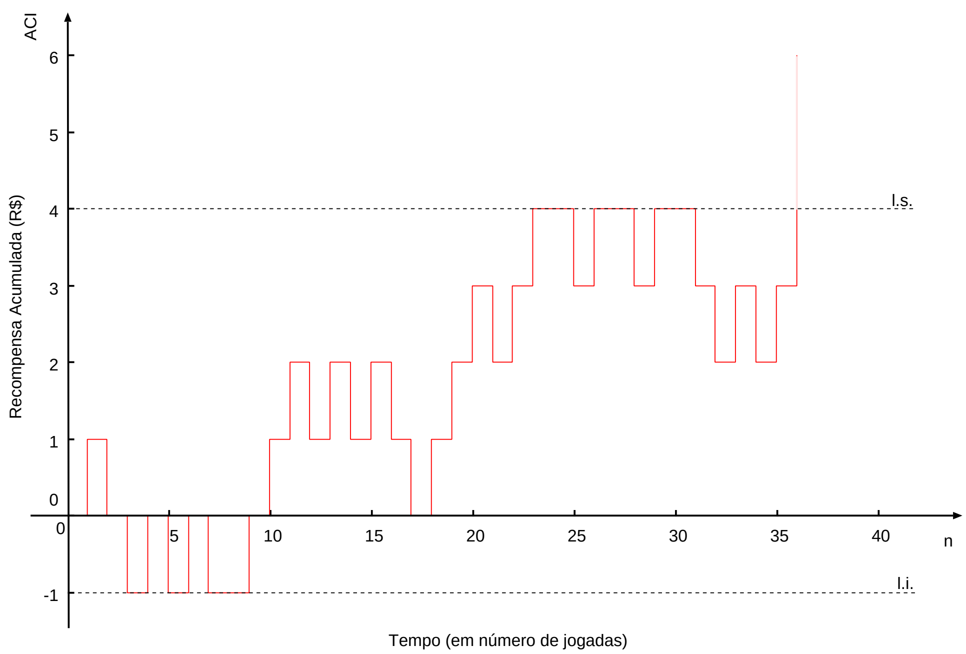
Recursão para o cálculo da Distribuição da Recompensa Acumulada

Seja $\hat{\Gamma}_s[n, \hat{r}]$ a probabilidade de a recompensa acumulada ser igual a $\hat{r}$, dado a ocorrência de n transições, e que o estado visitado após a última transição seja o

[7] de granularidade.



(a)



(b)

Figura 4.15: Evolução da recompensa acumulada no tempo: (a) sem quaisquer limites; (b) com os limites $l.i. = -1$ e $l.s. = 4$.

estado s. Podemos calcular $\hat{\Gamma}_s[n,\hat{r}]$ através da seguinte recursão:

para $n \geq 1$:

$$\hat{\Gamma}_s[n,\hat{r}] = \begin{cases} \displaystyle\sum_{\substack{s' \in \mathcal{S} \\ \rho(s',s) \leq 0}} \sum_{r' \leq \text{l.i.} - \rho(s',s)} p_{s's} \hat{\Gamma}_{s'}[n-1, r'] & \text{, se } \hat{r} = \text{l.i.} \\ \displaystyle\sum_{s' \in \mathcal{S}} p_{s's} \hat{\Gamma}_{s'}[n-1, \hat{r} - \rho(s',s)] & \text{, se l.i.} < \hat{r} < \text{l.s.} \\ \displaystyle\sum_{\substack{s' \in \mathcal{S} \\ \rho(s',s) \geq 0}} \sum_{r' \geq \text{l.s.} - \rho(s',s)} p_{s's} \hat{\Gamma}_{s'}[n-1, r'] & \text{, se } \hat{r} = \text{l.s.} \end{cases} \tag{4.11}$$

onde s' é o estado no qual a cadeia se encontra antes da transição ao estado S, r' é a recompensa acumulada imediatamente antes da última transição, $\rho(s',s)$ é a recompensa de impulso associada a transição (s',s) e $p_{s's}$ é a probabilidade da transição (s',s).

As condições iniciais são:

$$\hat{\Gamma}_s[0,r] = \begin{cases} \pi(s) & \text{, se } r = 0 \\ 0 & \text{, se } r \neq 0 \end{cases} \tag{4.12}$$

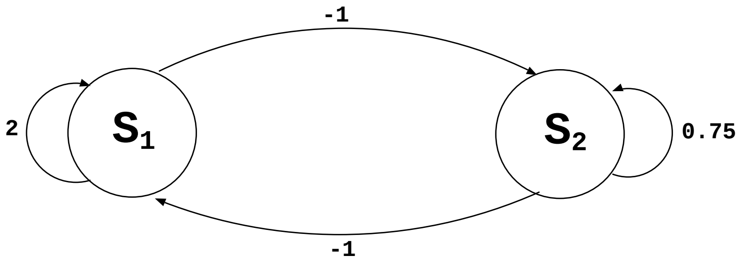
onde $\pi(s)$ é a probabilidade de a cadeia se encontrar, inicialmente, no estado s.

Podemos fazer as seguintes considerações com relação a recursão (4.11), quando a recompensa $\hat{r}$ atinge os valores do limite inferior e do limite superior:

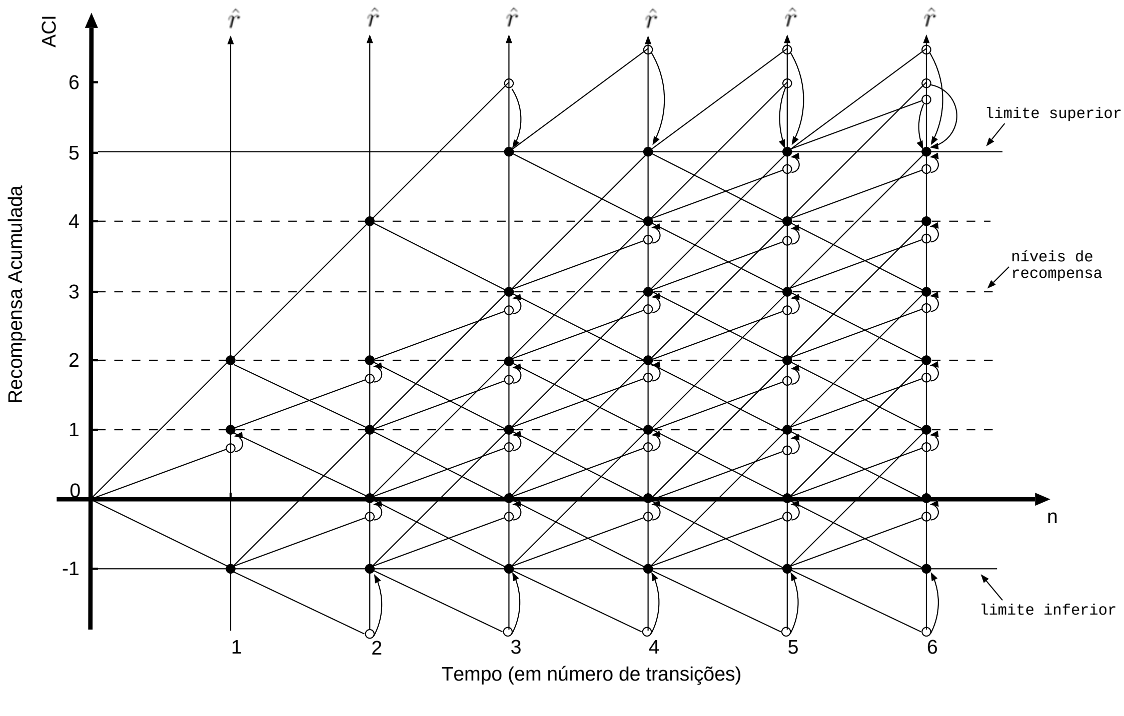
1. Para que a recompensa $\hat{r}$ atinja o valor $l.i.$ na n-ésima transição, o valor de r' até a transição $n-1$ deve assumir valores maiores que $l.i.$ e a recompensa recebida pela transição (s',s) deve ser, obrigatoriamente, menor que zero e ter valor absoluto maior que $r' - l.i.$. Isto explica a restrição $\rho(s',s) \leq 0$ do primeiro somatório, e a restrição $r' \leq \text{l.i.} - \rho(s',s)$ do segundo.

2. Da mesma maneira, para que a recompensa $\hat{r}$ atinja o valor $l.s.$ na n-ésima transição, o valor de r' até a transição $n-1$ deve assumir valores menores que $l.s.$ e a recompensa recebida pela transição (s',s) deve ser, obrigatoriamente, maior que zero e ter valor absoluto maior que $l.s. - r'$. Isto explica a restrição $\rho(s',s) \geq 0$ do primeiro somatório e a restrição $r' \leq \text{l.s.} - \rho(s',s)$ do segundo.

A recursão (4.11) é mais facilmente compreendida se a olharmos do início para o final. A Figura 4.16(b) ilustra o processo de cálculo da distribuição da recompensa acumulada pela cadeia apresentada na Figura 4.16(a). Os valores escolhidos para *l.i.*, *l.s.* e *gran* são, respectivamente, $-1, 5$ e 1.



(a)



(b)

Figura 4.16: Desenvolvimento da recursão (4.11): (a) Cadeia de Markov e suas recompensas; (b) caminho da recursão visto do início ao final.

Observando a Figura 4.16(b), fica evidente a razão pela qual são utilizados níveis de recompensa e sua importância. Note que se não os adotássemos, a quantidade de valores possíveis para a recompensa acumulada, após as seis transições, seria enorme, uma vez que seu crescimento é combinatório. Portanto, a idéia por trás deste método é limitar o número de valores da recompensa acumulada dentro do intervalo $[l.i., l.s.]$, de modo que seja viável calcular sua distribuição de probabilidades.

O cálculo é feito da seguinte forma. Na primeira transição, a cadeia pode ganhar qualquer uma das três recompensas existentes. Se o estado inicial é S_1, há duas possibilidades de recompensa: 2 ou -1. Se o estado inicial for S_2, 0.75 ou -1. Portanto, há três valores possíveis para a recompensa acumulada após a primeira transição: -1, 0.75 ou 2. Suas probabilidades são, respectivamente, $a_{12} + a_{21}; a_{22}$; e a_{11}, onde a_{ij} é a probabilidade da transição (S_i, S_j) na cadeia. No entanto, como 0.75 não corresponde a um dos níveis de recompensa existentes, teremos que aproximá-lo por um destes níveis. O nível mais próximo de 0.75 é 1. Logo, a nova distribuição de probabilidades da recompensa acumulada é

$$\text{para } n = 1: \quad \hat{\Gamma}[1, \hat{r}] = \begin{cases} a_{11} & , \text{ se } \hat{r} = 2 \\ a_{22} & , \text{ se } \hat{r} = 1 \\ a_{12} + a_{21} & , \text{ se } \hat{r} = -1 \\ 0 & , \text{ caso contrário.} \end{cases} \tag{4.13}$$

onde $\hat{\Gamma}[n, \hat{r}]$ é a probabilidade de $\text{ACI}(n) = \hat{r}$.

A distribuição para $n = 2$ é calculada a partir dos valores obtidos para $n = 1$. Se $\text{ACI}(1) = 2$, a cadeia tem que, obrigatoriamente, estar no estado S_1, e, portanto, só pode receber as recompensas 2, com probabilidade a_{11}, ou -1, com probabilidade a_{12}. Se $\text{ACI}(1) = 1$, a cadeia tem, necessariamente, que estar no estado S_2, e as recompensas possíveis de serem recebidas são 1, com probabilidade a_{22}, ou -1, com probabilidade a_{21}. Por último, se $\text{ACI}(1) = -1$, a cadeia pode se encontrar em qualquer um dos dois estados e, assim, receber qualquer uma das três recompensas.

Fazendo os cálculos das probabilidades, chegamos a

para $n = 2$:

$$\hat{\Gamma}[2,\hat{r}] = \begin{cases} \hat{\Gamma}[1,2] \cdot a_{11} & \text{, se } \hat{r} = 4 \\ \hat{\Gamma}[1,1] \cdot a_{22} & \text{, se } \hat{r} = 2 \\ \hat{\Gamma}[1,2] \cdot a_{12} + \hat{\Gamma}[1,-1] \cdot a_{11} & \text{, se } \hat{r} = 1 \\ \hat{\Gamma}[1,1] \cdot a_{21} + \hat{\Gamma}[1,-1] \cdot a_{22} & \text{, se } \hat{r} = 0 \\ \hat{\Gamma}[1,-1] \cdot (a_{12} + a_{21}) & \text{, se } \hat{r} = -1 \\ 0 & \text{, caso contrário.} \end{cases} \tag{4.14}$$

Procedendo de maneira análoga, podemos calcular a distribuição de ACI(6).

Uma visão mais completa, porém simples, do método discutido nesta subseção pode ser visto em [38] e as referências nele contidas. O trabalho que deu origem a recursão (4.11) pode ser visto em [37].

Modificações para Recompensas não Aditivas

O método discutido acima assume que as recompensas são aditivas. No entanto, em nosso caso de estudo, isso não é verdade. Os símbolos emitidos por nosso HMM correspondem a retornos diários do preço do petróleo. Quando expandirmos a cadeia oculta, estes passarão a ser as recompensas da cadeia de Markov equivalente. Se, em três dias consecutivos, o petróleo teve alta de $1.20\%, 2.53\%$ e 0.78%, é errado afirmar que a alta acumulada foi de $1.20\% + 2.53\% + 0.78\% = 4.51\%$. O valor correto é obtido multiplicando-se os fatores de retorno[8], e assim, a alta acumulada é, na realidade, de $1.012 \cdot 1.0253 \cdot 1.0078 = 1.0457$ o que corresponde a uma alta de 4.57%.

Assim, é necessários fazermos alguns pequenos ajustes na recursão (4.11), para que seja possível trabalharmos com esta restrição. Ao invés de somarmos ou sub-

[8]Definimos como *fator de retorno* o número pelo qual deve-se multiplicar o capital inicial para que seja obtido o capital final. Se foi realizado um lucro de 2%, sobre um capital inicial de R$100, o fator de retorno é de 1.02, uma vez que $100 \cdot 1.02 = 102$ reais que equivale ao capital final. Ele pode ser calculado pela equação $f_{r_t} = \frac{p_t - p_{t-1}}{p_{t-1}}$

trairmos as recompensas, teremos agora que multiplicá-las. Além disso, como retornos negativos têm fatores de retorno menores que 1 e os positivos, maiores que 1, devemos também modificar as restrições dos somatórios. A nova forma de (4.11) é

para $n \geq 1$:

$$\hat{\Gamma}_s[n,\hat{r}] = \begin{cases} \displaystyle\sum_{\substack{s' \in \mathcal{S} \\ \rho(s',s) \leq 1}} \sum_{r' \leq \text{l.i.} \cdot \rho(s',s)} p_{s's} \hat{\Gamma}_{s'}[n-1, r'] & \text{, se } \hat{r} = \text{l.i.} \\ \displaystyle\sum_{s' \in \mathcal{S}} p_{s's} \hat{\Gamma}_{s'}[n-1, \hat{r} \cdot \rho(s',s)] & \text{, se l.i.} < \hat{r} < \text{l.s.} \\ \displaystyle\sum_{\substack{s' \in \mathcal{S} \\ \rho(s',s) \geq 1}} \sum_{r' \geq \text{l.s.} \cdot \rho(s',s)} p_{s's} \hat{\Gamma}_{s'}[n-1, r'] & \text{, se } \hat{r} = \text{l.s.} \end{cases} \tag{4.15}$$

cujas condições iniciais são:

$$\hat{\Gamma}_s[0,r] = \begin{cases} \pi(s) & \text{, se } r = 1 \\ 0 & \text{, se } r \neq 1 \end{cases} \tag{4.16}$$

Apesar destas modificações, todos os conceitos discutidos acima permanecem válidos.

Capítulo 5

Resultados

Neste capítulo, apresentamos os resultados obtidos pela metodologia de previsão proposta no capítulo 4. Na seção 5.1, especificamos seus parâmetros, justificando-os, e introduzimos o modelo HMM construído. Na seção seguinte, descrevemos as métricas utilizadas na avaliação de nossa metodologia. Na seção 5.3, apresentamos os resultados. Finalmente, na seção 5.4, comparamos o desempenho de nossa metodologia frente a dois outros modelos de referência: o Replicador e o Passeio Aleatório.

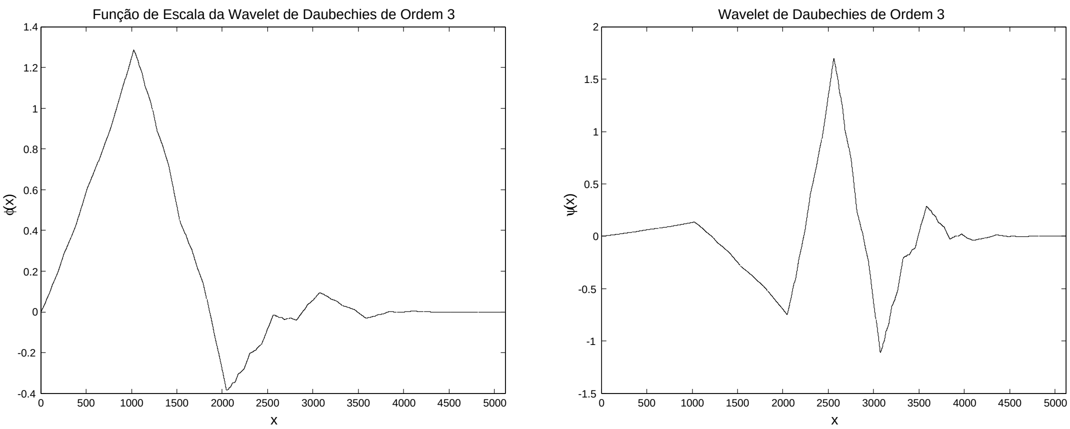
5.1 Especificação de Parâmetros

Esta seção está dividida em duas partes. Na primeira, apresentamos os parâmetros utilizados na suavização por wavelets. Em seguida, descrevemos o processo de construção e escolha dos parâmetros do HMM.

5.1.1 Parâmetros para Remoção de Ruídos por Wavelets

Como foi visto na seção 2.2.5, a remoção de ruídos de um sinal decomposto em wavelets depende de duas escolhas. A primeira consiste em selecionar uma base de wavelet, composta pelas funções $\phi(x)$ e $\psi(x)$, as quais serão utilizadas como funções base na aproximação do sinal. Já a segunda resume-se a definir a política de *thresholding*, ou seja, o método de remoção de coeficientes de wavelet.

Escolhemos trabalhar com a base de Daubechies de ordem 3, ilustrada na Figura 5.1. Decidimos por esta wavelet pois, dentre o conjunto das wavelets disponíveis no Matlab [13], é aquela cujo formato mais se assemelha à dinâmica dos preços do WTI. Reconhecemos que esta é uma escolha subjetiva, baseada apenas na inspeção visual das wavelets, e que seria necessário um estudo mais profundo das características de cada uma, para que pudéssemos avaliar, de forma precisa, qual destas seria mais adequada a nossa aplicação. No entanto, devido às restrições de tempo, tal estudo não pode ser realizado.



(a) (b)

Figura 5.1: Base de Daubechies de ordem 3 utilizada na suavização. (a) Função de escala $\phi(x)$, (b) e sua wavelet correspondente $\psi(x)$.

Conforme mencionamos na seção 4.2, utilizamos a função *wden()* do Matlab para suavizar nossos dados. Esta função tem a vantagem de permitir ao usuário especificar tanto a função de *thresholding* quanto o valor do *threshold* com as quais deseja trabalhar. Particularmente, escolhemos a função de *soft thresholding*, para evitarmos que possíveis descontinuidades abruptas pudessem surgir na série suavizada[1]. A regra de escolha do *threshold* pela qual optamos foi a *heursure*, que é uma variante

[1]Para relembrar a diferença entre as funções de *soft* e *hard thresholding*, sugerimos rever a Figura 2.27.

da regra SURE descrita na seção 2.2.5. Conduzimos um breve estudo para avaliar as diferentes regras disponíveis em *wden()* quanto ao resultado da suavização, porém, observamos que a série suavizada era praticamente idêntica, independente da regra utilizada. Mais detalhes sobre a função *wden()* podem ser encontrados em [13].

5.1.2 Especificação do HMM

A seguir, descrevemos detalhadamente as características do modelo HMM que construímos. Começamos explicando o método que empregamos para determinar a estrutura de sua cadeia de Markov. Em seguida, especificamos seus elementos (o número de estados, número de símbolos, janela de treinamento, etc.) e, por fim, descrevemos como foram obtidos os valores iniciais dos parâmetros do modelo.

Estrutura da Cadeia de Markov

Além do número de estados N, um modelo de Markov é caracterizado pela estrutura de sua cadeia. Até aqui, em todos os nossos exemplos, utilizamos cadeias completas, ou seja, cadeias nas quais cada estado pode ser alcançado partindo-se de qualquer outro estado, com uma transição apenas. Matematicamente, isto significa que $\forall i, j \; a_{ij} > 0$. No entanto, uma cadeia de Markov não precisa, necessariamente, ter esta estrutura. A Figura 5.2 ilustra alguns exemplos de cadeias com estruturas diferentes, comuns de serem encontradas na bibliografia. A escolha da estrutura a ser utilizada no modelo depende das características do fenômeno o qual se está estudado. Infelizmente, não existe um algoritmo que, dado um conjunto de características, construa uma cadeia ótima. Esta é uma tarefa que exige criatividade e bastante conhecimento sobre o processo que se deseja modelar. Assim, a fim de definirmos a estrutura de nosso HMM, realizamos um estudo para identificar as principais características presentes em nosso conjunto de dados.

A suavização da série de preços do WTI torna evidente o fato de que estes se movem, quase sempre, em direções bem definidas (para cima - preços em alta - ou para

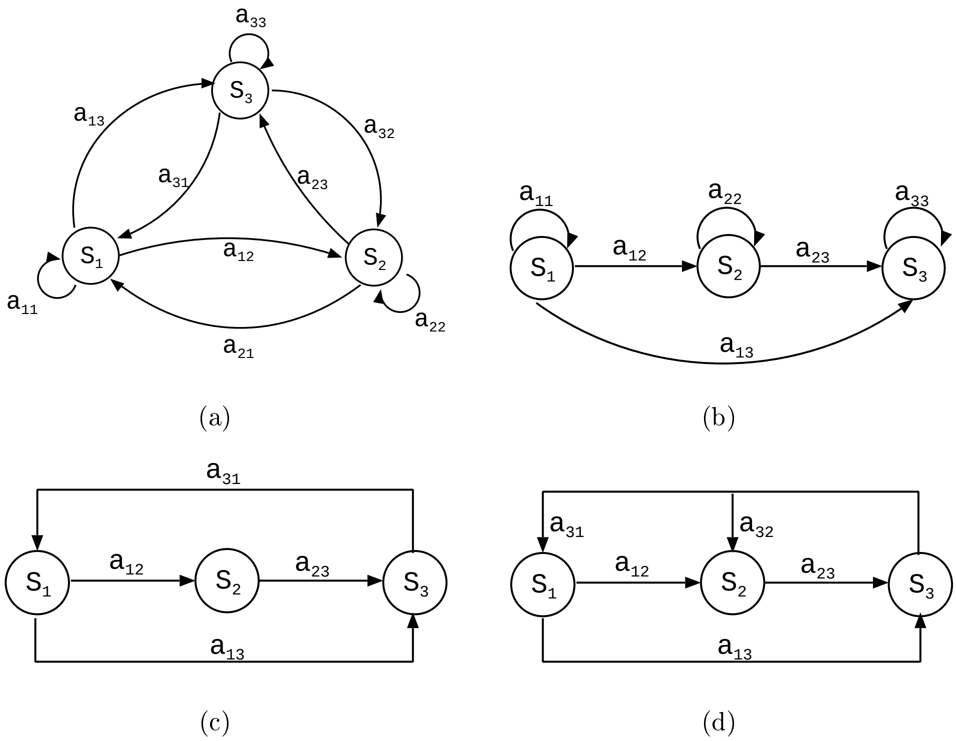


Figura 5.2: Diferentes tipos de cadeias de Markov com 3 estados: (a) cadeia completa, (b) cadeia *left-right*, (c) cadeia coxian, (d) cadeia coxian geral.

baixo - preços em queda), as quais são denominadas de tendências. A Figura 5.3 ilustra este fato. A primeira característica pela qual procuramos é se existe alguma periodicidade nestas tendências, como uma rápida inspeção visual da figura sugere. No entanto, a Figura 4.7 (que mostra a autocorrelação dos retornos diários calculados a partir da série suavizada de preços), sugere que não existe nenhuma periodicidade ou ciclo, de modo que, aparentemente, o momento em que determinada tendência ocorre não contém nenhuma informação sobre quando ela se repetirá novamente. Assim, passamos nosso foco de estudo para o tempo de duração das tendências, com o objetivo de tentarmos construir uma cadeia de Markov que reproduza-o adequadamente, para que assim, uma vez identificada uma tendência, possamos, a partir do modelo, calcular, por exemplo, a probabilidade de ela continuar por mais D dias.

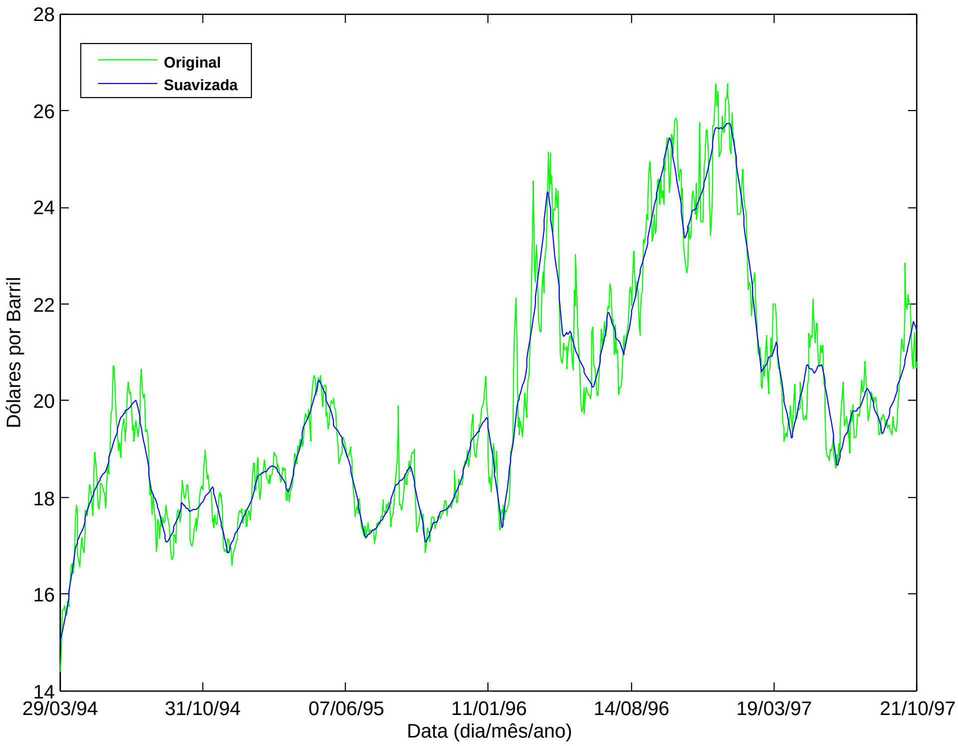


Figura 5.3: Comportamento do preço do WTI entre março de 1993 e outubro de 1997.

A Figura 5.4 mostra a freqüência relativa do tempo de duração (em dias) das tendências de alta e de baixa. Não incluímos ali os tempos de duração de 1 e 2 dias, pois consideramos que estas tendências de curtíssima duração são meramente ruídos não removidos pelo método de suavização em tendências mais longas. Observando ambos os gráficos, podemos notar que os movimentos de alta dos preços tendem a ser mais duradouros que os movimentos de baixa[2], o que indica que é possível que ambos tenham características distintas. Por este motivo, resolvemos modelá-los separadamente.

[2]Este fato é, frequentemente, observado em ativos financeiros de alta liquidez como ações e câmbio, e é de conhecimento geral.

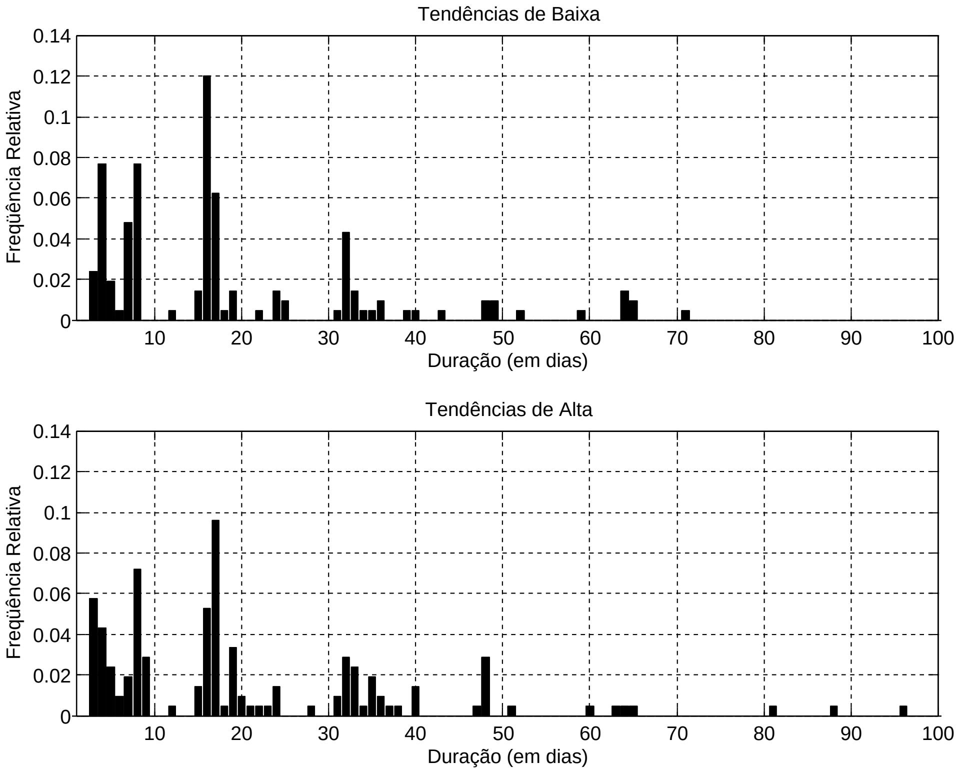


Figura 5.4: Freqüência relativa do tempo de duração, em dias, das tendências de alta e baixa do preço do WTI, no período de janeiro de 1986 a outubro de 2008.

Para melhor caracterizar a duração de uma tendência, decidimos buscar uma distribuição de probabilidades para representá-la. Utilizando o Matlab, testamos, inicialmente, as seguintes distribuições[3]: Exponencial, Gamma, Weibull, LogNormal e Pareto Generalizada. A qualidade do ajuste de cada uma foi verificada através de duas análises: uma visual e outra quantitativa. A análise visual consistiu em comparar os gráficos da função de distribuição acumulada complementar (CCDF, do inglês *complementary cumulative distribution function*) de cada distribuição testada ao gráfico da CCDF empírica. A análise quantitativa foi feita pela comparação do erro médio quadrático (MSE, do inglês *mean squared error*) calculado entre a distribuição empírica e as estimadas. Não conseguimos, no entanto, com nenhuma delas, obter uma descrição satisfatória. Passamos a investigar, então, uma outra classe de distribuições, conhecidas como distribuições *phase-type* (PH).

[3]A parametrização foi feita utilizando um estimador de máxima verossimilhança

Uma distribuição PH é definida como a distribuição do tempo decorrido até que um estado absorvente[4], S_{abs}, de uma cadeia de Markov, seja alcançado. Isto torna-a de grande interesse para nós, pois, por ser definida através de uma cadeia de Markov, uma vez encontrada uma parametrização adequada, encontra-se, consequentemente, a estrutura da cadeia que a gerou (que é justamente o que buscamos). Uma discussão mais completa sobre as distribuições PH pode ser encontrada no apêndice A.1.

Testamos as seguintes distribuições PH: Completa, Hiperexponencial, Soma de Exponenciais, Coxian e Coxian Geral. A parametrização de cada uma foi feita com *software* **EMpht** [39], que realiza uma estimação de parâmetros iterativa, utilizando um algoritmo EM. Calculando as mesmas métricas de avaliação de qualidade do ajuste descritas acima (comparação da CCDF e do MSE), chegamos a conclusão que, em geral, as distribuições PH (exceto a soma de exponenciais) apresentaram melhor ajuste aos dados empíricos se comparadas às distribuições testadas inicialmente, e que, dentre elas, a que apresentou resultados superiores foi a coxian geral. Isto foi verificado tanto para as tendências de alta quanto às de baixa.

Desta forma, decidimos que a cadeia de Markov de nosso HMM deveria ser composta de duas cadeias coxian geral: uma responsável por reproduzir as tendências de alta, e outra, por reproduzir as tendências de baixa. Quanto ao número de estados de cada cadeia, deixaremos para definí-lo mais adiante.

Resta-nos ainda resolver um problema: como unir ambas as partes, a fim de construirmos uma cadeia única. Lembre-se que a cadeia de Markov que gera uma distribuição PH contém um estado absorvente, e que, no caso de nossa aplicação, este estado indica fim da tendência. Assim, se conectarmos todas as transições que chegam a este estado aos estados da outra parte, pelos quais o processo pode se iniciar, estaremos essencialmente conectando o final de uma tendência (de alta, por exemplo) ao início da tendência seguinte (de baixa). A Figura 5.5 ilustra este processo. As setas em vermelho representam as transições que foram removidas da cadeia, e as em azul, as que foram adicionadas. Assim, procedendo da maneira

[4]Um estado é classificado como absorvente se não há como deixá-lo. Matematicamente, isto significa que $a_{ii} = 1$. O estado S_3 na Figura 5.2(b) é um exemplo de estado absorvente.

descrita, construímos a estrutura da cadeia de Markov do nosso modelo HMM.

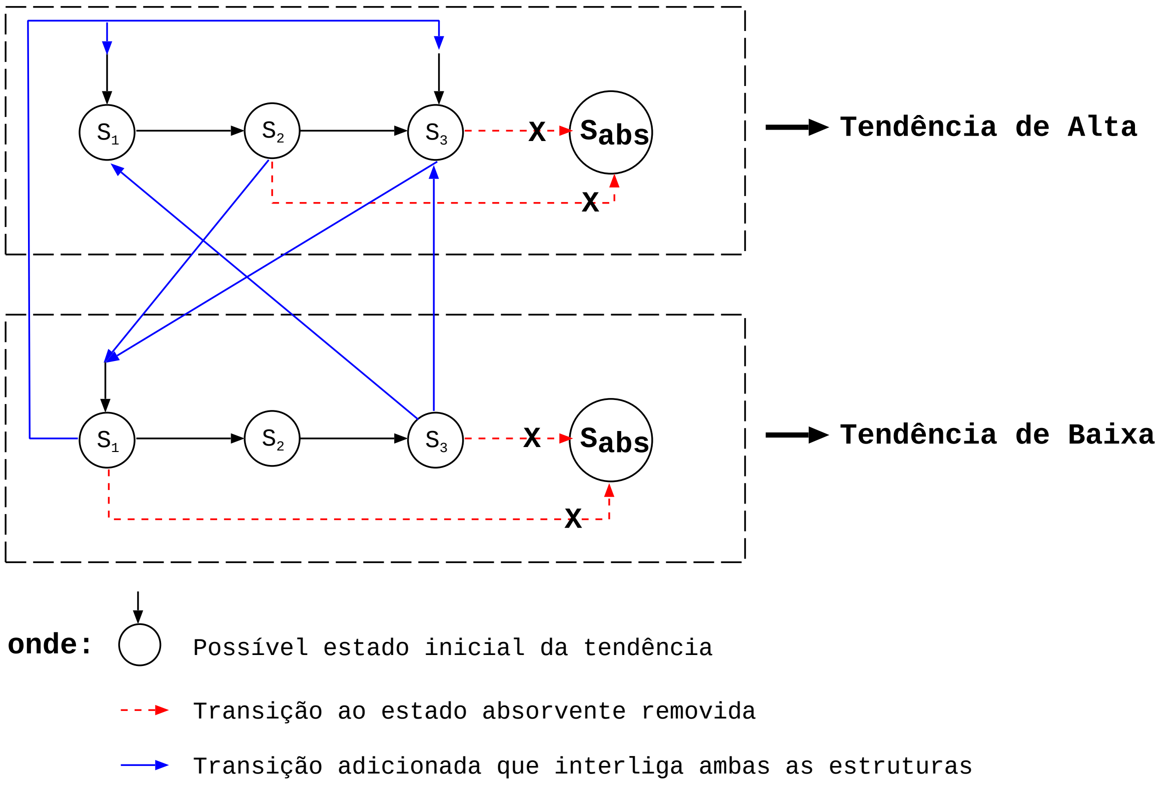


Figura 5.5: Exemplo do processo de união das estruturas construídas para as tendências de alta e baixa dos preços.

Antes de finalizarmos esta etapa, precisamos fazer uma ressalva. O programa **EMpht**, citado acima, faz o ajuste de uma distribuição *phase-type* contínua aos dados e, portanto, nos fornece como saída uma cadeia de Markov de tempo contínuo. No entanto, estamos trabalhando com um modelo de Markov de tempo discreto, de modo que foi necessário encontrar alguma maneira de se transformar esta cadeia contínua em uma discreta equivalente. Felizmente, existe uma técnica, conhecida como método de Uniformização [40], que faz exatamente isso. Por questões de coesão, discutimos este método e a maneira como o utilizamos, no apêndice A.2.

Elementos do HMM

Definida a estrutura geral da cadeia, precisamos, ainda, especificar o modelo HMM. Seus elementos são: o número de estados N, o número de símbolos M, o valor de cada faixa V às quais os símbolos estão associados (reveja a Figura 4.8), o tamanho do intervalo τ e a quantidade de amostras T que serão utilizadas no treinamento (ilustrados na Figura 4.10) e, finalmente, o intervalo de previsão Ψ, o tamanho do histórico recente H e o tamanho da janela de previsão F (apresentados na Figura 4.11). A seguir, especificamo-os individualmente.

F : Devido ao alto grau de incerteza presente na série de preços, não é possível utilizar uma janela de previsão muito extensa, de 100 dias por exemplo. No entanto, a alta volatilidade da série também não permite que esta janela seja curta demais, da ordem de 1, 3 ou 5 dias. Assim, realizamos um estudo empírico para definir seu tamanho adequado, e chegamos a conclusão que este encontra-se, aproximadamente, dentro do intervalo $[20, 30]$ dias. Assim, apresentaremos nossos resultados tanto para $F = 20$ quanto para $F = 30$ dias.

H : Observando a Figura 5.4, podemos notar que são raras as tendências cuja duração é superior a 60 dias. Por esta razão, decidimos utilizar este valor como tamanho do histórico recente utilizado na previsão, pois consideramos que ele é suficiente para identificarmos a tendência corrente e, assim, inferirmos o possível estado no qual a cadeia se encontra no momento da previsão.

Ψ : O intervalo de previsão foi escolhido como tendo o mesmo tamanho da janela de previsão, a fim de simplificar o modelo e seus resultados.

Os valores de N, M, V, τ e T foram obtidos experimentalmente, através de um estudo de análise de sensibilidade, cujos resultados podem ser vistos no apêndice A.3. Os valores selecionados encontram-se na Tabela 5.1. Conforme mencionado acima, fizemos experimentos usando duas janelas de previsão distintas, 20 e 30 dias, portanto, na tabela abaixo, estão especificados os valores dos elementos utilizados em cada caso.

Escolha dos Valores Iniciais dos Parâmetros do HMM

Conforme discutido ao final da seção 2.1.4, os valores iniciais dados aos parâmetros π, A e B têm influência sobre modelo resultante. Isto ocorre pois, apesar de o algoritmo EM (utilizado na estimação destes parâmetros) convergir para um máximo local da função verossimilhança, nada garante que este último é, também, um máximo global. É evidente que, a medida que o modelo se torna mais complexo (o que ocorre quando se aumenta o número de estados, o número de símbolos, ou a estrutura da cadeia), cresce o número de máximos locais da função de verossimilhança. Portanto, é de se esperar que modelos maiores sejam mais sensíveis a essas escolhas iniciais.

Na subseção anterior, descrevemos um método para estimar a estrutura da cadeia de Markov do modelo HMM, que consistia em ajustar uma distribuição PH ao conjunto dos tempos de duração das tendências. Devido à natureza desta distribuição, essencialmente o que estamos fazendo, ao aplicarmos este método, é encontrar os parâmetros (π e A) que melhor ajustam uma cadeia de Markov absorvente, específica, ao conjunto de dados. Assim, decidimos utilizar estes valores encontrados como os valores iniciais dos parâmetros π e A de nosso modelo.

Há apenas alguns detalhes aos quais devemos estar atentos. Lembremo-nos que nossa cadeia é dividida em duas partes, uma que corresponde as tendências de alta

Tabela 5.1: Especificação do modelo HMM.

	Tamanho da Janela de Previsão	
	20 dias	30 dias
N	8	8
M	4	2
V	1.00%	1.00%
T	600	900
τ	600	900

e outra que corresponde as tendências de baixa (como pode ser visto na Figura 5.5), e que o ajuste de cada uma é feito individualmente. Assim, como resultado, temos duas cadeias distintas, cA e cB, cada qual com seu conjunto de parâmetros. Evidentemente, ao uní-las, não podemos simplesmente concatenar os vetores π_{cA} e π_{cB} de modo a forma o vetor π da cadeia conectada. Se assim o fizéssemos, este último não seria um vetor de probabilidades, uma vez que a soma de seus elementos seria igual a 2.[5] É preciso que, após esta concatenação, seja feita a normalização do vetor π.

Um cuidado semelhante deve ser tomado com o parâmetro A. Conforme vimos, uma transição feita de um estado transiente para um estado absorvente deve ser substituída por I novas transições (onde I é o número de estados iniciais da outra cadeia), cada qual ligando o estado transiente em questão aos estados iniciais da outra. Evidentemente, os valores destas novas transições não podem ser iguais ao valor da transição removida (isto violaria as restrições probabilísticas). Logo, o valor de uma nova transição deve ser igual ao valor da transição antiga vezes a probabilidade do estado inicial ao qual ela se liga. Matematicamente, isto significa que $a_{ik} = a_{i,abs} \cdot \pi(k)$. A Figura 5.6 ilustra este processo.

[5] $\sum_{\forall i} \pi_{cA}(i) = 1$ e $\sum_{\forall j} \pi_{cB}(j) = 1$, logo, como $\pi = \pi_{cA} \cup \pi_{cB}$, $\sum_{\forall k} \pi(k) = 2$.

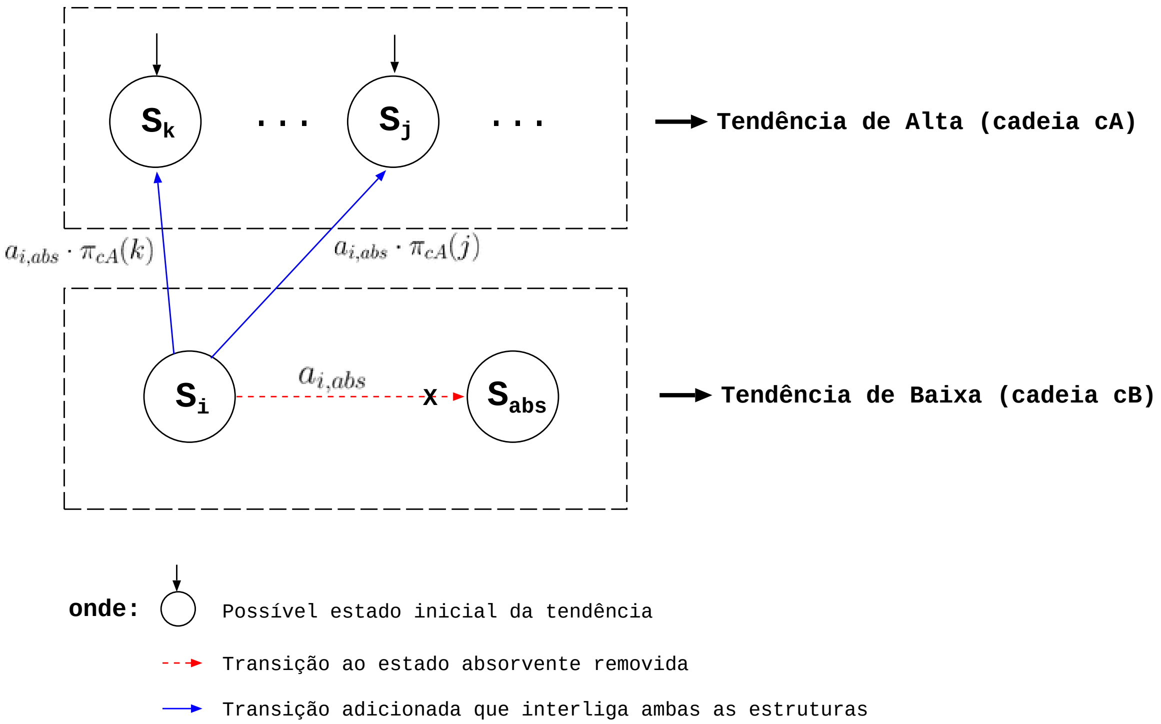


Figura 5.6: Exemplo do processo de união dos parâmetros das estruturas construídas para as tendências de alta e baixa dos preços.

Definidos os valores iniciais de π e A, resta-nos especificar o parâmetro B. A fim de definirmos seus valores iniciais, tentamos primeiro encontrar uma distribuição de probabilidades que reproduzisse adequadamente o conjunto de retornos de nossa série de preços suavizada. Naturalmente, dividimos os retornos em positivos e negativos, e ajustamos uma distribuição para cada conjunto. Foram testadas as seguintes distribuições: Exponencial, Gamma, Weibull, LogNormal e Pareto Generalizada. A avaliação da qualidade do ajuste foi feita utilizando as mesmas métricas citadas anteriormente, comparação do CCDF e do MSE. Para ambos os conjuntos (retornos negativos e positivos) a distribuição que melhor se ajustou aos dados foi a Gamma.

Encontrada uma distribuição para os retornos, podemos então definir a distribuição inicial de símbolos em cada estado e, consequentemente, os valores iniciais do parâmetro B. Suponha que queiramos definir estas probabilidades para um estado que pertença à cadeia que representa as tendências de alta. Naturalmente, os símbolos que representam retornos negativos terão probabilidade zero. Para os demais, sua probabilidade de emissão será igual a probabilidade do intervalo ao qual estão associados, calculada a partir da distribuição Gamma ajustada. A Figura 5.7 ilustra este processo. Procedendo de maneira análoga para cada estado, definimos, assim, os valores iniciais do parâmetro B.

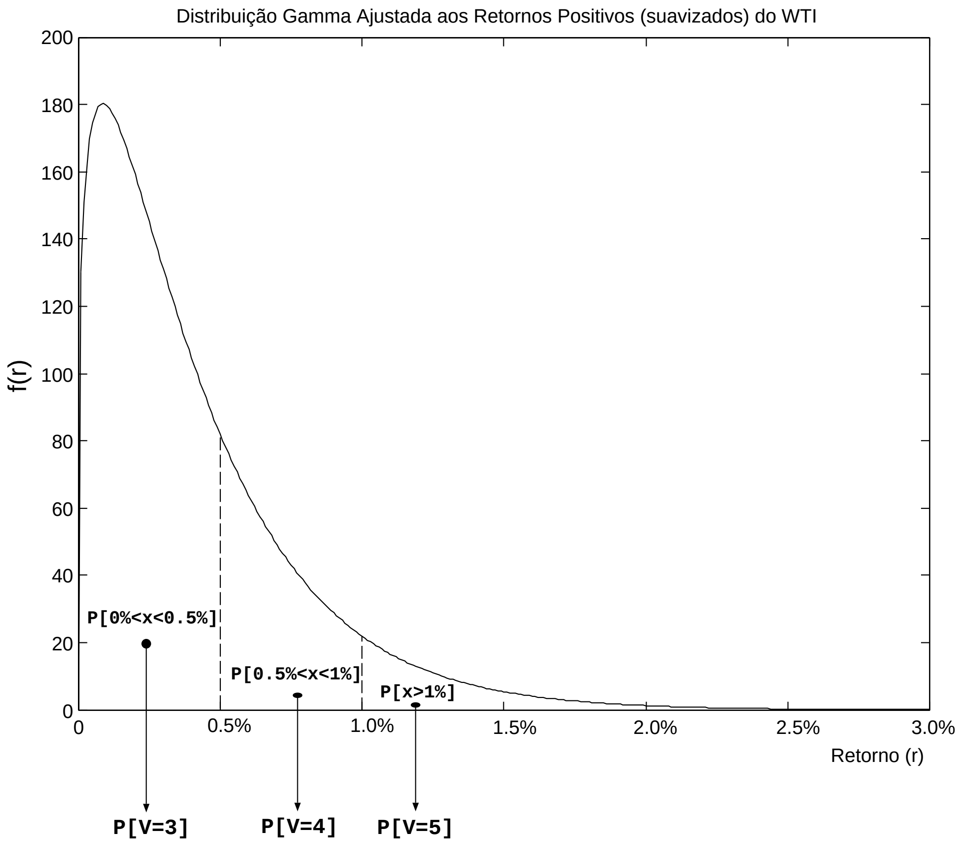


Figura 5.7: Definindo a distribuição dos símbolos do modelo HMM. Neste exemplo, estamos utilizando $M = 6$ símbolos e $V = 0.5\%$. Os três primeiros símbolos (que não estão ilustrados) têm probabilidade zero, pois representam retornos negativos.

5.2 Métricas de Avaliação da Metodologia

A metodologia proposta no capítulo 4 tem como objetivo prever a distribuição do retorno acumulado do preço do petróleo, ao final de uma janela de tempo futura de F dias. É de grande interesse para um investidor ter uma distribuição como esta disponível, pois isto lhe permite elaborar diferentes estratégias de investimento, assim como estimar o risco de suas decisões.

A fim de avaliarmos a metodologia e o modelo criados (quanto à qualidade de suas previsões), decidimos verificar os resultados auferidos por um investidor que utilizasse-os, em conjunto com a seguinte a estratégia de investimento: comprar petróleo no mercado *spot* se a média da distribuição calculada for positiva (o que indica que o modelo calcula que uma alta no preço é mais provável que uma queda) e vender se a média for negativa[6]. Decidimos fazer uso desta estratégia por sua simplicidade, e por representar um referencial para todas as outras, uma vez que, caso o investidor consiga obter lucros extraordinários utilizando-a, então, muito possivelmente, ele também o fará se construir outras estratégias mais complexas, que, por exemplo, façam uso de instrumentos financeiros como derivativos e opções.

Definida a estratégia, utilizamos as seguintes métricas para avaliar nossa metodologia:

- $\mathbf{D_{stat}}$: Indica a taxa de acerto da direção dos preços, ou seja, a fração de vezes que o modelo previu corretamente uma alta (ou queda) nos preços, no período considerando.
- **Retorno**: Indica o retorno obtido pelo investidor, ao utilizar determinada estratégia.

5.3 Resultados

Os ensaios a seguir foram realizados utilizando a série de preços diários, nominais, do WTI obtida na página da EIA [21]. Ela contém 5760 observações, e compreende o período que vai de 02 de janeiro de 1986 a 28 de outubro de 2008. Realizamos

[6]**Importante**: Vender não deve ser interpretado aqui como *short selling*, mas, simplesmente, como uma operação de venda no mercado *spot* à vista, caso o investidor possua a *commodity*, ou apenas não comprar, caso ele não a possua.

a estimação da estrutura do modelo e de seus parâmetros iniciais (cujos processos foram descritos na seção 5.1.2) utilizando as primeiras 4000 amostras da série. Por isso, consideramos que o investidor hipotético, descrito na seção anterior, inicia seus investimentos somente em agosto de 2001, após ter feito estas estimações iniciais. A seguir, apresentamos os resultados obtidos para duas janelas de previsão distintas, de 20 e 30 dias.

5.3.1 Janela de Previsão de 20 dias

Neste primeiro ensaio, consideramos que a janela de previsão do investidor é de 20 dias, ou seja, a cada 20 dias úteis ele reavalia sua posição no mercado. A estratégia de investimento utilizada por ele é aquela descrita na seção 5.2. Os valores dos elementos do HMM utilizados neste ensaio encontram-se especificados na Tabela 5.1.

A Figura 5.8 ilustra a evolução do capital do investidor (amostrado a cada 20 dias) no período de agosto de 2001 a outubro de 2008. Assumimos que seu investimento inicial é de 100 dólares, que, em nenhum momento, há adição de capital externo, e que não há custos de transação. Para efeito de comparação, a figura também mostra a evolução do capital de um investidor que tivesse optado pela estratégia $Buy-and-Hold$, ou seja, que tivesse investido 100 dólares comprando petróleo em agosto de 2001 e não houvesse mais mudado de posição. Ao ilustrarmos esta estratégia, estamos, na realidade, comparando o retorno auferido pela utilização de nosso modelo ao retorno do mercado.

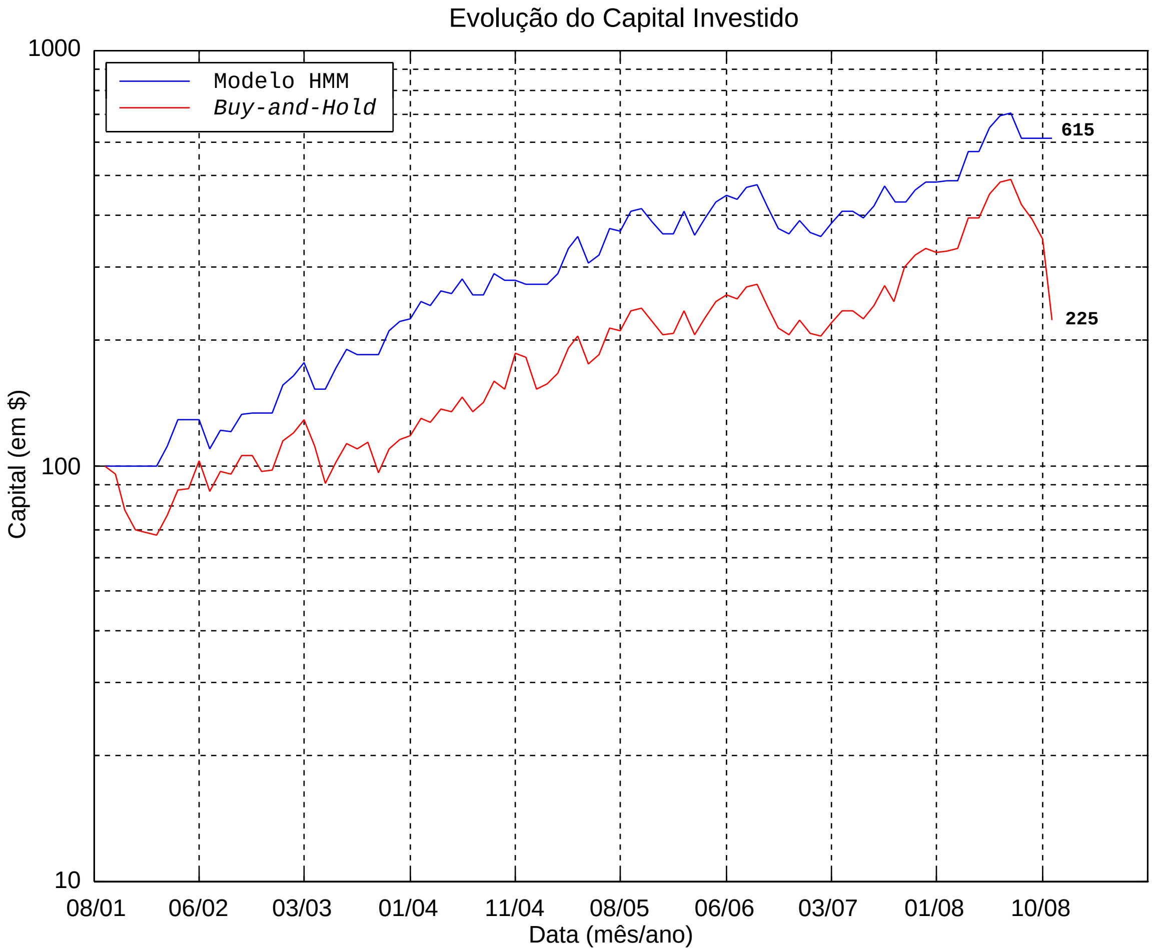


Figura 5.8: Evolução do capital investido com o HMM e com o *Buy-and-Hold*. A taxa de acerto D_{stat} do HMM foi de 63.33%.

No período considerado, o modelo HMM obteve uma taxa de acerto $D_{stat} = 63.33\%$, o que significa que, das 90 previsões feitas, o modelo conseguiu acertar a direção do preço do WTI em 57 delas. O capital final, possuído pelo investidor que utilizou o modelo, é de 615 dólares, o que representa uma rentabilidade de 515% em 7 anos. Neste mesmo período, a estratégia *Buy-and-Hold* apresentou uma rentabilidade de 125%, resultando em um capital final de 225 dólares.

A Tabela 5.2 compara a rentabilidade anual da estratégia que utiliza o HMM, à rentabilidade anual do mercado, ou seja, àquela obtida pelo *Buy-and-Hold*. Dela, podemos, primeiramente, observar que, utilizando o HMM, conseguimos acompanhar as altas do mercado, como ocorreu nos anos de 2002, 2004, 2005 e 2007. Apesar de, nestes anos, a rentabilidade do modelo ser menor que a do mercado, ela é, na maioria das vezes, bem próxima. Mas a grande vantagem do HMM aparece nos mercados em baixa. Observemos os anos de 2003 e 2008. Em ambos, a utilização do HMM trouxe uma substancial vantagem ao investidor, especialmente no ano de 2008, no qual houve uma aumento significativo dos preços do petróleo em seu início, seguido de uma queda brusca, do meio para o final do ano.

Tabela 5.2: Rentabilidades Anuais das Estratégias HMM e *Buy-and-Hold* entre 2002 e 2008.

Ano	Rentabilidade	
	HMM	*Buy-and-Hold*
2002	53.09%	56.40%
2003	46.14%	04.22%
2004	22.41%	33.36%
2005	38.89%	40.81%
2006	−00.33%	−00.33%
2007	33.31%	57.73%
2008	21.82%	−34.59%

A título de curiosidade, decidimos também verificar como o modelo se comportaria caso um investidor o tivesse utilizado desde o início da série (1986). Estamos conscientes que esta verificação ignora o fato de que os valores iniciais dos parâmetros do modelo foram estimados a partir de dados futuros. No entanto, como a reestimação de parâmetros é feita a cada 600 dias úteis, a influência de seus valores iniciais se faz presente apenas nas primeiras previsões. Além disso, como de 2001 até 2008 o preço do petróleo esteve, quase sempre, em alta, achamos válido verificar como o modelo se comportaria em outras situações (como uma longa baixa, por exemplo). Mesmo que a análise deste ensaio não possa ser utilizada para se formular uma conclusão definitiva, ela pode nos fornecer alguns indícios.

Assim, a Figura 5.9 ilustra a evolução do capital do investidor no período de maio de 1988 a outubro de 2008. Novamente assumimos que o investimento inicial feito foi de 100 dólares, que, em nenhum momento, houve adição de capital externo, e que não existem custos de transação.

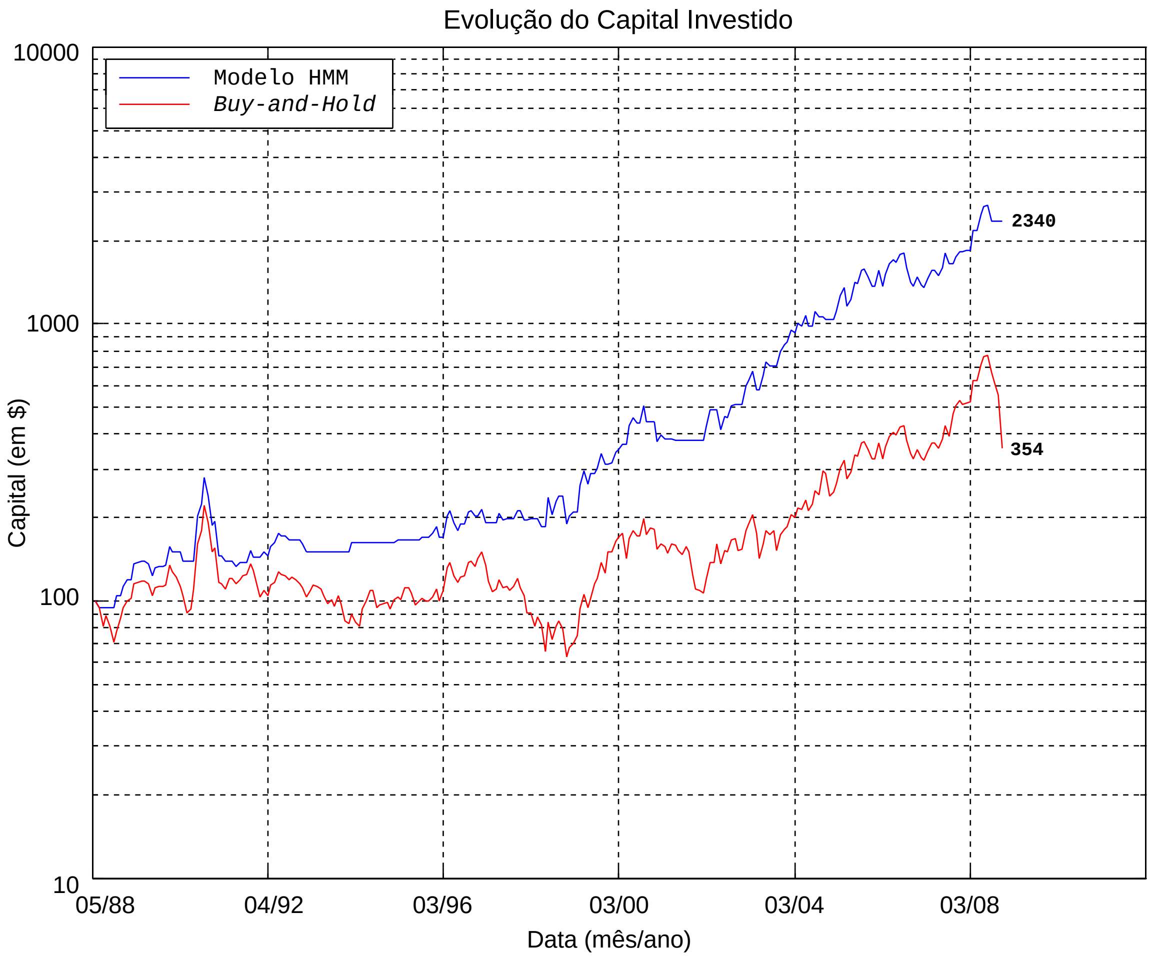


Figura 5.9: Evolução do capital investido com o HMM e com o *Buy-and-Hold*. A taxa de acerto D_{stat} do HMM foi de 60.07%.

No período considerado, a taxa de acerto D_{stat} do modelo HMM foi de 60.07%, o que significa que, das 258 previsões feitas, ele conseguiu acertar a direção do preço do WTI em 155 delas. O capital final, possuído pelo investidor que utilizou o modelo, é de 2340 dólares, o que representa uma rentabilidade de 2240% em aproximadamente 21 anos. Neste mesmo período, a estratégia *Buy-and-Hold* apresentou uma rentabilidade de 254%, resultando em um capital final de 354 dólares, ou seja, a rentabilidade do investidor que utilizou o HMM foi superior a do mercado em quase dez vezes.

A Tabela 5.3 é semelhante a Tabela 5.2, e compara a rentabilidade anual da estratégia que utiliza o HMM, à rentabilidade anual do mercado, no período testado. Em 13 dos 21 anos (ou em 61.90% das vezes), a rentabilidade do modelo foi igual ou superior a rentabilidade do mercado. Novamente, podemos observar como o modelo reduz significativamente as perdas nas épocas de queda do mercado, e acompanha razoavelmente as altas.

Por fim, a Figura 5.10 ilustra a evolução, no tempo, da distância entre o capital acumulado por ambas as estratégias. Esta distância é calculada como a diferença entre os logaritmos do capital acumulado pelo modelo e pelo *Buy-and-Hold*. Observando-a, fica evidente a tendência de crescimento, no longo prazo, desta distância, o que indica que o modelo consegue, constantemente (no longo prazo), auferir resultados superiores aos do mercado.

Tabela 5.3: Rentabilidades Anuais das Estratégias HMM e *Buy-and-Hold* entre 1988 e 2008.

Ano	Rentabilidade	
	HMM	*Buy-and-Hold*
1988	15.61%	−03.42%
1989	25.24%	27.50%
1990	38.99%	30.48%
1991	−28.82%	−32.76%
1992	10.65%	01.91%
1993	−05.59%	−27.16%
1994	08.52%	25.20%
1995	15.25%	09.95%
1996	11.67%	32.55%
1997	−05.70%	−31.84%
1998	04.22%	−31.20%
1999	47.74%	112.28%
2000	21.83%	03.72%
2001	02.95%	−25.33%
2002	53.09%	56.40%
2003	46.14%	04.22%
2004	22.41%	33.36%
2005	38.89%	40.81%
2006	−0.33%	−00.33%
2007	33.31%	57.73%
2008	21.82%	−34.59%

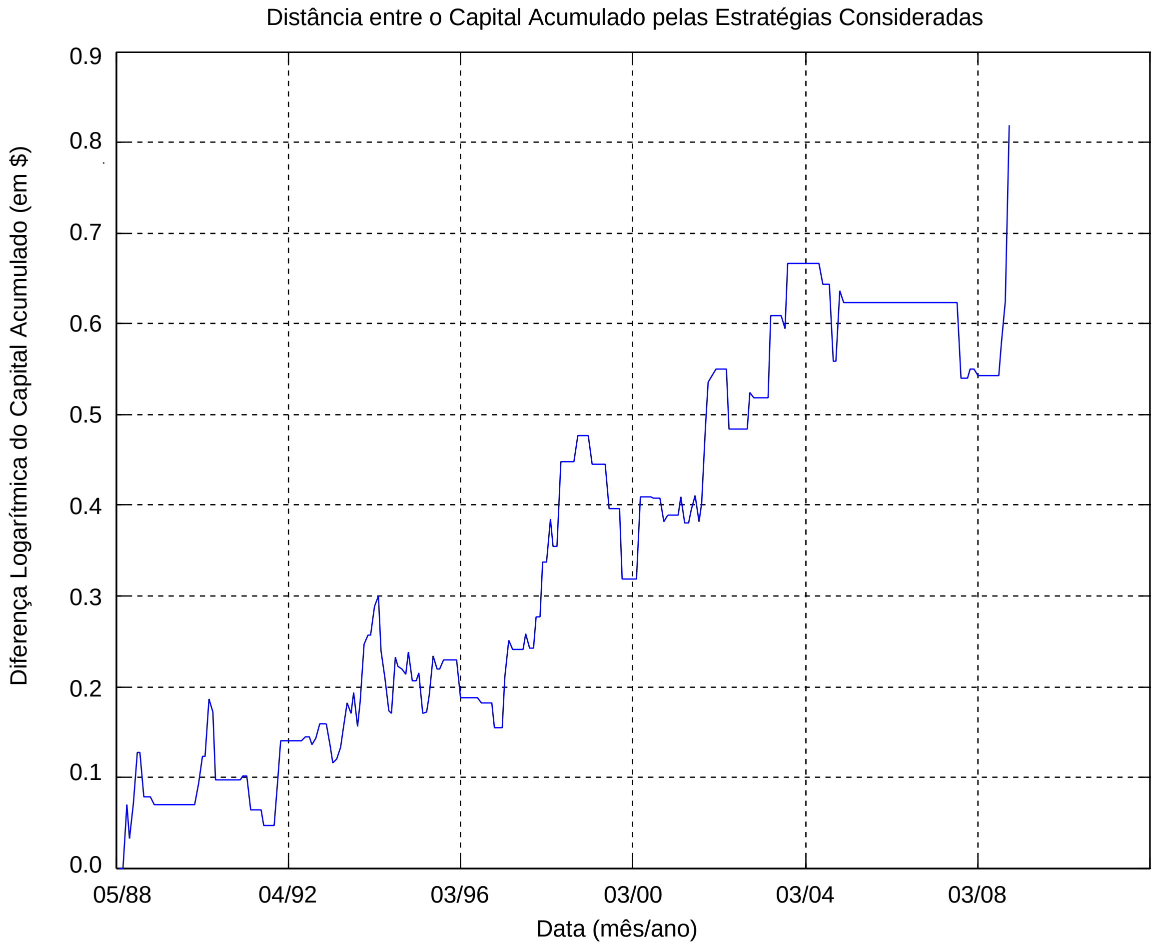


Figura 5.10: Distância entre o capital acumulado pela estratégia que utiliza o modelo HMM e o acumulado pela estratégia *Buy-and-Hold*.

5.3.2 Janela de Previsão de 30 dias

Nesta seção, repetimos os ensaios realizados na seção anterior, só que, desta vez, utilizamos uma janela de previsão de 30 dias. Estes novos experimentos foram realizados com o objetivo de verificarmos a capacidade de previsão da metodologia e do modelo desenvolvidos, ao serem utilizados com um horizonte de previsão um pouco mais longo. Os resultados aqui apresentados seguem a mesma forma daqueles apresentados na seção 5.3.1. Os parâmetros utilizados encontram-se especificados na seção 5.1.2.

A Figura 5.11 ilustra a evolução do capital (amostrado a cada 30 dias) de dois investidores distintos, um que utiliza o HMM e outro que utiliza o *Buy-and-Hold*, no período de outubro de 2001 a outubro de 2008. Novamente assumimos que o investimento inicial feito foi de 100 dólares, que, em nenhum momento, houve adição de capital externo, e que não existem custos de transação.

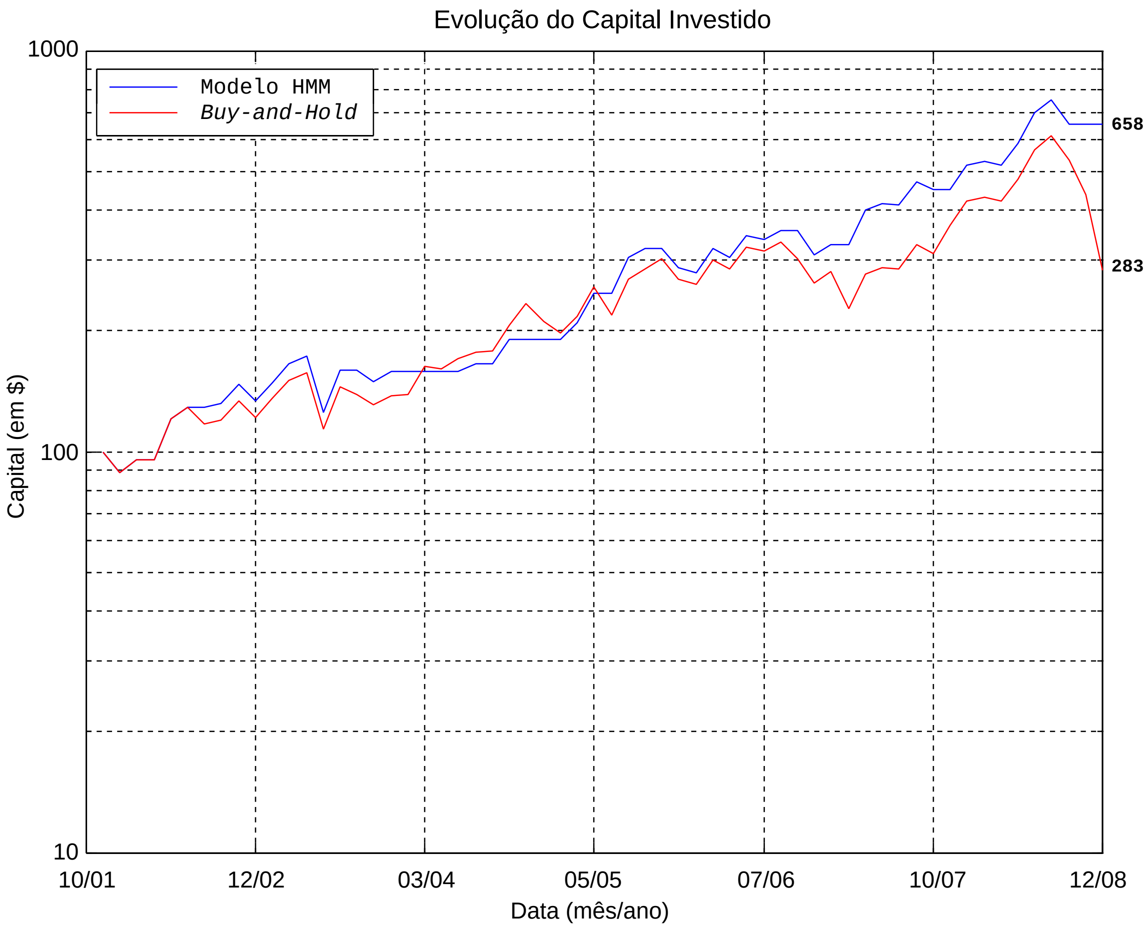


Figura 5.11: Evolução do capital investido com o HMM e com o *Buy-and-Hold*. A taxa de acerto D_{stat} do HMM foi de 66.10%.

No período considerado, o modelo HMM obteve uma taxa de acerto D_{stat} = 66.10%, o que significa que, das 59 previsões feitas, o modelo conseguiu acertar a direção do preço do WTI em 39 delas. O capital final, possuído pelo investidor que utilizou o modelo HMM, é de 658 dólares, o que representa uma rentabilidade de 558% em 7 anos. Neste mesmo período, a estratégia *Buy-and-Hold* apresentou uma rentabilidade de 183%, resultando em um capital final de 283 dólares.

A Tabela 5.4 compara a rentabilidade anual da estratégia que utiliza o HMM, à rentabilidade anual do mercado, ou seja, àquela obtida pelo *Buy-and-Hold*. Novamente, podemos observar que, utilizando o HMM, conseguimos acompanhar bem as altas no mercado (como ocorreu nos anos de 2002, 2004, 2005 e 2007), e nos posicionar favoravelmente durante as baixas, evitando assim perdas substanciais (como em 2006 e 2008).

Tabela 5.4: Rentabilidades Anuais das Estratégias HMM e *Buy-and-Hold* entre 2002 e 2008.

Ano	Rentabilidade	
	HMM	*Buy-and-Hold*
2002	71.69%	56.40%
2003	02.77%	04.22%
2004	20.12%	33.36%
2005	54.16%	40.81%
2006	11.99%	−00.33%
2007	61.61%	57.73%
2008	23.71%	−34.59%

Pelos mesmos motivos descritos na seção 5.3.1, decidimos também verificar o comportamento do modelo HMM nos anos anteriores à 2002. A Figura 5.12 ilustra a evolução do capital do investidor no período de julho de 1989 a outubro de 2008. Novamente assumimos que o investimento inicial feito foi de 100 dólares, que, em nenhum momento, houve adição de capital externo, e que não existem custos de transação.

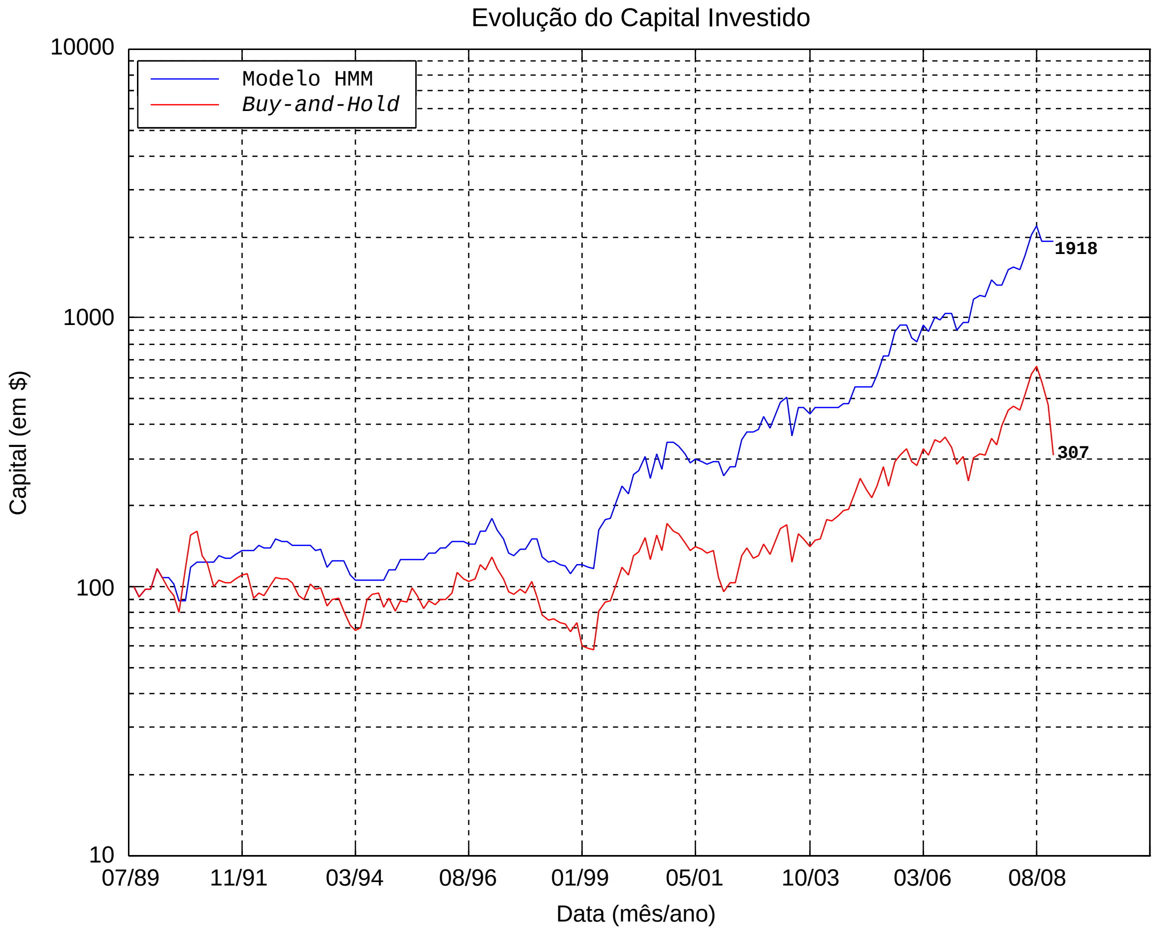


Figura 5.12: Evolução do capital investido com o HMM e com o *Buy-and-Hold*. A taxa de acerto D_{stat} do HMM foi de 61.11%.

No período considerado, a taxa de acerto D_{stat} do modelo HMM foi de 61.11%, o que significa que, das 162 previsões feitas, ele conseguiu acertar a direção do preço do WTI em 99 delas. O capital final, possuído pelo investidor que utilizou o modelo HMM, é de 1918 dólares, o que representa uma rentabilidade de 1818% em aproximadamente 20 anos. Neste mesmo período, a estratégia *Buy-and-Hold* apresentou uma rentabilidade de 207%, resultando em um capital final de 307 dólares.

A Tabela 5.5 é semelhante a Tabela 5.3, e compara a rentabilidade anual da estratégia que utiliza o HMM, à rentabilidade anual do mercado, no período testado. Em 13 dos 19 anos (ou em 68.42% das vezes), a rentabilidade do modelo foi igual ou superior a rentabilidade do mercado. Novamente, podemos observar como o modelo, na maior parte dos anos, reduz significativamente as perdas nas épocas de queda do mercado, e acompanha razoavelmente as altas.

Tabela 5.5: Rentabilidades Anuais das Estratégias HMM e *Buy-and-Hold* entre 1989 e 2008.

Ano	Rentabilidade	
	HMM	*Buy-and-Hold*
1990	15.52%	30.48%
1991	14.25%	−32.76%
1992	01.68%	01.91%
1993	−25.35%	−27.16%
1994	15.37%	25.20%
1995	20.46%	09.95%
1996	19.23%	32.55%
1997	−19.59%	−31.84%
1998	−16.15%	−31.20%
1999	112.28%	112.28%
2000	10.37%	03.72%
2001	−05.92%	−25.33%
2002	71.69%	56.40%
2003	02.77%	04.22%
2004	20.12%	33.36%
2005	54.16%	40.81%
2006	11.99%	−00.33%
2007	61.61%	57.73%
2008	23.71%	−34.59%

Por fim, a Figura 5.13 ilustra evolução, no tempo, da distância entre o capital acumulado por ambas as estratégias. Observando-a, fica evidente a tendência de crescimento, no longo prazo, desta distância, o que indica que o modelo consegue, no longo prazo, auferir resultados superiores aos do mercado.

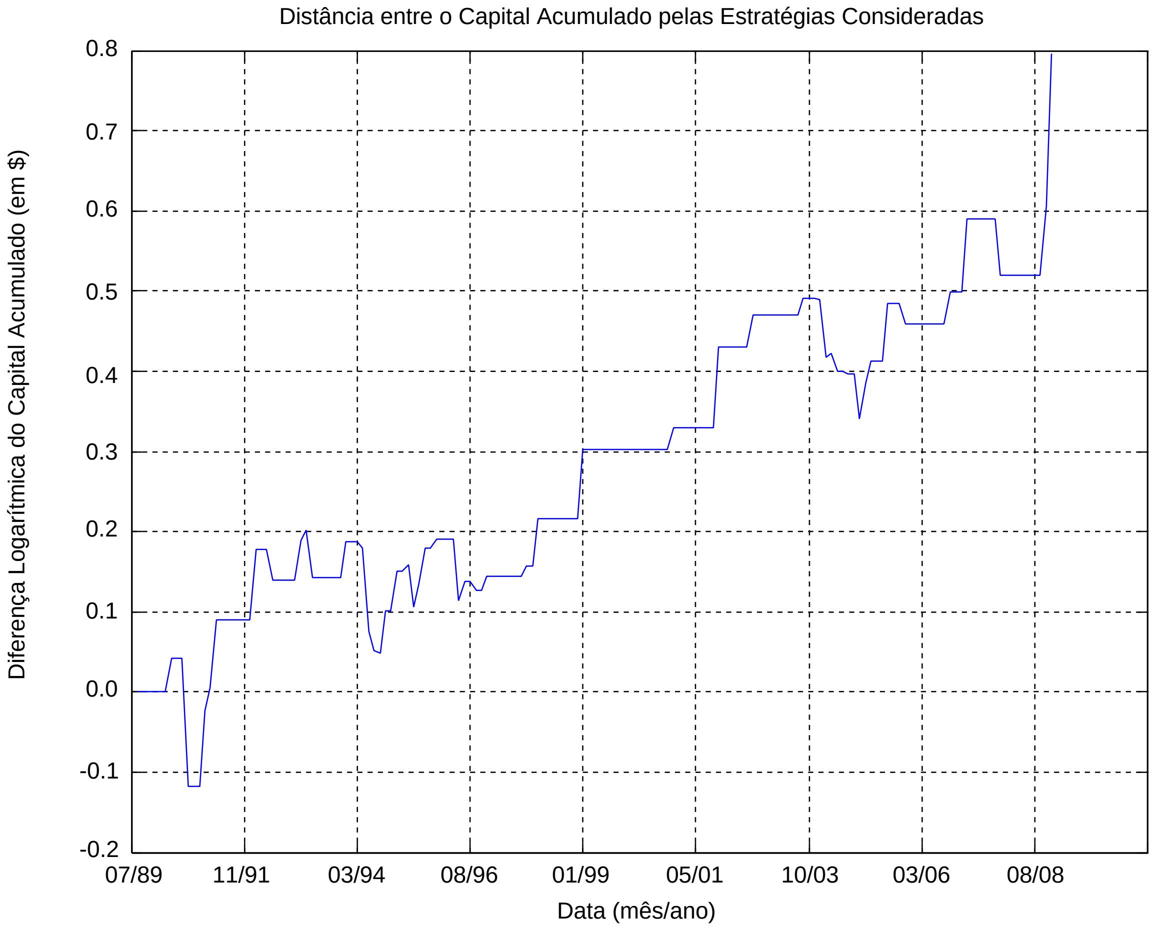


Figura 5.13: Distância entre o capital acumulado pela estratégia que utiliza o modelo HMM e o acumulado pela estratégia *Buy-and-Hold*.

5.3.3 Influência dos Valores Iniciais dos Parâmetros do HMM

O método que utilizamos para encontrar valores iniciais para os parâmetros π e A, descrito na seção 5.1.2, sofre de um empecilho. Conforme mostramos, estes valores são obtidos do ajuste feito a partir de uma distribuição PH (no caso, uma coxian geral) ao conjunto dos tempos de duração das tendências. Este ajuste, por sua vez, é feito pelo programa **EMpht**, através de um algoritmo EM; ou seja, o resultado do ajuste depende dos valores iniciais dados a distribuição PH. Assim, a fim de verificarmos a influência que o resultado deste ajuste pode ter sobre nossos resultados finais, decidimos fazer uma análise de sensibilidade dos últimos em relação aos primeiros.

O programa **EMpht**, por padrão, inicializa aleatoriamente os parâmetros da distribuição PH que será ajustada. Assim, decidimos fazer 101 ajustes diferentes com o programa e, a partir de cada um, utilizando o método de estimação de parâmetros iniciais descrito na seção 5.1.2, obtivemos 101 conjuntos distintos de valores iniciais de parâmetros para nosso modelo HMM. Realizamos, então, 101 vezes nossos experimentos, e verificamos a variabilidade dos resultados finais. Os resultados desta verificação podem ser vistos no apêndice A.3.

Infelizmente, observamos que estes ajustes iniciais têm, sim, influência considerável nos resultados finais[7], como pode ser visto na Figura 5.14. Nela, é ilustrada a taxa de acerto (D_{stat}) obtida por cada um dos 101 experimentos realizados, para seis configurações distintas do HMM (nas quais varia, apenas, o número de estados), e também sua média e seu desvio padrão, considerando uma janela de previsão de 20 dias, e que o período de investimento tem início em agosto de 2001 e se encerra em outubro de 2008. A variabilidade dos resultados obtidos ilustra como o HMM é sensível aos valores iniciais de seus parâmetros.

[7]Verificamos então, se havia alguma relação entre a qualidade do ajuste feita pelo EMpht e qualidade do modelo resultante, ou seja, se os ajustes com maior *likelihood* resultavam nos melhores modelos. No entanto, não observamos qualquer relação óbvia.

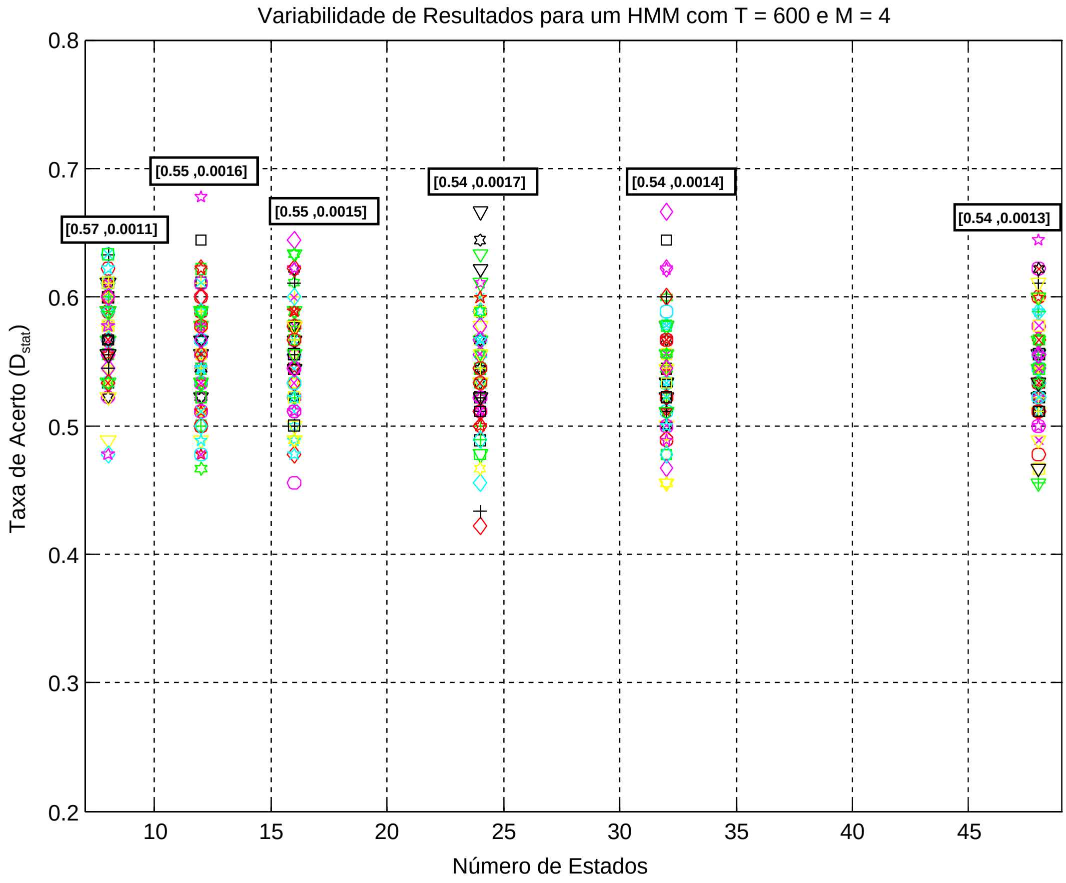


Figura 5.14: Taxa de acerto obtida por cada um dos 101 experimentos realizados para seis configurações distintas do HMM. A média e o desvio padrão de cada configuração são apresentados, respectivamente, dentro das caixas pretas.

Assim, os resultados apresentados nas seções 5.3.1 e 5.3.2 são os melhores que obtivemos, em termos de rentabilidade ao investidor, dada a configuração do modelo HMM escolhida (número de estados, janela de treinamento, etc.) para as duas janelas de previsão consideradas, de 20 e 30 dias.

5.4 Comparação com Outras Estratégias

Em separado, os resultados apresentados acima não são suficientes para convencer um investidor a adotar a metodologia proposta neste trabalho. Para isto, precisamos também mostrar que ela tem desempenho superior à outras estratégias existentes, especialmente as mais simples. Evidentemente, não nos cabe, aqui, fazer uma análise comparativa entre todos os modelos preditivos e estratégias de investimento já elaborados. Por isso, decidimos fazer tal comparação considerando apenas dois métodos de previsão clássicos, extremamente simples: o preditor replicativo e o passeio aleatório.

Selecionamos estes dois, pois consideramo-os uma referência. Se não é possível prever os movimentos do preço do petróleo, então qualquer estratégia (ou modelo preditivo) elaborada terá desempenho, no máximo, igual a ambos.

5.4.1 Preditor Replicativo

Dado uma série temporal, a estratégia do preditor replicativo assume que a melhor previsão para a próxima amostra é a amostra atual. Assim, a previsão de y_t é $\hat{y}_t = y_{t-1}$. Em nosso caso, como estamos querendo prever o retorno acumulado pelo WTI em um intervalo futuro de F dias, o valor previsto por esta estratégia será igual ao seu retorno acumulado nos F dias anteriores.

A Figura 5.15 ilustra a evolução do capital de um investidor que adotasse a estratégia do preditor replicativo, entre agosto de 2001 e outubro de 2008, utilizando uma janela de previsão de 20 dias. Novamente, assumimos que é feito um investimento inicial de 100 dólares, que, em nenhum momento, há adição de capital externo, e que não existem custos de transação.

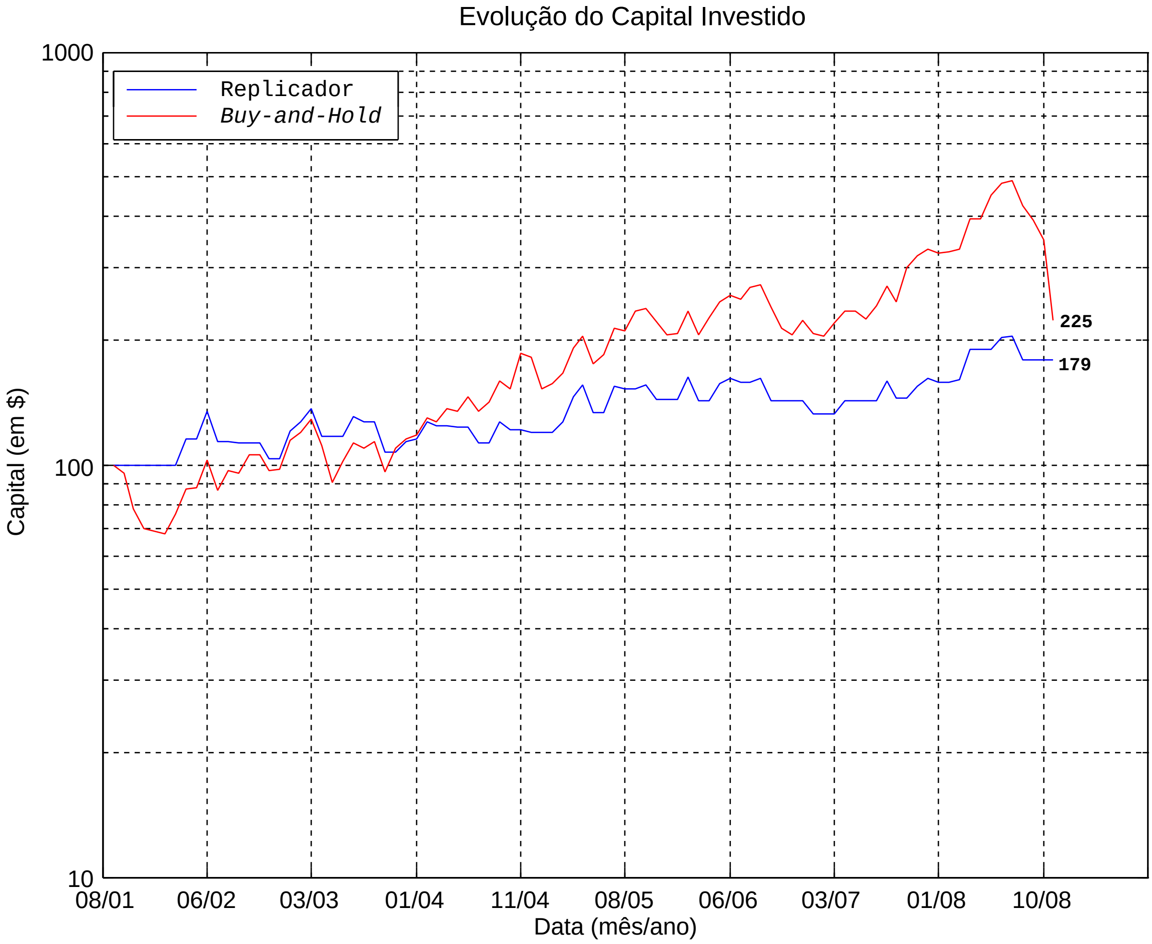


Figura 5.15: Evolução do capital investido com o preditor replicativo e com o *Buy-and-Hold*. A taxa de acerto D_{stat} do preditor replicativo foi de 48.89%.

No período considerado, o preditor replicativo obteve uma taxa de acerto $D_{stat} =$ 48.89%, o que significa que, das 90 previsões feitas, ele conseguiu acertar a direção do preço do WTI em apenas 44 delas. O capital final, possuído pelo investidor que utilizou esta estratégia, é de 179 dólares, o que representa uma rentabilidade acumulada de 79% em 7 anos. Neste mesmo período, a estratégia *Buy-and-Hold* apresentou uma rentabilidade de 125%, resultando em um capital final de 225 dólares.

A Tabela 5.6 compara a rentabilidade anual da estratégia replicadora à rentabilidade anual do HMM (considerando $F = 20$ dias) e à do mercado, no período testado. Observe que em apenas 1 dos 7 anos, a rentabilidade da estratégia replicadora foi igual ou superior a rentabilidade do mercado, e que, em nenhum momento, ela conseguiu superar a rentabilidade do HMM.

Tabela 5.6: Rentabilidades Anuais das Estratégias Replicadora, HMM e *Buy-and-Hold* entre 2002 e 2008

Ano	Rentabilidade		
	Replicador	HMM	*Buy-and-Hold*
2002	18.01%	53.09%	56.40%
2003	−03.08%	46.14%	04.22%
2004	04.59%	22.41%	33.36%
2005	27.14%	38.89%	40.81%
2006	−08.44%	−00.33%	−00.33%
2007	14.02%	33.31%	57.73%
2008	12.98%	21.82%	−34.59%

A exemplo do que fizemos na seção 5.3.1, decidimos também verificar os resultados desta estratégia replicadora caso utilizassemo-na desde o início da série. A Figura 5.16 ilustra a evolução do capital de um investidor que adotasse a estratégia do preditor replicativo, entre maio de 1988 e outubro de 2008, utilizando uma janela de previsão de 20 dias. Novamente, assumimos que é feito um investimento inicial de 100 dólares, que, em nenhum momento, há adição de capital externo, e que não existem custos de transação.

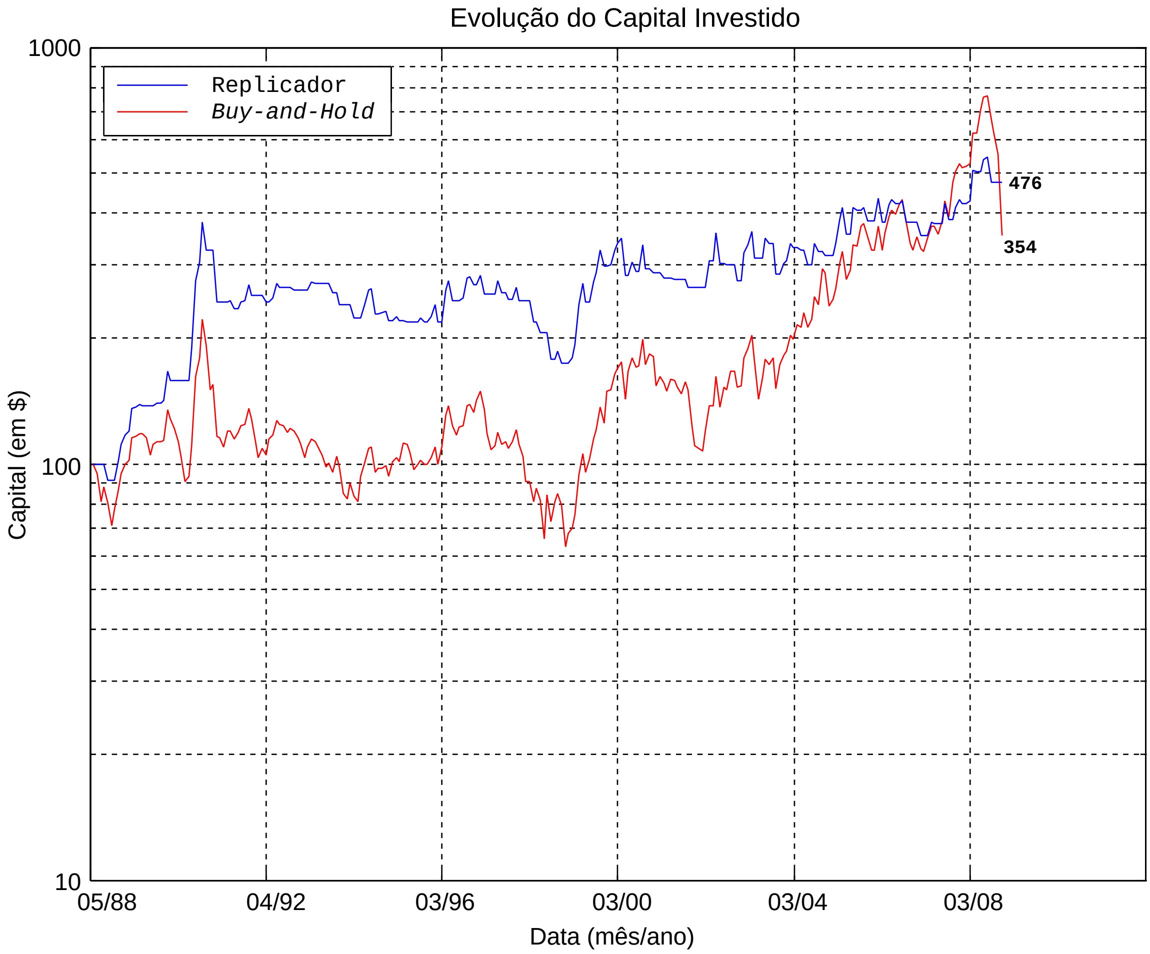


Figura 5.16: Evolução do capital investido com o preditor replicativo e com o *Buy-and-Hold*. A taxa de acerto D_{stat} do preditor replicativo foi de 51.94%.

No período considerado, a taxa de acerto D_{stat} do preditor replicativo foi de 51.94%, o que significa que, das 258 previsões feitas, ele conseguiu acertar a direção do preço do WTI em apenas 134 delas. O capital final, possuído pelo investidor que utilizou esta estratégia, é de 476 dólares, o que representa uma rentabilidade de 376% em, aproximadamente, 21 anos. Neste mesmo período, a estratégia *Buy-and-Hold* apresentou uma rentabilidade de 254%, resultando em um capital final de 354 dólares.

A Tabela 5.7 compara a rentabilidade anual da estratégia replicadora à rentabilidade anual do HMM (considerando $F = 20$ dias) e à do mercado, no período testado. Observe que em apenas 10 dos 21 anos (ou em 47.62% das vezes), a rentabilidade da estratégia replicadora foi igual ou superior a rentabilidade do mercado, enquanto que, para o HMM, esta taxa foi de 61.90% (ou 13) anos. Além disso, a rentabilidade do HMM supera a do Replicador em 16 dos 21 anos em questão (ou 76.19% das vezes).

Tabela 5.7: Rentabilidades Anuais das Estratégias Replicadora, HMM e *Buy-and-Hold* entre 1988 e 2008

Ano	Rentabilidade		
	Replicador	HMM	*Buy-and-Hold*
1988	13.43%	15.61%	−03.42%
1989	34.68%	25.24%	27.50%
1990	113.58%	38.99%	30.48%
1991	−22.08%	−28.82%	−32.76%
1992	03.01%	10.65%	01.91%
1993	−07.76%	−05.59%	−27.16%
1994	−08.82%	08.52%	25.20%
1995	06.17%	15.25%	09.95%
1996	14.45%	11.67%	32.55%
1997	−11.19%	−05.70%	−31.84%
1998	−28.83%	04.22%	−31.20%
1999	65.50%	47.74%	112.28%
2000	−00.23%	21.83%	03.72%
2001	−08.01%	02.95%	−25.33%
2002	18.01%	53.09%	56.40%
2003	−03.08%	46.14%	04.22%
2004	04.59%	22.41%	33.36%
2005	27.14%	38.89%	40.81%
2006	−08.44%	−00.33%	−00.33%
2007	14.02%	33.31%	57.73%
2008	12.98%	21.82%	−34.59%

5.4.2 Passeio Aleatório

O modelo de passeio aleatório, discutido na seção 3.2.1, considera que cada mudança sucessiva em uma série temporal y_t é extraída, independentemente, de uma distribuição de probabilidade com média 0 e variança σ^2. Se, de fato, o preço do petróleo segue um passeio aleatório, então é impossível prever seus movimentos futuros, e, portanto, qualquer rentabilidade significativa que houvesse sido auferida por um investidor, através da utilização de um modelo matemático, seria conseqüência da sorte. Assim, decidimos verificar os resultados obtidos por um modelo de passeio aleatório, a fim de averiguar a possibilidade de que os resultados conseguidos com o HMM tenham sido obra do acaso.

Resolvemos, então, construir a seguinte estratégia de investimento: a cada F dias (onde F é o tamanho da janela de previsão) o investidor joga uma moeda, não viciada, e, se o resultado for *cara*, ele compra petróleo no mercado *spot*, caso tenha capital disponível; se for coroa, ele vende no mercado *spot* todo óleo que possuir. A posição adotada não é alterada até que se passem os F dias.

Realizamos, ao todo, 10000 ensaios, utilizando uma janela de previsão $F = 20$ dias. A fim de compararmos os resultados deste modelo, aos obtidos pelo HMM, consideramos que os investimentos são realizados entre agosto de 2001 e outubro de 2008. Assumimos, novamente, que o investidor dispõe de um capital inicial de 100 dólares, e que não há, em nenhum momento, introdução de capital externo. Assumimos, também, que não existem custos de transação.

Conforme discutido na seção 5.3.3, os resultados do HMM são sensíveis à escolha de seus parâmetros iniciais. Por esta razão, não basta, simplesmente, compararmos os resultados de um único ensaio, aos obtidos pelo passeio aleatório, sob o risco de que este ensaio tenha, por acaso, apresentado bons resultados. É necessário, também, verificarmos se, em geral, o HMM apresenta resultados superiores aos do passeio aleatório. Assim, utilizamos as métricas colhidas na análise de sensibilidade feita para o HMM, apresentada no apêndice A.3, e comparamo-as as calculadas a partir dos ensaios feitos para o passeio aleatório. Estas métricas são: (1) taxa de acerto média (valor médio observado para a estatística D_{stat}); (2) variância da taxa de acerto; (3) probabilidade de $D_{stat} \geq 57\%$; (4) probabilidade da rentabilidade da metodologia ser maior que a do mercado (ou seja, ser superior a estratégia *Buy-and-Hold*); (5) probabilidade de se auferir rentabilidade negativa; (6) e o 90º percentil da rentabilidade. A justificativa para a utilização de cada uma destas estatísticas pode ser encontrada no apêndice A.3.

Na Tabela 5.8, é feita esta comparação[8]. Observando-a, fica evidente a superioridade do HMM em relação ao modelo de passeio aleatório.

Tabela 5.8: Comparação entre os Resultados do HMM e do Passeio Aleatório

Métrica	Passeio Aleatório	HMM
$E[D_{stat}]$	0.5009	0.5745
$Var[D_{stat}]$	$2.80e-03$	$1.10e-03$
$P[D_{stat} \geq 0.57]$	0.0895	0.5248
$P[\text{Rentabilidade} \geq \text{BH}]$	0.2225	0.9307
$P[\text{Perda}]$	0.2104	0.0

[8]A configuração do HMM utilizada é aquela especificada em 5.1.2.

Capítulo 6

Conclusão

Vimos, no capítulo 1, a importância de se construir um modelo de previsão do preço do petróleo. Vimos também que, devido a enorme quantidade de fatores que o afetam, a construção de modelos de regressão preditivos é uma tarefa difícil, principalmente devido ao fato de que, nestes, a previsão do preço está, normalmente, condicionada a previsão das próprias variáveis explanatórias, como é o caso do modelo apresentado na seção 3.1.

Esta dificuldade inerente aos modelos de regressão estimulou a busca de formas alternativas de se tentar fazer esta previsão, utilizando, por exemplo, modelos de séries temporais. Os modelos de séries temporais lineares, como o AR, o ARMA e o ARIMA, discutidos na seção 3.2, não apresentaram resultados preditivos satisfatórios, como verificado em [26], sugerindo que a série de preço do petróleo está longe de ser um processo linear.

Recentemente, devido ao aumento do poder computacional, uma nova classe de modelos preditivos não lineares vem sendo testada, dentre os quais podemos citar as Redes Neurais. Os resultados apresentados por [25, 29, 26, 30] sugerem que, estes sim, conseguem capturar um pouco da dinâmica dos preços, e apresentam capacidade preditiva satisfatória. Dentro deste contexto, esta dissertação teve como objetivo propor a utilização de um outro modelo não linear de previsão de séries temporais, ainda pouco utilizado na área financeira, conhecido como modelo de Markov oculto

(HMM), e verificar sua eficácia na previsão do comportamento do preço do petróleo.

Elaboramos uma metodologia de previsão que consiste de, basicamente, duas etapas. Na primeira, devido a alta volatilidade existente, suavizamos a série de preços utilizando as Wavelets, com o objetivo de ressaltar suas reais tendências e, assim, destacar seus movimentos significativos. Em seguida, esta série suavizada é fornecida como entrada ao HMM, e utilizada para estimar seus parâmetros. Determinados estes valores, o HMM é, então, utilizado para fazer as previsões. Decidimos por prever a distribuição completa do retorno acumulado pelo preço do petróleo, ao invés de, somente, seu valor esperado, pois entendemos que a primeira traz muito mais informações úteis ao investidor, e permite que ele, por exemplo, possa fazer uma análise de risco de suas decisões.

Os resultados apresentados nas seções 5.3 e 5.4 encorajam a utilização do HMM na previsão dos movimentos do preço do petróleo e, certamente, de outros ativos financeiros (como preço de ações ou outras *commodities*). Verificamos que, utilizando-o, é possível identificar os movimentos significativos dos preços (suas tendências) e, a partir destes, calcular a probabilidade de que, no futuro próximo, se mantenham ou se invertam.

Entretanto, ainda há um aspecto que deve ser visto com muito cuidado e estudado mais a fundo. Conforme comentamos ao longo desta dissertação, o HMM é sensível a escolha dos valores iniciais de seus parâmetros. Apesar de termos elaborado uma metodologia para escolhe-lôs, ela não é conclusiva, ou seja, não nos fornece, sempre, um conjunto ideal de valores. Por esta razão, foi necessário fazermos diversos ensaios para uma mesma configuração do HMM, a fim de escolhermos o conjunto de valores iniciais dos parâmetros que levam aos melhores resultados. Assim, acreditamos que o próximo passo deste trabalho é estudar cuidadosamente a diferença entre estes valores, identificar a razão pela qual uns resultam em bons modelos, e outros não, e, assim, elaborar um método que otimize-os.

Por fim, achamos importante ressaltar que o modelo desenvolvido neste trabalho, assim como qualquer outro modelo matemático, jamais deve ser utilizado como única ferramenta de análise de um processo. O objetivo de um modelo deve ser o de auxiliar seu usuário, fornecendo informações extras, e não substituí-lo como tomador de decisões, especialmente no caso de decisões de investimento que, como vimos, podem depender de inúmeros fatores.

6.1 Trabalhos Futuros

Acreditamos que este trabalho pode servir como ponto de origem à vários outros. Dentre os trabalhos futuros que visualizamos, podemos destacar:

- Elaborar um algoritmo para otimizar a escolha dos valores iniciais dos parâmetros do HMM;
- Construir novas estruturas para a cadeia de Markov do HMM, e verificar se estas melhoram sua capacidade preditiva;
- Estudar em mais detalhes as características de diferentes Wavelets, a fim de determinar qual delas é a mais adequada para se utilizar na suavização da série de preços do petróleo;
- Fazer uma análise de sensibilidade mais completa dos elementos do HMM, para se identificar, com maior precisão, sua configuração ideal;
- Existe um novo tipo de HMM, chamado de HMM hierárquico, no qual o processo de emissão de símbolos é governado por uma cadeia de Markov, ao invés de uma simples distribuição de probabilidades. Assim, cada estado oculto da cadeia contém, dentro de si, uma outra cadeia de Markov, que é responsável pela emissão dos símbolos. Uma cadeia como esta foi, recentemente, utilizada por [33, 34] para prever a taxa de perda de pacotes na Internet, observada por uma aplicação de transmissão de voz em tempo real. Seu objetivo, ao utilizar este tipo de cadeia, foi capturar, com a cadeia oculta, os estados de longo prazo das perdas e, com as cadeias hierárquicas (que se encontram dentro de

cada estado oculto), as variações de curto prazo nas estatísticas de perda. Entendemos que um modelo como este é perfeitamente plausível de ser utilizado com séries financeiras. A exemplo de [33, 34], poderíamos tentar capturar, com os estados ocultos, as tendências de longo prazo da série e, com as cadeias hierárquicas, a dinâmica de curto prazo em cada um destes estados.

Apêndice A

Conceitos Teóricos Auxiliares

A.1 Distribuições *Phase-Type*

Uma distribuição *Phase-Type* (PH) consiste em uma mistura de distribuições exponenciais, e é caracterizada por uma cadeia de Markov finita absorvente[1]. O número N de fases da distribuição é igual ao número de estados transientes[2] da cadeia associada ao processo. A distribuição PH é utilizada para representar variáveis aleatórias que medem o tempo decorrido até que o estado absorvente da cadeia seja atingido.

Assim, uma distribuição PH é definida pelo vetor $\pi_i = P[q_1 = S_i],\ 1 \leq i \leq N$, que representa a probabilidade do processo ser iniciado no estado i, e pela matriz

$$Q = \begin{bmatrix} \mathbf{T} & \mathbf{t} \\ (0, \cdots, 0) & 0 \end{bmatrix} \tag{A.1}$$

onde $\mathbf{T}$ é uma matriz $(N \times N)$ que representa as transições entre os estados transientes, e $\mathbf{t}$ é o vetor $(N \times 1)$ que contém o valor das transições dos estados transientes ao absorvente. $\mathbf{T}$ e $\mathbf{t}$ estão relacionados pela equação $\mathbf{t} = -\mathbf{T} \cdot 1$.

[1] Uma cadeia de Markov é dita absorvente se contém um estado absorvente.

[2] Um estado é classificado como transiente se a probabilidade de nunca se retornar a ele, após tê-lo deixado, é maior que zero. Logo estado absorvente não é transiente.

A Figura A.1 ilustra alguns exemplos de cadeias de Markov que representam distribuições PH. Uma discussão mais detalhada sobre esta classe de distribuições, assim como suas propriedades, pode ser vista em [41, 42, 39].

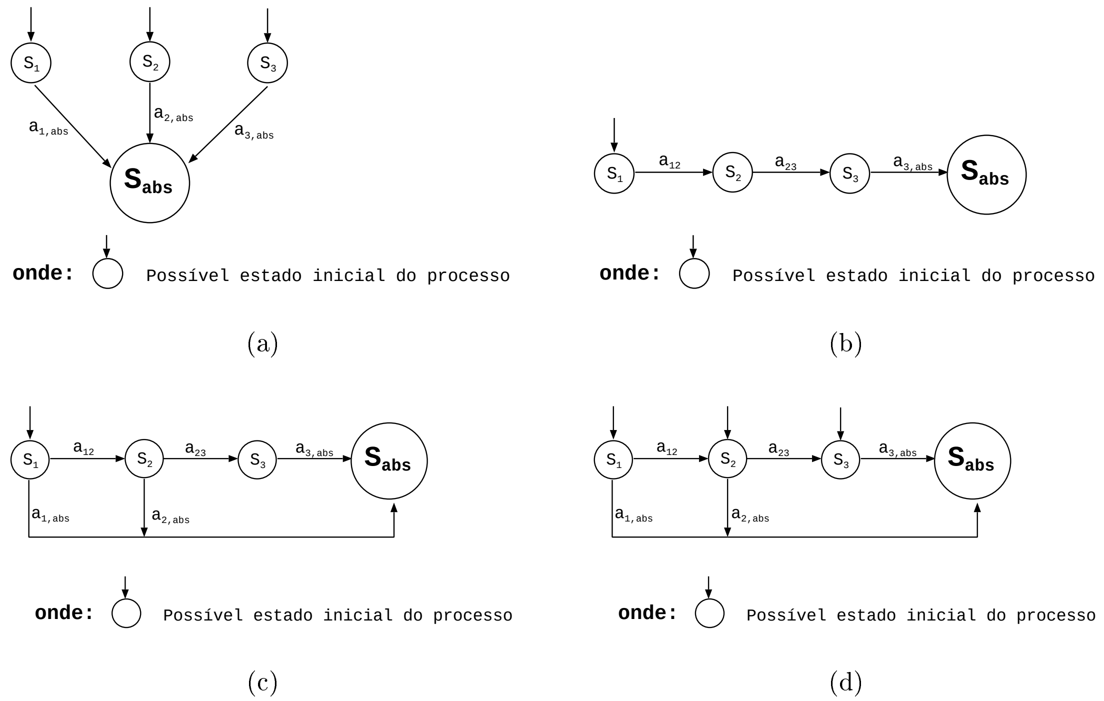


Figura A.1: Exemplos de Distribuições *Phase-Type*. (a) Hiperexponencial (b) Soma de Exponenciais (c) Coxian (d) Coxian Geral.

A.2 Técnica de Uniformização

A técnica de uniformização foi proposta por Jensen, em 1953 [43]. Seu objetivo é transformar uma cadeia de Markov de tempo contínuo em uma cadeia de Markov análoga, de tempo discreto. Isto é de grande utilidade pois, em muitos casos, o cálculo de diversas medidas de interesse é muito mais simples de ser feito em cadeias de tempo discreto, do que em cadeias de tempo contínuo. A seguir, descrevemos esta técnica [40].

Suponha a cadeia de Markov de tempo contínuo $\mathcal{X}$, composta por N estados, cuja matriz de transição Q é dada por

$$Q = \{q_{ij}\} = \begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1N} \\ q_{21} & -q_2 & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & -q_N \end{bmatrix}$$

Para $i \neq j$, q_{ij} é a taxa exponencial associada a transição que leva do estado S_i ao estado S_j, e $q_i = \sum_{j \neq i} q_{ij}$ define a taxa exponencial de saída do estado S_i[3]. Em $\mathcal{X}$, a probabilidade de se transicionar para o estado S_j, dado que houve uma transição partindo de S_i, é $a_{ij} = q_{ij}/q_i$, e o tempo médio de permanência no estado S_i é exponencialmente distribuído com taxa q_i.

Agora considere as seguintes transformações em cada estado S_i de $\mathcal{X}$:

1. é criada uma transição recorrente, ou seja, que permite a cadeia transicionar de S_i de volta a S_i;
2. é definida uma nova taxa de saída $\Lambda_i \geq q_i$ para o estado S_i, de modo que, após permanecer um tempo exponencialmente distribuído com taxa Λ_i em S_i, a cadeia faz uma transição (para o mesmo, ou para outro estado).

Desse modo, as probabilidades de transição, nesta nova cadeia $\mathcal{X}_i^*$ são dadas por

$$a_{ij}^* = q_{ij}/\Lambda_i \tag{A.2}$$

$$a_{ii}^* = 1 - q_i/\Lambda_i \tag{A.3}$$

Em $\mathcal{X}_i^*$, as probabilidades de se transicionar de S_i para S_j, dado que ocorreu uma transição $S_j \neq S_i$, são idênticas as observadas na cadeia $\mathcal{X}$. Além disso, é possível demonstrar que, em ambas as cadeias, as distribuições do tempo total de permanência em um mesmo estado também são as mesmas. Este fato pode ser verificado da seguinte maneira: suponha que a cadeia $\mathcal{X}_i^*$ encontre-se em S_i, e que, a partir deste estado, ela tenha realizado n transições de volta ao mesmo estado,

[3]Àqueles que desejam relembrar o significado destas taxas, sugerimos que revejam a seção 2.1.2

antes de deixá-lo pela transição $n+1$. Como vimos anteriormente, a cada transição, a cadeia permanece um tempo em S_i, que é exponencialmente distribuído com taxa Λ_i. Portanto, a distribuição do tempo total de permanência em S_i é dada pela soma de n distribuições exponenciais independentes e identicamente distribuídas, o que, por sua vez, define uma distribuição do tipo Erlang, cuja função densidade é dada por

$$E(t; n, \Lambda_i) = \frac{\Lambda_i(\Lambda_i t)^{n-1} e^{\Lambda_i t}}{(n-1)!} \qquad \text{para } t > 0 \tag{A.4}$$

A função (A.4) está condicionada ao fato de que conhecemos o número n de transições feitas ao mesmo estado. Para encontrarmos a distribuição geral, basta descondicioná-la, multiplicando-a pela probabilidade de terem ocorrido n transições. Conforme vimos na seção 2.1.1, a probabilidade de terem ocorrido n transições consecutivas a um mesmo estado é dada por uma distribuição geométrica, com parâmetro p, onde p é a probabilidade de se deixar o estado. Assim, teremos que

$$\begin{aligned} P(T = t) &= \sum_{n=0}^{\infty} \frac{\Lambda_i(\Lambda_i t)^{n-1} e^{\Lambda_i t}}{(n-1)!} \cdot (1 - q_i/\Lambda_i)^{n-1} q_i/\Lambda_i \\ &= q_i e^{-q_i t} \end{aligned} \tag{A.5}$$

ou seja, a distribuição do tempo total de permanência em S_i é exponencial com taxa q_i. Assim, verificamos que as cadeias $\mathcal{X}$ e $\mathcal{X}_i^*$ são equivalentes.

Ao invés de escolhermos uma taxa Λ_i para cada estado, podemos escolher uma única taxa $\Lambda \geq \max(q_i)$ e fazer as mesmas modificações descritas acima, gerando a cadeia $\mathcal{X}^*$. Deste modo, estaremos uniformizando as taxas exponenciais que descrevem o tempo até a ocorrência de uma transição em $\mathcal{X}^*$. É possível provar que ambas as cadeias $\mathcal{X}$ e $\mathcal{X}^*$ são equivalentes. Uma discussão simples sobre a técnica de uniformização pode ser vista em [38, 36]. Para uma descrição mais completa do método, sugerimos [40, 44].

Apesar de a cadeia $\mathcal{X}^*$ ter uma estrutura discreta, ela ainda é uma cadeia contínua, uma vez que o tempo decorrido até uma transição sua é exponencialmente distribuído com taxa Λ. Como sabemos, o modelo com o qual trabalhamos nesta dissertação é discreto, e, portanto, a princípio, não podemos utilizar $\mathcal{X}^*$ diretamente.

No entanto, como os dados que utilizamos para gerar a cadeia $\mathcal{X}$ são discretos[4], estamos interessados em fazer uma descrição do tempo pelo número de transições. Assim, se uniformizarmos $\mathcal{X}$ com a taxa $\Lambda = 1$ (isso implica que $\max(q_i) \leq 1$), podemos ignorar o conceito de tempo contínuo em $\mathcal{X}^*$, e utiliza-la, diretamente, como uma cadeia discreta $\mathcal{X}^d$. Estatisticamente, $\mathcal{X}$, $\mathcal{X}^*$ e $\mathcal{X}^d$ são equivalentes.

A.3 Análise de Sensibilidade aos Parâmetros do HMM

Nesta seção, mostramos os resultados obtidos pela metodologia de previsão proposta, para diferentes configurações do modelo HMM. Conforme dito na seção 5.1.2, foi a partir desta análise empírica que escolhemos os valores de N, M, V, τ e T do modelo HMM.

As Tabelas a seguir apresentam os resultados alcançados pelas diferentes configurações avaliadas, para duas janelas de previsão distintas: 20 e 30 dias. Para cada uma, realizamos 101 ensaios distintos, cada um contendo valores iniciais de π e A diferentes dos demais.

Devido à sensibilidade observada no desempenho do modelo com relação aos valores iniciais dos seus parâmetros, decidimos avaliar cada configuração pelas seguintes métricas: (1) taxa de acerto média (valor médio observado para a estatística D_{stat}); (2) variância da taxa de acerto; (3) probabilidade de $D_{stat} \geq 57\%$; (4) probabilidade da rentabilidade da metodologia ser maior que a do mercado (ou seja, ser superior a estratégia *Buy-and-Hold*); (5) probabilidade de se auferir rentabilidade negativa; (6) e o 90º percentil da rentabilidade.

As duas primeiras métricas citadas no parágrafo anterior indicam tanto o desempenho esperado ao se adotar a configuração quanto sua variabilidade. Observando-as, é possível fazer uma rápida avaliação de cada alternativa. A terceira métrica indica a probabilidade de, adotada àquela configuração, se conseguir uma taxa de acerto significativa. Escolhemos o valor desta taxa como 57% pois observamos ser

[4] O processo de geração de $\mathcal{X}$ pode ser visto em 5.1.2.

este o limite inferior que garante que a rentabilidade do modelo seja superior a do *Buy-and-Hold*. A quarta métrica mostra o desempenho da configuração quanto a sua rentabilidade, quando comparada a rentabilidade do mercado. A quinta indica a probabilidade de o investidor ter perdas ao adotar o modelo. Por fim, a sexta nos permite comparar, entre si, os melhores modelos (em termos de rentabilidade) de cada configuração (no caso, os 10% melhores).

Observando os resultados da análise feita, apresentados nas tabelas a seguir, fica evidente a diferença de qualidade entre as configurações do HMM. Por exemplo, se compararmos as configurações apresentadas nas Tabelas A.4 e A.3, podemos notar a superioridade da primeira em relação a segunda, tanto em termos da taxa média de acerto, quanto em termos da rentabilidade obtida.

Tabela A.1: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 2 Símbolos, com $T = 600$ e $T = 800$.

	Janela de Treinamento: 600 dias		Janela de Treinamento: 800 dias	
8 estados	$E[D_{stat}] = 0.5441$ $Var[D_{stat}] = 7.32e-04$	$P[D_{stat} \geq 0.57] = 0.1485(15/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.8218(83/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.2277	$E[D_{stat}] = 0.5255$ $Var[D_{stat}] = 3.46e-04$	$P[D_{stat} \geq 0.57] = 0.0(0/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6535(66/101)$ $P[\text{Perda}] = 0.0(0/101)$ 90º percentil : 3.2223
12 estados	$E[D_{stat}] = 0.5469$ $Var[D_{stat}] = 9.32e-04$	$P[D_{stat} \geq 0.57] = 0.2277(23/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6535(66/101)$ $P[\text{Perda}] = 0.0(0/101)$ 90º percentil : 3.3185	$E[D_{stat}] = 0.5305$ $Var[D_{stat}] = 5.67e-04$	$P[D_{stat} \geq 0.57] = 0.0693(7/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6436(65/101)$ $P[\text{Perda}] = 0.0(0/101)$ 90º percentil : 3.2516
16 estados	$E[D_{stat}] = 0.5448$ $Var[D_{stat}] = 1.15e-03$	$P[D_{stat} \geq 0.57] = 0.1980(20/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6337(64/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 3.4526	$E[D_{stat}] = 0.5238$ $Var[D_{stat}] = 5.78e-04$	$P[D_{stat} \geq 0.57] = 0.0495(5/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5644(57/101)$ $P[\text{Perda}] = 0.0(0/101)$ 90º percentil : 3.3762
24 estados	$E[D_{stat}] = 0.5517$ $Var[D_{stat}] = 1.17e-03$	$P[D_{stat} \geq 0.57] = 0.2574(26/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6733(68/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.8556	$E[D_{stat}] = 0.5209$ $Var[D_{stat}] = 8.45e-04$	$P[D_{stat} \geq 0.57] = 0.0594(6/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5347(54/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.6081
32 estados	$E[D_{stat}] = 0.5498$ $Var[D_{stat}] = 1.67e-03$	$P[D_{stat} \geq 0.57] = 0.2772(28/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5743(58/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.1742	$E[D_{stat}] = 0.5165$ $Var[D_{stat}] = 1.21e-03$	$P[D_{stat} \geq 0.57] = 0.0495(5/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5347(54/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.4771
48 estados	$E[D_{stat}] = 0.5380$ $Var[D_{stat}] = 1.43e-03$	$P[D_{stat} \geq 0.57] = 0.1983(17/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.8452	$E[D_{stat}] = 0.5309$ $Var[D_{stat}] = 1.34e-03$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5941(60/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.2384

Tabela A.2: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 2 Símbolos, com $T = 1000$ e $T = 1200$.

	Janela de Treinamento: 1000 dias		Janela de Treinamento: 1200 dias	
8 estados	$E[D_{stat}] = 0.5248$ $Var[D_{stat}] = 5.77e-04$	$P[D_{stat} \geq 0.57] = 0.0495(5/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5644(57/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.5239	$E[D_{stat}] = 0.5609$ $Var[D_{stat}] = 9.42e-04$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5545(56/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 5.0113
12 estados	$E[D_{stat}] = 0.5206$ $Var[D_{stat}] = 6.32e-04$	$P[D_{stat} \geq 0.57] = 0.0297(3/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.3395	$E[D_{stat}] = 0.5730$ $Var[D_{stat}] = 1.27e-03$	$P[D_{stat} \geq 0.57] = 0.5050(51/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5644(57/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.9621
16 estados	$E[D_{stat}] = 0.5128$ $Var[D_{stat}] = 8.26e-04$	$P[D_{stat} \geq 0.57] = 0.0000(0/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2178(22/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.1954	$E[D_{stat}] = 0.5708$ $Var[D_{stat}] = 1.29e-03$	$P[D_{stat} \geq 0.57] = 0.4752(48/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5842(59/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.7081
24 estados	$E[D_{stat}] = 0.5209$ $Var[D_{stat}] = 1.20e-03$	$P[D_{stat} \geq 0.57] = 0.0693(7/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2178(22/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.2211	$E[D_{stat}] = 0.5613$ $Var[D_{stat}] = 1.28e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4158(42/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.2642
32 estados	$E[D_{stat}] = 0.5176$ $Var[D_{stat}] = 7.92e-04$	$P[D_{stat} \geq 0.57] = 0.0297(3/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1881(19/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.3247	$E[D_{stat}] = 0.5554$ $Var[D_{stat}] = 1.64e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4356(44/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.3760
48 estados	$E[D_{stat}] = 0.5206$ $Var[D_{stat}] = 1.65e-03$	$P[D_{stat} \geq 0.57] = 0.1188(12/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2475(25/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.5827	$E[D_{stat}] = 0.5420$ $Var[D_{stat}] = 2.21e-03$	$P[D_{stat} \geq 0.57] = 0.2277(23/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.5950

Tabela A.3: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 2 Símbolos, com $T = 2000$ e $T = 2400$.

	Janela de Treinamento: 2000 dias	Janela de Treinamento: 2400 dias
8 estados	$E[D_{stat}] = 0.5102$ $Var[D_{stat}] = 6.52e-04$ $P[D_{stat} \geq 0.57] = 0.0000(0/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.0396(4/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 2.6033	$E[D_{stat}] = 0.5266$ $Var[D_{stat}] = 3.60e-04$ $P[D_{stat} \geq 0.57] = 0.0198(2/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3366(34/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 2.8459
12 estados	$E[D_{stat}] = 0.5195$ $Var[D_{stat}] = 8.01e-04$ $P[D_{stat} \geq 0.57] = 0.0396(4/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.0891(9/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 2.7904	$E[D_{stat}] = 0.5325$ $Var[D_{stat}] = 4.93e-04$ $P[D_{stat} \geq 0.57] = 0.0792(8/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2574(26/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 2.8797
16 estados	$E[D_{stat}] = 0.5240$ $Var[D_{stat}] = 1.31e-03$ $P[D_{stat} \geq 0.57] = 0.1188(12/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1584(16/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.0935	$E[D_{stat}] = 0.5389$ $Var[D_{stat}] = 7.04e-04$ $P[D_{stat} \geq 0.57] = 0.1188(12/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3366(34/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.0473
24 estados	$E[D_{stat}] = 0.5333$ $Var[D_{stat}] = 1.71e-03$ $P[D_{stat} \geq 0.57] = 0.1287(13/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2772(28/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 3.2418	$E[D_{stat}] = 0.5365$ $Var[D_{stat}] = 9.50e-04$ $P[D_{stat} \geq 0.57] = 0.1980(20/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4257(43/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.9923
32 estados	$E[D_{stat}] = 0.5330$ $Var[D_{stat}] = 1.54e-03$ $P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1683(17/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 3.1704	$E[D_{stat}] = 0.5313$ $Var[D_{stat}] = 1.50e-03$ $P[D_{stat} \geq 0.57] = 0.2178(22/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4158(42/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.7866
48 estados	$E[D_{stat}] = 0.5356$ $Var[D_{stat}] = 1.68e-03$ $P[D_{stat} \geq 0.57] = 0.1485(15/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1683(17/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.1362	$E[D_{stat}] = 0.5183$ $Var[D_{stat}] = 1.98e-03$ $P[D_{stat} \geq 0.57] = 0.1287(13/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3168(32/101)$ $P[\text{Perda}] = 0.0990(10/101)$ 90º percentil : 3.2528

Tabela A.4: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 4 Símbolos, com $T = 600$ e $T = 800$.

	Janela de Treinamento: 600 dias		Janela de Treinamento: 800 dias	
8 estados	$E[D_{stat}] = 0.5745$ $Var[D_{stat}] = 1.10e-03$	$P[D_{stat} \geq 0.57] = 0.5248(53/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.9307(94/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 5.2455	$E[D_{stat}] = 0.5595$ $Var[D_{stat}] = 8.67e-04$	$P[D_{stat} \geq 0.57] = 0.4158(42/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.8713(88/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.6550
12 estados	$E[D_{stat}] = 0.5537$ $Var[D_{stat}] = 1.55e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7327(74/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.3711	$E[D_{stat}] = 0.5444$ $Var[D_{stat}] = 1.38e-03$	$P[D_{stat} \geq 0.57] = 0.2772(28/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6832(69/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.3318
16 estados	$E[D_{stat}] = 0.5526$ $Var[D_{stat}] = 1.48e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7030(71/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.2584	$E[D_{stat}] = 0.5450$ $Var[D_{stat}] = 1.31e-03$	$P[D_{stat} \geq 0.57] = 0.2871(29/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7129(72/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.4356
24 estados	$E[D_{stat}] = 0.5432$ $Var[D_{stat}] = 1.67e-03$	$P[D_{stat} \geq 0.57] = 0.1683(17/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6535(66/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.0641	$E[D_{stat}] = 0.5414$ $Var[D_{stat}] = 1.60e-03$	$P[D_{stat} \geq 0.57] = 0.2277(23/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6238(63/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.4535
32 estados	$E[D_{stat}] = 0.5389$ $Var[D_{stat}] = 1.41e-03$	$P[D_{stat} \geq 0.57] = 0.1584(16/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5941(60/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.6627	$E[D_{stat}] = 0.5291$ $Var[D_{stat}] = 1.42e-03$	$P[D_{stat} \geq 0.57] = 0.1584(16/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5941(60/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.0158
48 estados	$E[D_{stat}] = 0.5427$ $Var[D_{stat}] = 1.34e-03$	$P[D_{stat} \geq 0.57] = 0.1782(18/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6337(64/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.0927	$E[D_{stat}] = 0.5437$ $Var[D_{stat}] = 1.36e-03$	$P[D_{stat} \geq 0.57] = 0.2178(22/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7426(75/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.5160

Tabela A.5: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 4 Símbolos, com $T = 1000$ e $T = 1200$.

	Janela de Treinamento: 1000 dias		Janela de Treinamento: 1200 dias	
8 estados	$E[D_{stat}] = 0.5417$ $Var[D_{stat}] = 4.24e-04$	$P[D_{stat} \geq 0.57] = 0.0495(5/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4455(45/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.5745	$E[D_{stat}] = 0.5389$ $Var[D_{stat}] = 1.17e-03$	$P[D_{stat} \geq 0.57] = 0.0396(4/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4158(42/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.6076
12 estados	$E[D_{stat}] = 0.5215$ $Var[D_{stat}] = 8.06e-04$	$P[D_{stat} \geq 0.57] = 0.0099(1/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2673(27/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.3040	$E[D_{stat}] = 0.5281$ $Var[D_{stat}] = 1.02e-03$	$P[D_{stat} \geq 0.57] = 0.0396(4/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.7858
16 estados	$E[D_{stat}] = 0.5190$ $Var[D_{stat}] = 8.01e-04$	$P[D_{stat} \geq 0.57] = 0.0099(1/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2970(30/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.2891	$E[D_{stat}] = 0.5340$ $Var[D_{stat}] = 1.08e-03$	$P[D_{stat} \geq 0.57] = 0.0792(8/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4851(49/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.1906
24 estados	$E[D_{stat}] = 0.5287$ $Var[D_{stat}] = 9.98e-04$	$P[D_{stat} \geq 0.57] = 0.0693(7/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 3.7123	$E[D_{stat}] = 0.5296$ $Var[D_{stat}] = 1.07e-03$	$P[D_{stat} \geq 0.57] = 0.0792(8/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.3956
32 estados	$E[D_{stat}] = 0.5362$ $Var[D_{stat}] = 1.24e-03$	$P[D_{stat} \geq 0.57] = 0.1188(12/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4752(48/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.0528	$E[D_{stat}] = 0.5407$ $Var[D_{stat}] = 1.67e-03$	$P[D_{stat} \geq 0.57] = 0.2376(24/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5050(51/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 5.0002
48 estados	$E[D_{stat}] = 0.5260$ $Var[D_{stat}] = 1.16e-03$	$P[D_{stat} \geq 0.57] = 0.0594(6/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2970(30/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.9812	$E[D_{stat}] = 0.5524$ $Var[D_{stat}] = 1.77e-03$	$P[D_{stat} \geq 0.57] = 0.2871(29/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6337(64/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 5.2725

Tabela A.6: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 4 Símbolos, com $T = 2000$ e $T = 2400$.

	Janela de Treinamento: 2000 dias		Janela de Treinamento: 2400 dias	
8 estados	$E[D_{stat}] = 0.5440$ $Var[D_{stat}] = 8.36e-04$	$P[D_{stat} \geq 0.57] = 0.1089(11/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4257(43/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.8396	$E[D_{stat}] = 0.5451$ $Var[D_{stat}] = 5.39e-04$	$P[D_{stat} \geq 0.57] = 0.1485(15/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4752(48/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.5143
12 estados	$E[D_{stat}] = 0.5228$ $Var[D_{stat}] = 1.51e-03$	$P[D_{stat} \geq 0.57] = 0.0594(6/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3762(38/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.9264	$E[D_{stat}] = 0.5244$ $Var[D_{stat}] = 1.17e-03$	$P[D_{stat} \geq 0.57] = 0.1089(11/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3861(39/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 3.3956
16 estados	$E[D_{stat}] = 0.5243$ $Var[D_{stat}] = 1.52e-03$	$P[D_{stat} \geq 0.57] = 0.0693(7/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3168(32/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 3.6221	$E[D_{stat}] = 0.5284$ $Var[D_{stat}] = 1.10e-03$	$P[D_{stat} \geq 0.57] = 0.1188(12/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4950(50/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 3.5858
24 estados	$E[D_{stat}] = 0.5403$ $Var[D_{stat}] = 8.74e-04$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5149(52/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.1256	$E[D_{stat}] = 0.5348$ $Var[D_{stat}] = 1.22e-03$	$P[D_{stat} \geq 0.57] = 0.1584(16/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5149(52/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.0277
32 estados	$E[D_{stat}] = 0.5366$ $Var[D_{stat}] = 1.21e-03$	$P[D_{stat} \geq 0.57] = 0.0990(10/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4950(50/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.0128	$E[D_{stat}] = 0.5337$ $Var[D_{stat}] = 1.06e-03$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5050(51/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.9679
48 estados	$E[D_{stat}] = 0.5554$ $Var[D_{stat}] = 1.41e-03$	$P[D_{stat} \geq 0.57] = 0.2970(30/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5149(52/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.5391	$E[D_{stat}] = 0.5393$ $Var[D_{stat}] = 1.16e-03$	$P[D_{stat} \geq 0.57] = 0.2376(24/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5941(60/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.3968

Tabela A.7: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 6 Símbolos, com $T = 600$ e $T = 800$.

	Janela de Treinamento: 600 dias		Janela de Treinamento: 800 dias	
8 estados	$E[D_{stat}] = 0.5384$ $Var[D_{stat}] = 1.84e-03$	$P[D_{stat} \geq 0.57] = 0.2178(22/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3069(31/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 3.4799	$E[D_{stat}] = 0.5221$ $Var[D_{stat}] = 2.54e-03$	$P[D_{stat} \geq 0.57] = 0.1881(19/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3861(39/101)$ $P[\text{Perda}] = 0.1089(11/101)$ 90º percentil : 4.0574
12 estados	$E[D_{stat}] = 0.5329$ $Var[D_{stat}] = 1.78e-03$	$P[D_{stat} \geq 0.57] = 0.2376(24/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.7751	$E[D_{stat}] = 0.5211$ $Var[D_{stat}] = 1.58e-03$	$P[D_{stat} \geq 0.57] = 0.1287(13/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2376(24/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.2855
16 estados	$E[D_{stat}] = 0.5288$ $Var[D_{stat}] = 2.27e-03$	$P[D_{stat} \geq 0.57] = 0.1782(18/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4950(50/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 3.7616	$E[D_{stat}] = 0.5293$ $Var[D_{stat}] = 1.71e-03$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3861(39/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.8425
24 estados	$E[D_{stat}] = 0.5333$ $Var[D_{stat}] = 2.19e-03$	$P[D_{stat} \geq 0.57] = 0.1980(20/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5149(52/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 4.6468	$E[D_{stat}] = 0.5105$ $Var[D_{stat}] = 2.46e-03$	$P[D_{stat} \geq 0.57] = 0.1287(13/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3069(31/101)$ $P[\text{Perda}] = 0.0792(8/101)$ 90º percentil : 3.5740
32 estados	$E[D_{stat}] = 0.5340$ $Var[D_{stat}] = 1.97e-03$	$P[D_{stat} \geq 0.57] = 0.1980(20/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6436(65/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.1376	$E[D_{stat}] = 0.5233$ $Var[D_{stat}] = 2.21e-03$	$P[D_{stat} \geq 0.57] = 0.1683(17/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4851(49/101)$ $P[\text{Perda}] = 0.1089(11/101)$ 90º percentil : 3.7947
48 estados	$E[D_{stat}] = 0.5349$ $Var[D_{stat}] = 2.14e-03$	$P[D_{stat} \geq 0.57] = 0.2178(22/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5842(59/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.1854	$E[D_{stat}] = 0.5195$ $Var[D_{stat}] = 2.02e-03$	$P[D_{stat} \geq 0.57] = 0.1287(13/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4257(43/101)$ $P[\text{Perda}] = 0.0693(7/101)$ 90º percentil : 3.5403

Tabela A.8: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 6 Símbolos, com $T = 1000$ e $T = 1200$.

	Janela de Treinamento: 1000 dias		Janela de Treinamento: 1200 dias	
8 estados	$E[D_{stat}] = 0.5154$ $Var[D_{stat}] = 2.21e-03$	$P[D_{stat} \geq 0.57] = 0.0792(8/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1089(11/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 2.8511	$E[D_{stat}] = 0.5205$ $Var[D_{stat}] = 3.38e-03$	$P[D_{stat} \geq 0.57] = 0.1782(18/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$ $P[\text{Perda}] = 0.0891(9/101)$ 90º percentil : 4.2546
12 estados	$E[D_{stat}] = 0.5096$ $Var[D_{stat}] = 1.46e-03$	$P[D_{stat} \geq 0.57] = 0.0198(2/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1188(12/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 2.9529	$E[D_{stat}] = 0.5243$ $Var[D_{stat}] = 2.24e-03$	$P[D_{stat} \geq 0.57] = 0.1485(15/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3762(38/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.8789
16 estados	$E[D_{stat}] = 0.5064$ $Var[D_{stat}] = 1.76e-03$	$P[D_{stat} \geq 0.57] = 0.0198(2/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1485(15/101)$ $P[\text{Perda}] = 0.0990(10/101)$ 90º percentil : 3.0442	$E[D_{stat}] = 0.5369$ $Var[D_{stat}] = 2.19e-03$	$P[D_{stat} \geq 0.57] = 0.2673(27/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.3428
24 estados	$E[D_{stat}] = 0.5124$ $Var[D_{stat}] = 1.78e-03$	$P[D_{stat} \geq 0.57] = 0.0693(7/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2277(23/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 3.3342	$E[D_{stat}] = 0.5312$ $Var[D_{stat}] = 1.99e-03$	$P[D_{stat} \geq 0.57] = 0.1881(19/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3267(33/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.1672
32 estados	$E[D_{stat}] = 0.5161$ $Var[D_{stat}] = 1.55e-03$	$P[D_{stat} \geq 0.57] = 0.0693(7/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2475(25/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.2204	$E[D_{stat}] = 0.5467$ $Var[D_{stat}] = 3.19e-03$	$P[D_{stat} \geq 0.57] = 0.2871(29/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4653(47/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.7640
48 estados	$E[D_{stat}] = 0.5231$ $Var[D_{stat}] = 2.06e-03$	$P[D_{stat} \geq 0.57] = 0.1089(11/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3069(31/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.4993	$E[D_{stat}] = 0.5505$ $Var[D_{stat}] = 2.66e-03$	$P[D_{stat} \geq 0.57] = 0.3069(31/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 5.0887

Tabela A.9: Resultados Experimentais para uma janela de previsão de 20 dias: HMM de 6 Símbolos, com $T = 2000$ e $T = 2400$.

	Janela de Treinamento: 2000 dias		Janela de Treinamento: 2400 dias	
8 estados	$E[D_{stat}] = 0.5384$ $Var[D_{stat}] = 2.29e-03$	$P[D_{stat} \geq 0.57] = 0.2376(24/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 5.4041	$E[D_{stat}] = 0.5341$ $Var[D_{stat}] = 2.94e-03$	$P[D_{stat} \geq 0.57] = 0.3465(35/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5248(53/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 4.7646
12 estados	$E[D_{stat}] = 0.5206$ $Var[D_{stat}] = 3.21e-03$	$P[D_{stat} \geq 0.57] = 0.1584(16/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2772(28/101)$ $P[\text{Perda}] = 0.0792(8/101)$ 90º percentil : 4.0044	$E[D_{stat}] = 0.5107$ $Var[D_{stat}] = 2.13e-03$	$P[D_{stat} \geq 0.57] = 0.1089(11/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3267(33/101)$ $P[\text{Perda}] = 0.0891(9/101)$ 90º percentil : 2.9040
16 estados	$E[D_{stat}] = 0.5324$ $Var[D_{stat}] = 2.07e-03$	$P[D_{stat} \geq 0.57] = 0.1881(19/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3366(34/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.1504	$E[D_{stat}] = 0.5258$ $Var[D_{stat}] = 1.63e-03$	$P[D_{stat} \geq 0.57] = 0.1287(13/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.2516
24 estados	$E[D_{stat}] = 0.5291$ $Var[D_{stat}] = 2.13e-03$	$P[D_{stat} \geq 0.57] = 0.1980(20/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3069(31/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.1670	$E[D_{stat}] = 0.5179$ $Var[D_{stat}] = 2.02e-03$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3762(38/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 3.4972
32 estados	$E[D_{stat}] = 0.5365$ $Var[D_{stat}] = 2.14e-03$	$P[D_{stat} \geq 0.57] = 0.1782(18/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2970(30/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.0874	$E[D_{stat}] = 0.5305$ $Var[D_{stat}] = 2.17e-03$	$P[D_{stat} \geq 0.57] = 0.2079(21/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0792(8/101)$ 90º percentil : 4.3630
48 estados	$E[D_{stat}] = 0.5465$ $Var[D_{stat}] = 2.45e-03$	$P[D_{stat} \geq 0.57] = 0.3168(32/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4950(50/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.7355	$E[D_{stat}] = 0.5382$ $Var[D_{stat}] = 2.56e-03$	$P[D_{stat} \geq 0.57] = 0.2772(28/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6238(63/101)$ $P[\text{Perda}] = 0.0792(8/101)$ 90º percentil : 4.6138

Tabela A.10: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 2 Símbolos, com $T = 600$ e $T = 900$.

	Janela de Treinamento: 600 dias		Janela de Treinamento: 900 dias	
8 estados	$E[D_{stat}] = 0.5561$ $Var[D_{stat}] = 1.34e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7525(76/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.4689	$E[D_{stat}] = 0.6058$ $Var[D_{stat}] = 1.07e-03$	$P[D_{stat} \geq 0.57] = 0.9010(91/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5050(51/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.9669
12 estados	$E[D_{stat}] = 0.5568$ $Var[D_{stat}] = 1.98e-03$	$P[D_{stat} \geq 0.57] = 0.3168(32/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6337(64/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.7947	$E[D_{stat}] = 0.5984$ $Var[D_{stat}] = 2.13e-03$	$P[D_{stat} \geq 0.57] = 0.8020(81/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6535(66/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.5797
16 estados	$E[D_{stat}] = 0.5660$ $Var[D_{stat}] = 1.49e-03$	$P[D_{stat} \geq 0.57] = 0.4257(43/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7327(74/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.3952	$E[D_{stat}] = 0.5855$ $Var[D_{stat}] = 1.69e-03$	$P[D_{stat} \geq 0.57] = 0.6931(70/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3861(39/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.7576
24 estados	$E[D_{stat}] = 0.5614$ $Var[D_{stat}] = 1.39e-03$	$P[D_{stat} \geq 0.57] = 0.3564(36/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.7228(73/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.6624	$E[D_{stat}] = 0.5852$ $Var[D_{stat}] = 2.62e-03$	$P[D_{stat} \geq 0.57] = 0.6832(69/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5842(59/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.7481
32 estados	$E[D_{stat}] = 0.5606$ $Var[D_{stat}] = 1.94e-03$	$P[D_{stat} \geq 0.57] = 0.3861(39/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6337(64/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.0886	$E[D_{stat}] = 0.5729$ $Var[D_{stat}] = 2.56e-03$	$P[D_{stat} \geq 0.57] = 0.5743(58/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4455(45/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.1312
48 estados	$E[D_{stat}] = 0.5612$ $Var[D_{stat}] = 1.88e-03$	$P[D_{stat} \geq 0.57] = 0.3663(37/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5446(55/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 3.8071	$E[D_{stat}] = 0.5818$ $Var[D_{stat}] = 2.55e-03$	$P[D_{stat} \geq 0.57] = 0.6040(61/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6535(66/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.9611

Tabela A.11: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 2 Símbolos, com $T = 1200$ e $T = 2100$.

	Janela de Treinamento: 1200 dias		Janela de Treinamento: 2100 dias	
8 estados	$E[D_{stat}] = 0.6298$ $Var[D_{stat}] = 1.27e-03$	$P[D_{stat} \geq 0.57] = 0.9703(98/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6436(65/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.0101	$E[D_{stat}] = 0.5694$ $Var[D_{stat}] = 1.91e-03$	$P[D_{stat} \geq 0.57] = 0.5545(56/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1584(16/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 2.8918
12 estados	$E[D_{stat}] = 0.6038$ $Var[D_{stat}] = 2.42e-03$	$P[D_{stat} \geq 0.57] = 0.8020(81/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5743(58/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.4511	$E[D_{stat}] = 0.5637$ $Var[D_{stat}] = 2.23e-03$	$P[D_{stat} \geq 0.57] = 0.4950(50/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1881(19/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.1339
16 estados	$E[D_{stat}] = 0.6124$ $Var[D_{stat}] = 1.56e-03$	$P[D_{stat} \geq 0.57] = 0.8812(89/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5842(59/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.1831	$E[D_{stat}] = 0.5622$ $Var[D_{stat}] = 2.15e-03$	$P[D_{stat} \geq 0.57] = 0.4356(44/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2079(21/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.2106
24 estados	$E[D_{stat}] = 0.5892$ $Var[D_{stat}] = 1.89e-03$	$P[D_{stat} \geq 0.57] = 0.7228(73/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4455(45/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.7298	$E[D_{stat}] = 0.5571$ $Var[D_{stat}] = 2.84e-03$	$P[D_{stat} \geq 0.57] = 0.4653(47/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2574(26/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.4653
32 estados	$E[D_{stat}] = 0.5986$ $Var[D_{stat}] = 2.42e-03$	$P[D_{stat} \geq 0.57] = 0.7624(77/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5446(55/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.7603	$E[D_{stat}] = 0.5692$ $Var[D_{stat}] = 2.25e-03$	$P[D_{stat} \geq 0.57] = 0.5149(52/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2772(28/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.5670
48 estados	$E[D_{stat}] = 0.5879$ $Var[D_{stat}] = 2.42e-03$	$P[D_{stat} \geq 0.57] = 0.6733(68/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4851(49/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.8319	$E[D_{stat}] = 0.5734$ $Var[D_{stat}] = 1.68e-03$	$P[D_{stat} \geq 0.57] = 0.5743(58/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.5245

Tabela A.12: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 2 Símbolos, com $T = 2400$.

	Janela de Treinamento: 2400 dias	
8 estados	$E[D_{stat}] = 0.5766$	$P[D_{stat} \geq 0.57] = 0.5050(51/101)$
	$Var[D_{stat}] = 1.65e-03$	$P[\text{Rentabilidade} \geq \text{BH}] = 0.2079(21/101)$
		$P[\text{Perda}] = 0.0000(0/101)$
		90º percentil : 3.2618
12 estados	$E[D_{stat}] = 0.5592$	$P[D_{stat} \geq 0.57] = 0.3762(38/101)$
	$Var[D_{stat}] = 1.40e-03$	$P[\text{Rentabilidade} \geq \text{BH}] = 0.2178(22/101)$
		$P[\text{Perda}] = 0.0000(0/101)$
		90º percentil : 3.0517
16 estados	$E[D_{stat}] = 0.5667$	$P[D_{stat} \geq 0.57] = 0.5743(58/101)$
	$Var[D_{stat}] = 1.78e-03$	$P[\text{Rentabilidade} \geq \text{BH}] = 0.2871(29/101)$
		$P[\text{Perda}] = 0.0000(0/101)$
		90º percentil : 3.4357
24 estados	$E[D_{stat}] = 0.5580$	$P[D_{stat} \geq 0.57] = 0.4950(50/101)$
	$Var[D_{stat}] = 2.11e-03$	$P[\text{Rentabilidade} \geq \text{BH}] = 0.3168(32/101)$
		$P[\text{Perda}] = 0.0099(1/101)$
		90º percentil : 3.4635
32 estados	$E[D_{stat}] = 0.5702$	$P[D_{stat} \geq 0.57] = 0.6238(63/101)$
	$Var[D_{stat}] = 1.53e-03$	$P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$
		$P[\text{Perda}] = 0.0000(0/101)$
		90º percentil : 3.4576
48 estados	$E[D_{stat}] = 0.5734$	$P[D_{stat} \geq 0.57] = 0.6238(63/101)$
	$Var[D_{stat}] = 1.80e-03$	$P[\text{Rentabilidade} \geq \text{BH}] = 0.5050(51/101)$
		$P[\text{Perda}] = 0.0000(0/101)$
		90º percentil : 4.0616

Tabela A.13: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 4 Símbolos, com $T = 600$ e $T = 900$.

	Janela de Treinamento: 600 dias		Janela de Treinamento: 900 dias	
8 estados	$E[D_{stat}] = 0.5257$ $Var[D_{stat}] = 1.54e-03$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 2.7632	$E[D_{stat}] = 0.5801$ $Var[D_{stat}] = 1.72e-03$	$P[D_{stat} \geq 0.57] = 0.7228(73/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2277(23/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.0877
12 estados	$E[D_{stat}] = 0.5182$ $Var[D_{stat}] = 1.89e-03$	$P[D_{stat} \geq 0.57] = 0.1485(15/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2277(23/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 2.6372	$E[D_{stat}] = 0.5741$ $Var[D_{stat}] = 2.46e-03$	$P[D_{stat} \geq 0.57] = 0.5743(58/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3267(33/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.9100
16 estados	$E[D_{stat}] = 0.5269$ $Var[D_{stat}] = 1.75e-03$	$P[D_{stat} \geq 0.57] = 0.1188(12/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3168(32/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 2.8510	$E[D_{stat}] = 0.5882$ $Var[D_{stat}] = 2.66e-03$	$P[D_{stat} \geq 0.57] = 0.6634(67/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3762(38/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.2284
24 estados	$E[D_{stat}] = 0.5356$ $Var[D_{stat}] = 2.13e-03$	$P[D_{stat} \geq 0.57] = 0.1782(18/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3564(36/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 2.8310	$E[D_{stat}] = 0.5704$ $Var[D_{stat}] = 2.54e-03$	$P[D_{stat} \geq 0.57] = 0.4752(48/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3861(39/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.9130
32 estados	$E[D_{stat}] = 0.5351$ $Var[D_{stat}] = 2.72e-03$	$P[D_{stat} \geq 0.57] = 0.2079(21/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4158(42/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 3.3012	$E[D_{stat}] = 0.5709$ $Var[D_{stat}] = 3.36e-03$	$P[D_{stat} \geq 0.57] = 0.4752(48/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.8021
48 estados	$E[D_{stat}] = 0.5416$ $Var[D_{stat}] = 2.88e-03$	$P[D_{stat} \geq 0.57] = 0.2772(28/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3762(38/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 3.1029	$E[D_{stat}] = 0.5639$ $Var[D_{stat}] = 3.51e-03$	$P[D_{stat} \geq 0.57] = 0.5149(52/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3861(39/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.1115

Tabela A.14: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 4 Símbolos, com $T = 1200$ e $T = 2100$.

	Janela de Treinamento: 1200 dias		Janela de Treinamento: 2100 dias	
8 estados	$E[D_{stat}] = 0.5951$ $Var[D_{stat}] = 2.26e-03$	$P[D_{stat} \geq 0.57] = 0.7129(72/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4851(49/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.4137	$E[D_{stat}] = 0.5920$ $Var[D_{stat}] = 1.34e-03$	$P[D_{stat} \geq 0.57] = 0.8020(81/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4752(48/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.8650
12 estados	$E[D_{stat}] = 0.5818$ $Var[D_{stat}] = 2.30e-03$	$P[D_{stat} \geq 0.57] = 0.6535(66/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4158(42/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.6973	$E[D_{stat}] = 0.5677$ $Var[D_{stat}] = 2.74e-03$	$P[D_{stat} \geq 0.57] = 0.5347(54/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.6879
16 estados	$E[D_{stat}] = 0.6036$ $Var[D_{stat}] = 2.13e-03$	$P[D_{stat} \geq 0.57] = 0.8119(82/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5446(55/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.3911	$E[D_{stat}] = 0.5800$ $Var[D_{stat}] = 2.86e-03$	$P[D_{stat} \geq 0.57] = 0.6337(64/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4455(45/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.0700
24 estados	$E[D_{stat}] = 0.5914$ $Var[D_{stat}] = 2.25e-03$	$P[D_{stat} \geq 0.57] = 0.6733(68/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6238(63/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.9367	$E[D_{stat}] = 0.5954$ $Var[D_{stat}] = 2.60e-03$	$P[D_{stat} \geq 0.57] = 0.7030(71/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5149(52/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 5.0034
32 estados	$E[D_{stat}] = 0.5946$ $Var[D_{stat}] = 3.26e-03$	$P[D_{stat} \geq 0.57] = 0.7129(72/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5842(59/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.7280	$E[D_{stat}] = 0.5941$ $Var[D_{stat}] = 3.19e-03$	$P[D_{stat} \geq 0.57] = 0.7228(73/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5446(55/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.2548
48 estados	$E[D_{stat}] = 0.5946$ $Var[D_{stat}] = 3.40e-03$	$P[D_{stat} \geq 0.57] = 0.7129(72/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4356(44/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.3943	$E[D_{stat}] = 0.6120$ $Var[D_{stat}] = 2.25e-03$	$P[D_{stat} \geq 0.57] = 0.8614(87/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.6535(66/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.2887

Tabela A.15: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 4 Símbolos, com $T = 2400$.

	Janela de Treinamento: 2400 dias	
8 estados	$E[D_{stat}] = 0.5925$ $Var[D_{stat}] = 1.53e-03$	$P[D_{stat} \geq 0.57] = 0.7030(71/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5248(53/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.1603
12 estados	$E[D_{stat}] = 0.5503$ $Var[D_{stat}] = 2.59e-03$	$P[D_{stat} \geq 0.57] = 0.3663(37/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2178(22/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.1062
16 estados	$E[D_{stat}] = 0.5748$ $Var[D_{stat}] = 1.82e-03$	$P[D_{stat} \geq 0.57] = 0.5545(56/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4257(43/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.0573
24 estados	$E[D_{stat}] = 0.5781$ $Var[D_{stat}] = 1.88e-03$	$P[D_{stat} \geq 0.57] = 0.6436(65/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5248(53/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.6082
32 estados	$E[D_{stat}] = 0.5727$ $Var[D_{stat}] = 2.51e-03$	$P[D_{stat} \geq 0.57] = 0.6040(61/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 3.9896
48 estados	$E[D_{stat}] = 0.5904$ $Var[D_{stat}] = 2.75e-03$	$P[D_{stat} \geq 0.57] = 0.6634(67/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5842(59/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 4.5332

Tabela A.16: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 6 Símbolos, com $T = 600$ dias e $T = 900$ dias.

	Janela de Treinamento: 600 dias		Janela de Treinamento: 900 dias	
8 estados	$E[D_{stat}] = 0.5434$ $Var[D_{stat}] = 1.76e-03$	$P[D_{stat} \geq 0.57] = 0.1386(14/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2574(26/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 2.6194	$E[D_{stat}] = 0.5437$ $Var[D_{stat}] = 3.49e-03$	$P[D_{stat} \geq 0.57] = 0.4158(42/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1287(13/101)$ $P[\text{Perda}] = 0.0792(8/101)$ 90º percentil : 3.4072
12 estados	$E[D_{stat}] = 0.5554$ $Var[D_{stat}] = 1.58e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$ $P[\text{Perda}] = 0.0000(0/101)$ 90º percentil : 3.0945	$E[D_{stat}] = 0.5491$ $Var[D_{stat}] = 3.41e-03$	$P[D_{stat} \geq 0.57] = 0.3663(37/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 3.7065
16 estados	$E[D_{stat}] = 0.5482$ $Var[D_{stat}] = 2.57e-03$	$P[D_{stat} \geq 0.57] = 0.2871(29/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0693(7/101)$ 90º percentil : 3.2587	$E[D_{stat}] = 0.5548$ $Var[D_{stat}] = 4.20e-03$	$P[D_{stat} \geq 0.57] = 0.4455(45/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4356(44/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.7583
24 estados	$E[D_{stat}] = 0.5475$ $Var[D_{stat}] = 2.55e-03$	$P[D_{stat} \geq 0.57] = 0.2970(30/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 3.2028	$E[D_{stat}] = 0.5476$ $Var[D_{stat}] = 3.86e-03$	$P[D_{stat} \geq 0.57] = 0.3762(38/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4257(43/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.8450
32 estados	$E[D_{stat}] = 0.5530$ $Var[D_{stat}] = 2.73e-03$	$P[D_{stat} \geq 0.57] = 0.3267(33/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 3.1579	$E[D_{stat}] = 0.5427$ $Var[D_{stat}] = 4.96e-03$	$P[D_{stat} \geq 0.57] = 0.4356(44/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3960(40/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.7653
48 estados	$E[D_{stat}] = 0.5450$ $Var[D_{stat}] = 3.55e-03$	$P[D_{stat} \geq 0.57] = 0.2673(27/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4059(41/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 3.5946	$E[D_{stat}] = 0.5481$ $Var[D_{stat}] = 5.30e-03$	$P[D_{stat} \geq 0.57] = 0.4455(45/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4356(44/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 4.7700

Tabela A.17: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 6 Símbolos, com $T = 1200$ dias e $T = 2100$ dias.

	Janela de Treinamento: 1200 dias		Janela de Treinamento: 2100 dias	
8 estados	$E[D_{stat}] = 0.5305$ $Var[D_{stat}] = 5.26e-03$	$P[D_{stat} \geq 0.57] = 0.4059(41/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0693(7/101)$ 90º percentil : 3.3701	$E[D_{stat}] = 0.5466$ $Var[D_{stat}] = 7.40e-03$	$P[D_{stat} \geq 0.57] = 0.5050(51/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4059(41/101)$ $P[\text{Perda}] = 0.0891(9/101)$ 90º percentil : 3.9218
12 estados	$E[D_{stat}] = 0.5300$ $Var[D_{stat}] = 5.34e-03$	$P[D_{stat} \geq 0.57] = 0.3465(35/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2079(21/101)$ $P[\text{Perda}] = 0.0990(10/101)$ 90º percentil : 3.2851	$E[D_{stat}] = 0.5326$ $Var[D_{stat}] = 4.10e-03$	$P[D_{stat} \geq 0.57] = 0.3069(31/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.1881(19/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.3800
16 estados	$E[D_{stat}] = 0.5677$ $Var[D_{stat}] = 5.17e-03$	$P[D_{stat} \geq 0.57] = 0.5743(58/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5248(53/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 5.1579	$E[D_{stat}] = 0.5633$ $Var[D_{stat}] = 5.12e-03$	$P[D_{stat} \geq 0.57] = 0.5545(56/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.7626
24 estados	$E[D_{stat}] = 0.5649$ $Var[D_{stat}] = 4.84e-03$	$P[D_{stat} \geq 0.57] = 0.5347(54/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0099(1/101)$ 90º percentil : 4.6080	$E[D_{stat}] = 0.5477$ $Var[D_{stat}] = 5.58e-03$	$P[D_{stat} \geq 0.57] = 0.4455(45/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4158(42/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 4.9667
32 estados	$E[D_{stat}] = 0.5677$ $Var[D_{stat}] = 4.51e-03$	$P[D_{stat} \geq 0.57] = 0.5149(52/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3663(37/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.4814	$E[D_{stat}] = 0.5689$ $Var[D_{stat}] = 5.37e-03$	$P[D_{stat} \geq 0.57] = 0.6139(62/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5446(55/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.5425
48 estados	$E[D_{stat}] = 0.5791$ $Var[D_{stat}] = 4.13e-03$	$P[D_{stat} \geq 0.57] = 0.6238(63/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4554(46/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 4.8411	$E[D_{stat}] = 0.5719$ $Var[D_{stat}] = 6.15e-03$	$P[D_{stat} \geq 0.57] = 0.6238(63/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.5149(52/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 4.7701

Tabela A.18: Resultados Experimentais para uma janela de previsão de 30 dias: HMM de 6 Símbolos, com $T = 2400$ dias.

	Janela de Treinamento: 2400 dias	
8 estados	$E[D_{stat}] = 0.5378$ $Var[D_{stat}] = 6.74e-03$	$P[D_{stat} \geq 0.57] = 0.4356(44/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3267(33/101)$ $P[\text{Perda}] = 0.0594(6/101)$ 90º percentil : 3.3670
12 estados	$E[D_{stat}] = 0.5345$ $Var[D_{stat}] = 3.73e-03$	$P[D_{stat} \geq 0.57] = 0.3069(31/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.2574(26/101)$ $P[\text{Perda}] = 0.0396(4/101)$ 90º percentil : 3.3190
16 estados	$E[D_{stat}] = 0.5581$ $Var[D_{stat}] = 3.69e-03$	$P[D_{stat} \geq 0.57] = 0.4455(45/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3465(35/101)$ $P[\text{Perda}] = 0.0198(2/101)$ 90º percentil : 4.0202
24 estados	$E[D_{stat}] = 0.5487$ $Var[D_{stat}] = 3.77e-03$	$P[D_{stat} \geq 0.57] = 0.4257(43/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3762(38/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.2763
32 estados	$E[D_{stat}] = 0.5607$ $Var[D_{stat}] = 3.26e-03$	$P[D_{stat} \geq 0.57] = 0.5347(54/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.3564(36/101)$ $P[\text{Perda}] = 0.0297(3/101)$ 90º percentil : 4.1745
48 estados	$E[D_{stat}] = 0.5691$ $Var[D_{stat}] = 4.60e-03$	$P[D_{stat} \geq 0.57] = 0.5347(54/101)$ $P[\text{Rentabilidade} \geq \text{BH}] = 0.4059(41/101)$ $P[\text{Perda}] = 0.0495(5/101)$ 90º percentil : 4.4778

Referências Bibliográficas

[1] HULL, JOHN C., *Options, Futures and Other Derivatives (7th Edition)*. Prentice Hall, May 2008.

[2] RABINER, L.R., "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, v. 77, n. 2, pp. 257–286, Feb 1989.

[3] CAPPÉ, O., "Ten years of HMMs", Mar 2001, http://www.tsi.enst.fr/∼cappe/docs/hmmbib.html acessado em 18/02/2009.

[4] MUNTZ, R. AND DE SOUZA E SILVA, E. A., *Modeling with Markov Chains*. Notas de Aula, 2008.

[5] DRAKE, A. W., *Fundamentals of Applied Probability Theory - Chapter 5*. Mcgraw-Hill College, 1967.

[6] BAUM, L. E., PETRIE, T., SOULES, G. AND WEISS, N., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", *The Annals of Mathematical Statistics*, v. 41, n. 1, pp. 164–171, 1970.

[7] DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 39, n. 1, pp. 1–38, 1977.

[8] BAUM, L. E. AND SELL, G. R., "Growth functions for transformations on manifolds", *Pacific Journal of Mathematics*, v. 27, n. 2, pp. 211–227, 1968.

[9] BAUM, L. E. AND EAGON, J. A., "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology", *Bulletin of the American Mathematical Society*, v. 73, n. 3, pp. 360–363, 1967.

[10] BILMES, J., *A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*, Tech. rep., University of Berkeley, 1997.

[11] GRAPS, A., "An Introduction to Wavelets", *IEEE Computational Science and Engineering*, v. 02, n. 2, pp. 50–61, 1995.

[12] VIDAKOVIC, B. AND MULLER, P., "Wavelets for Kids", http://www2.isye.gatech.edu/∼brani/wp/kidsA.pdf acessado em 18/02/2009.

[13] "Matlab", http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/ acessado em 18/02/2009.

[14] MALLAT, S. G., "A theory for multiresolution signal decomposition: the wavelet representation", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 11, n. 7, pp. 674–693, Jul 1989.

[15] MEYER, Y., "Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs", *Astérisque*, , n. 145-146, pp. 4, 209–223, 1987, Séminaire Bourbaki, Vol. 1985/86.

[16] VALENS, C., "A Really Friendly Guide to Wavelets", 1999, http://pagesperso-orange.fr/polyvalens/clemens/download/arfgtw.pdf acessado em 18/02/2009.

[17] PINDYCK, R. S., "The dynamics of commodity spot and futures markets: a primer", *The Energy Journal*, v. 22, n. 3, 2001.

[18] YE, M., ZYREN, J. AND SHORE, J., "Forecasting short-run crude oil price using high- and low-inventory variables", *Energy Policy*, v. 34, n. 17, pp. 2736–2743, November 2006.

[19] YE, M., ZYREN, J. AND SHORE, J., “A monthly crude oil spot price forecasting model using relative inventories”, *International Journal of Forecasting*, v. 21, n. 3, pp. 491–501, 2005.

[20] YE, M., ZYREN, J. AND SHORE, J., “Forecasting crude oil spot price using OECD petroleum inventory levels”, *Journal International Advances in Economic Research*, v. 8, n. 4, pp. 324–333, November 2002.

[21] “Energy Information Administration”, http://www.eia.doe.gov acessado em 18/02/2009.

[22] MERINO, A. AND ORTIZ, Á., “Explaining the so-called "price premium"in oil markets”, *OPEC Review*, v. 29, n. 2, pp. 133–152, Junho 2005.

[23] BROCKWELL, P. J. AND DAVIS, R. A., *Introduction to time series and forecasting*. Springer Texts in Statistics, 2002.

[24] PINDYCK, R. S., RUBINFELD, D. L., *Econometria - Modelos e Previsões*. Elsevier, 2004.

[25] SHAMBORA, W. E. AND ROSSITER, R., “Are there exploitable inefficiencies in the futures market for oil?” *Energy Economics*, v. 29, n. 1, pp. 18–27, January 2007.

[26] XIE, W., YU, L., XU, S. AND WANG, S., “A New Method for Crude Oil Price Forecasting Based on Support Vector Machines”. In: *International Conference on Computational Science (4)*, pp. 444–451, 2006.

[27] MORANA, C., “A semiparametric approach to short-term oil price forecasting”, *Energy Economics*, v. 23, n. 3, pp. 325–338, 2001.

[28] HAYKIN, S., *Redes Neurais: Princípios e prática*. 2a ed. Bookman, 2001.

[29] WANG, S., YU, L. AND LAI, K. K., “A Novel Hybrid AI System Framework for Crude Oil Price Forecasting”. In: *CASDMKM*, pp. 233–242, 2004.

[30] SHI, S. AND WEIGEND, A. S., "Taking time seriously: hidden Markov experts applied to financial engineering", *Computational Intelligence for Financial Engineering (CIFEr), 1997., Proceedings of the IEEE/IAFE 1997*, v. , n. , pp. 244–252, 23-25 Mar 1997.

[31] ZHANG, Y., *Prediction of financial time series with hidden markov models*, Master's Thesis, B.Eng. Shandong University - School of Computing Science, Maio 2004.

[32] DE SOUZA E SILVA, E., LEÃO, R. M. M. AND RATTON, D., "An integrated modeling environment for computer systems and networks", *ACM SIGMETRICS Performance Evaluation Review*, v. 36, pp. 1, 2009.

[33] SILVEIRA FILHO, F. AND DE SOUZA E SILVA, E., "Um método de previsão de perdas de pacotes com aplicações à transmissão de mídia contínua". In: *Anais do XXIV Simpósio Brasileiro de Redes de Computadores (SBRC'2006)*, v. II, pp. 879–894, Junho 2006.

[34] SILVEIRA FILHO, F. AND DE SOUZA E SILVA, E., "Modeling the short-term dynamics of packet losses". In: *Performance Evaluation Review*, v. 34, n. 3, pp. 27–29, December 2006.

[35] DE SOUZA E SILVA, E. A. AND GAIL, H. R., "Calculating availability and performability measures of repairable computer systems using randomization", *J. ACM*, v. 36, n. 1, pp. 171–193, 1989.

[36] SILVA, A. P. C., *Métodos computacionais para modelos markovianos com recompensas*, Ph.D. Thesis, Universidade Federal do Rio de Janeiro - COPPE - Programa de Engenharia de Sistemas e Computação, Outubro 2006.

[37] DE SOUZA E SILVA, E. A. AND GAIL, R., "An algorithm to calculate transient distributions of cummulative rate and impulse based reward". In: *Communications in Statistics - Stochastic Models*, v. 14, n. 3, pp. 509–536, 1998.

[38] SILVA, A. P. C., *Métodos de solução para modelos markovianos com recompensa*, Master's Thesis, Universidade Federal do Rio de Janeiro COPPE - Programa de Engenharia de Sistemas e Computação, Outubro 2001.

[39] "EMpht", http://home.imf.au.dk/asmus/dl/EMusersguide.ps acessado em 18/02/2009.

[40] DE SOUZA E SILVA, E. A. AND GAIL, H. R., "The Uniformization Method in Performability Analysis", In: *Performability Modeling: Techniques and Tools*.

[41] "Phase-Type distributions", http://www.cs.wm.edu/∼riska/PhD-thesis-html/node19.html acessado em 18/02/2009.

[42] "Brief tutorial on phase-type distributions", http://www.cs.cmu.edu/∼osogami/thesis/html/node38.html acessado em 18/02/2009.

[43] JENSEN, A., "Markov chains as an aid in the study of Markov processes", *Skandinavsk Aktuarietidskrift*, v. 36, pp. 87–91, 1953.

[44] DE SOUZA E SILVA, E. A. AND GAIL, H. R., "Computational Probability", chap. Transient Solutions for Markov Chains, Kluwer Academic Publishers, 2000.