

**UNIVERSIDADE FEDERAL DE PELOTAS**  
Centro de Desenvolvimento Tecnológico  
Programa de Pós-Graduação em Computação



Dissertação

**Aprendizado de Máquina na Detecção de Tendências em Bolsa De Valores**

**Marcelo Rubens Giesel**

Pelotas, 2015

**Marcelo Rubens Giesel**

**Aprendizado de Máquina na Detecção de Tendências em Bolsa De Valores**

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação

Orientador: Prof. Dr. Ricardo Matsumura Araujo  
Coorientador: Prof. Dr. Thyago Carvalho Marques

Pelotas, 2015

Universidade Federal de Pelotas / Sistema de Bibliotecas  
Catalogação na Publicação

G455a Giesel, Marcelo Rubens

Aprendizado de máquina na detecção de tendências  
em bolsa de valores / Marcelo Rubens Giesel ; Ricardo  
Matsumura Araujo, orientador ; Thyago Carvalho Marques,  
coorientador. — Pelotas, 2015.

105 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação  
em Computação, Centro de Desenvolvimento Tecnológico,  
Universidade Federal de Pelotas, 2015.

1. Aprendizado de máquina. 2. Inteligência artificial. 3.  
Bolsa de valores. I. Araujo, Ricardo Matsumura, orient. II.  
Marques, Thyago Carvalho, coorient. III. Título.

CDD : 005





UNIVERSIDADE FEDERAL DE PELOTAS  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
CENTRO DE DESENVOLVIMENTO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
COMPUTAÇÃO



PPGC  
Programa de Pós-Graduação  
em Ciência da Computação  
Universidade Federal de Pelotas

ATA DE APRECIÇÃO SOBRE A DEFESA DE DISSERTAÇÃO DE MESTRADO

NOME DO ESTUDANTE		MATRÍCULA
Marcelo Rubens Giesel		13102745
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO		MESTRADO
MEMBROS DA BANCA EXAMINADORA	TÍTULO	ASSINATURA
Mario Duarte Canever	Doutor	
Marilton Sanchotene Aguiar	Doutor	
Paulo Roberto Ferreira Jr	Doutor	

APRECIÇÃO SOBRE A DISSERTAÇÃO

NÃO SIGILOSA

Em 01 de setembro de 2015 os membros acima nomeados para a defesa da Dissertação do aluno **ALUNO**, matriculado no Programa de Pós-Graduação em Computação, consideraram aprovado, estabelecendo o título definitivo da Dissertação como sendo " o mesmo ", e estabelecendo um prazo máximo de 30 dias para as correções e entrega da versão definitiva.

DADOS PESSOAIS DOS MEMBROS DA BANCA EXAMINADORA

NOME COMPLETO	CPF	ANO NASCIMENTO	TITULAÇÃO		
			Área	Local	Ano
Mario Duarte Canever	62351206991		Administração	Wageningen University, Holanda	2006
Marilton Sanchotene Aguiar	78878659053	1975	Computação	UFRGS	2004
Paulo Roberto Ferreira Jr	61699098034	1976	Computação	UFRGS	2008

1ª Via – Coordenador do Curso; 2ª Via – Orientador; 3ª Via – PRPPG.

**DISTRIBUIÇÃO A CARGO DA COORDENAÇÃO DO PROGRAMA.**



**Ao seu Anselmo e dona Justina.**

## AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – pela concessão da bolsa para esta pesquisa.

Ao Programa de Pós-Graduação em Computação pela oportunidade e pelo acolhimento.

Aos professores, técnicos administrativos e demais funcionários do programa. Em especial ao Ricardo, Marilton, Paulo e Martha, que estiveram mais próximos de minha trajetória. Obrigado pelo suporte, apoio e direções.

Ao Ricardo, meu orientador, pela paciência, cuidado, orientação e parceria. Ao Thyago, meu co-orientador, pelo auxílio em um momento muito importante.

À minha família, por ter me incentivado todo o tempo, principalmente quando a pesquisa se tornava exaustiva.

À dona Ilva, pelo acolhimento e cuidados, sempre feitos com muito carinho.

Ao Edegar e Igor, pela amizade e companheirismo.

À minha namorada, Cibele, por ser minha parceira em tudo e por ser *expert* em ouvir desabafos.

Aos colegas de mestrado que compartilharam conhecimento e vivência ao longo do curso.

Aos amigos de graduação que souberam partilhar dos questionamentos que surgiam vez ou outra ao longo da pesquisa.

À proteção e provisão divinas.

## RESUMO

GIESEL, Marcelo Rubens. **Aprendizado de Máquina na Detecção de Tendências em Bolsa De Valores**. 2015. 105 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2015.

Este trabalho tem por objetivo verificar a viabilidade de utilização do algoritmo de detecção de tendências de NIKOLOV (2012) no mercado de ações brasileiro. São realizadas identificações das diferenças entre os ambientes da aplicação original do algoritmo e do ambiente da bolsa de valores, e sugeridas quatro adaptações ao algoritmo. Em uma das adaptações verificou-se resultado positivo, retorno financeiro acima do *benchmark* utilizado, que foi a técnica de operação de cruzamento de médias móveis. O trabalho é finalizado com sugestões de melhorias e de outros tipos de testes.

**Palavras-chave:** Aprendizado de máquina, Inteligência artificial, Bolsa de valores.

## **ABSTRACT**

GIESEL, Marcelo Rubens. **Learning Machine in the Detection of Trends in the Stock Market**. 2015. 105 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2015.

This work aims to verify the feasibility of using the Nikolov's trending detection algorithm (NIKOLOV, 2012) in the Brazilian stock market. We identify the differences between the environments of the original application of the algorithm and the stock exchange environment, and suggest four adaptations to the algorithm. In one adaptation we have got a positive result, a financial return higher than the benchmark used, which was the operation technique of moving averages intersection. The work ends with suggestions for improvements and other types of tests.

**Keywords:** Learning Machine, Artificial Intelligence, Stock Market.



## LISTA DE FIGURAS

1	Fonte latente $t_{j*}$ , média que minimiza a distância entre o sinal de referência $r$ e a observação $s$ . Os pontilhados são as fontes latentes $t_j$ . Retirado de NIKOLOV (2012) . . . . .	20
2	Comparando um pedaço do comprido sinal de referência com uma curta observação. Retirado de NIKOLOV (2012) . . . . .	22
3	Definição de tendência da Abordagem 1. . . . .	32
4	Cruzamento de médias móveis de períodos 9 e 21 . . . . .	34
5	(a) Distribuição dos resultados de mil execuções de sete configurações da Abordagem 1, na fase de validação. (b) Ampliação do mesmo gráfico. . . . .	39
6	(a) Distribuição dos resultados de mil execuções de sete configurações da Abordagem 1, na fase de testes. (b) Ampliação do mesmo gráfico. . . . .	39
7	Boxplots dos retornos, separados por variável, da carteira teórica com os 10 ativos para a Abordagem 2. . . . .	42
8	Boxplots dos retornos, separados por variável, da carteira teórica com os 10 ativos para a Abordagem 3. . . . .	43
9	Boxplots dos retornos, separados por variável, da carteira teórica com os 10 ativos para a Abordagem 4. . . . .	44

## LISTA DE TABELAS

1	Comparação das características dos dois ambientes. . . . .	27
2	Ativos selecionados para o trabalho, referentes às maiores empresas de cada setor da economia. . . . .	28
3	Retornos das Abordagens 2, 3 e 4 (A2, A3 e A4), do cruzamento das médias móveis (MM), média e desvio padrão de cada estratégia na fase de teste. . . . .	41

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>15</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>17</b>
<b>3.1</b>	<b>Algoritmo de Nikolov</b>	<b>17</b>
3.1.1	A abordagem	17
3.1.2	Motivação	18
3.1.3	Modelo Estocástico	19
3.1.4	Detecção	19
<b>4</b>	<b>METODOLOGIA</b>	<b>23</b>
<b>4.1</b>	<b>Objetivos</b>	<b>23</b>
<b>4.2</b>	<b>Adaptação do algoritmo</b>	<b>24</b>
4.2.1	Características dos dois ambientes	24
<b>4.3</b>	<b>Seleção das ações e do período histórico</b>	<b>27</b>
<b>4.4</b>	<b>Obtenção e ajustes das séries temporais</b>	<b>28</b>
<b>4.5</b>	<b>Divisão do histórico em 3 conjuntos</b>	<b>29</b>
<b>5</b>	<b>ABORDAGENS</b>	<b>31</b>
<b>5.1</b>	<b>Abordagem 1</b>	<b>31</b>
<b>5.2</b>	<b>Abordagem 2</b>	<b>33</b>
<b>5.3</b>	<b>Abordagem 3</b>	<b>35</b>
<b>5.4</b>	<b>Abordagem 4</b>	<b>37</b>
<b>6</b>	<b>RESULTADOS</b>	<b>38</b>
<b>6.1</b>	<b>Abordagem 1</b>	<b>38</b>
<b>6.2</b>	<b>Abordagens 2, 3 e 4</b>	<b>40</b>
6.2.1	Validação	40
6.2.2	Teste	41
<b>7</b>	<b>DISCUSSÕES E TRABALHOS FUTUROS</b>	<b>46</b>
	<b>REFERÊNCIAS</b>	<b>49</b>
	<b>APÊNDICE A TABELAS DA VALIDAÇÃO DA ABORDAGEM 2</b>	<b>51</b>
<b>A.1</b>	<b>gama</b>	<b>51</b>
<b>A.2</b>	<b>theta</b>	<b>52</b>
<b>A.3</b>	<b>detectionsLimit</b>	<b>53</b>
<b>A.4</b>	<b>NSMOOTH</b>	<b>54</b>

A.5	NREF	55
A.6	NOBS	56
<b>APÊNDICE B BOXPLOTS DA VALIDAÇÃO DA ABORDAGEM 2</b>		<b>57</b>
B.1	gama	57
B.2	theta	59
B.3	detectionsLimit	61
B.4	NSMOOTH	63
B.5	NREF	65
B.6	NOBS	67
<b>APÊNDICE C TABELAS DA VALIDAÇÃO DA ABORDAGEM 3</b>		<b>69</b>
C.1	gama	69
C.2	theta	70
C.3	detectionsLimit	71
C.4	NSMOOTH	72
C.5	NREF	73
C.6	NOBS	74
<b>APÊNDICE D BOXPLOTS DA VALIDAÇÃO DA ABORDAGEM 3</b>		<b>75</b>
D.1	gama	75
D.2	theta	77
D.3	detectionsLimit	79
D.4	NSMOOTH	81
D.5	NREF	83
D.6	NOBS	85
<b>APÊNDICE E TABELAS DA VALIDAÇÃO DA ABORDAGEM 4</b>		<b>87</b>
E.1	gama	87
E.2	theta	88
E.3	detectionsLimit	89
E.4	NSMOOTH	90
E.5	NREF	91
E.6	NOBS	92
<b>APÊNDICE F BOXPLOTS DA VALIDAÇÃO DA ABORDAGEM 4</b>		<b>93</b>
F.1	gama	93
F.2	theta	95
F.3	detectionsLimit	97
F.4	NSMOOTH	99
F.5	NREF	101
F.6	NOBS	103

# 1 INTRODUÇÃO

Uma empresa, ao abrir seu capital em uma bolsa de valores, divide sua sociedade em várias frações de igual valor, e cada fração é chamada de ação. O preço da ação tende a refletir o que o mercado espera quanto ao desempenho futuro da empresa que a emitiu (JÚNIOR; RIGO; CHEROBIM, 2002). Ele é alterado frequentemente devido ao processo de compra e venda, e estas alterações ocorrem por diversos motivos. Os operadores baseiam-se nos balanços da empresa para fazer esta estimativa, utilizando dados como lucro líquido, dívida, financiamentos, e outros. Uma empresa que, fisicamente, está bem estruturada e tende a crescer tende a ter elevado também o valor de sua ação. A análise da empresa inclui fatores econômicos, como alteração de taxa de juros, incentivos fiscais e auxílio governamental. Crises internacionais, alianças comerciais entre governos, eleições, rumores e pura especulação também exercem influência no preço das ações. Estas alterações de preço oferecem oportunidades de negócios. Com a finalidade de obtenção de lucro, operadores de vários tipos fazem uso de estratégias variadas de aplicações para longo, médio ou curto prazo.

Segundo BRITO (2005), operações de compra e venda de ações aproveitando as flutuações dos preços estão entre os tópicos básicos das operações no mercado acionário. Quando gerencia-se uma carteira de ações estas flutuações são normalmente analisadas para longos períodos, compreendendo semanas, meses e anos. Mas também estão incluídas as operações que aproveitam as oscilações bruscas de mercado em um curto espaço de tempo, sejam estas oscilações com duração de dias, horas ou minutos. Uma característica que independe do período analisado é o fato de se ter uma oscilação de preços por um determinado tempo.

Para um operador sozinho é praticamente impossível analisar todos os fatores que influenciam a variação do preço da ação e tomar proveito disto. As variáveis são muitas e difíceis de serem analisadas. Um estudo cuidadoso de todas é um processo oneroso, podendo não ser viável financeiramente e o resultado pode sair tardiamente, quando não há possibilidade de aproveitar o movimento. Pode-se escolher, então, um sub-conjunto de variáveis para serem analisadas, a fim de diminuir o tempo de análise. A escolha deste conjunto pode ser útil, mas também pode resultar em indi-

cações contraditórias. Facilitando este processo, PINHEIRO (2005) afirma que não é necessário todas estas análises, pois todo fator que afeta o preço de uma ação já está descontado no próprio preço. Os preços sofrem alterações e, para o analista, não importa o motivo da movimentação (crise europeia, parceria com a China, alta dos juros, etc), basta analisar o gráfico da ação, pois este já traduz o comportamento do mercado.

Uma das abordagens mais utilizadas a tomar proveito desta afirmação é a análise técnica, que baseia-se na ideia de que preços se movem de acordo com padrões não estacionários, porém identificáveis (ADRIÃO, 2009). Seguindo um modelo, um padrão na série temporal dos preços, os operadores que utilizam esta abordagem decidem se realizam operações de compra ou venda acreditando em um movimento. Por exemplo, se acreditam que haverá um aumento no preço dos ativos eles compram para depois poder vender quando o preço estiver mais elevado, lucrando com a diferença dos preços.

Dentro da análise técnica, existem dezenas de identificadores de tendências, como médias móveis, canais de Donchian e HiLo Activator (MALHEIROS, 2011). As médias móveis, em especial, são um dos indicadores mais populares, servindo também como base para quase todos os outros indicadores (NORONHA, 2009). O objetivo delas é suavizar o gráfico facilitando a visualização de tendências, indicando seu início e término. Ao eliminar ruídos facilita ainda a identificação da direção do movimento de massa (ELDER, 2004). Mas apesar de ser de fácil uso e amplamente utilizada esta técnica é atrasada, no sentido de demorar a realizar as indicações.

Além disso, ainda são verificados muitos falsos positivos, situações onde o indicador técnico sinaliza a existência de uma tendência mas que acaba não se confirmando. JOHANN; VIDAL; TÁRTARO (2010), por exemplo, obtiveram um percentual de operações lucrativas de 46% com a estratégia mais rentável de médias móveis. Essa baixa taxa de acerto faz com que os lucros obtidos sejam consumidos pelas operações negativas, diminuindo a rentabilidade final apresentada pela estratégia. Isto motiva a busca por ferramentas mais precisas.

Este trabalho tem como objetivo levantar possibilidades de melhoramento dos resultados apresentados pelas técnicas clássicas em operações realizadas no mercado financeiro brasileiro, especificamente na Bolsa de Valores, Mercadorias e Futuros de São Paulo – Bovespa<sup>1</sup>. Para isso, é feito um levantamento dos estudos relacionados à detecção de tendências em séries temporais. Parte deste levantamento é mostrado no Capítulo 2, que apresenta alguns dos trabalhos lidos juntamente com comentários a respeito da abordagem tomada e resultados obtidos nos mesmos. No Capítulo 3 é apresentada em detalhes a técnica escolhida para ser aplicada ao cenário deste trabalho. NIKOLOV (2012) elaborou um algoritmo de aprendizagem de máquina sem

---

<sup>1</sup><http://www.bmfbovespa.com.br/>



modelo para fazer inferências na série temporal, a fim de poder separar “eventos” de “não-eventos” dada uma quantidade suficiente de exemplos históricos. O Capítulo 4 apresenta as diferenças entre o ambiente original onde a técnica foi aplicada e o ambiente que utilizamos neste trabalho, as adaptações necessárias ao algoritmo, dificuldades encontradas e o processo de testes. O Capítulo 5 detalha as diferentes abordagens de adaptação do algoritmo, incluindo justificativas. O Capítulo 6 apresenta as conclusões sobre a aplicabilidade do algoritmo de Nikolov ao cenário proposto, a detecção e tendências em ações no mercado financeiro brasileiro. No Capítulo 7, por fim, discorreremos sobre todo o processo e mostramos possibilidades de trabalhos futuros.

## 2 TRABALHOS RELACIONADOS

Para obter melhores resultados nas operações, foi necessário fazer um levantamento sobre os trabalhos que realizam detecção/predição em séries temporais. Consideramos que o mercado de ações segue o modelo técnico baseado em tendências puramente gráficas (ARTHUR et al., 1996). Os trabalhos aqui apresentados procuram fazer as análises utilizando diferentes abordagens, que vão desde análise técnica tradicional até o uso de redes neurais. Eles serviram como base para o entendimento das formas de detecção de valores e auxiliaram na escolha da estratégia.

Alguns dos estudos apresentados neste capítulo são baseados no Jogo da Minoria (CHALLET; ZHANG, 1997). Este jogo, basicamente, é composto por vários agentes, cada um com duas estratégias, que devem escolher um grupo dentre dois a cada rodada, e os agentes que escolheram o grupo que ficou com o menor número de agentes recebe uma bonificação. Assim, os agentes têm por objetivo escolher o grupo da minoria a cada rodada baseando-se em suas estratégias. O jogo contém agentes heterogêneos com racionalidade limitada agindo de forma indutiva, e apresenta características interessantes para estudo em várias áreas onde há competição por recursos, inclusive mercado financeiro.

O trabalho de JOHANN; VIDAL; TÁRTARO (2010) apresenta uma análise da detecção de tendências na bolsa de valores brasileira utilizando métodos tradicionais para este tipo de operação. O estudo utiliza a técnica de cruzamento de médias móveis, um dos indicadores técnicos mais populares. A escolha desta técnica se deve ao fato de ela facilitar a visualização da tendência no gráfico e por servir de base para a maioria dos indicadores técnicos. Por ser muito simples, apresenta alguns pontos negativos, como atraso na detecção e uma porcentagem não muito alta, para os estudos do tipo, de operações lucrativas, mas ainda assim seu uso a longo prazo se mostra vantajoso. O trabalho obteve resultados positivos nas duas estratégias de cruzamento de médias móveis, ou seja, um retorno lucrativo sobre o capital investido. O estudo levou em consideração um período de quatro anos onde o mercado não apresentou tendências definidas, proporcionando mais operações de compra e venda e uma taxa de 46% de operações positivas na estratégia mais rentável. Serve como um *benchmark* para

obtenção de resultados melhores.

Já FERREIRA et al. (2003) fizeram uma análise do comportamento de séries temporais em simulações do mercado financeiro. Implementaram um modelo da dinâmica de mercado baseado em agentes que apresentava um comportamento complexo. O modelo permitia duas categorias de agentes, produtores – que usam o mercado para troca de mercadorias – e especuladores – que visam obtenção de lucro a partir da flutuação de preços –, além de permitir que especuladores optem por não realizar operações. Estas duas características fizeram com que o modelo ficasse com comportamento próximo ao real.

O trabalho de CAMARGO (2006) envolvia estudar a troca de informações locais em um Jogo da Minoria numa rede complexa. A ideia através dessa aproximação é considerar que alguns dos agentes que jogam o jogo imitam alguns de seus vizinhos que eles acreditam ser mais bem informados. Assim um agente, a cada rodada, pode escolher utilizar ou uma de suas estratégias ou a melhor estratégia do seu vizinho mais bem informado. Quando a imitação se expande pela rede, pode ocorrer que a estratégia de um agente muito bem informado seja imitada por outros agentes, que por sua vez são imitados por outros agentes, propagando a estratégia pela rede e gerando assim o chamado efeito manada. Este efeito de manada é verificado com frequência em tendências.

Já RIBEIRO (2005) estudou o comportamento do Jogo da Minoria com dois grupos de agentes, um cujos agentes não evoluíam suas estratégias, estáticas, e outro cujos agentes utilizavam algoritmo de aprendizado para alterar suas estratégias. Verificou-se que os agentes que faziam uso de aprendizado para alterar suas estratégias conseguiam, em determinadas condições, identificar padrões nas escolhas dos agentes com estratégias estáticas. A eficiência na identificação no geral depende do algoritmo utilizado, mas quando a quantidade de agentes que alteram as estratégias é pequena em relação à quantidade dos agentes que não alteram suas estratégias, aqueles conseguem bons resultados independentemente do algoritmo utilizado.

GRILLI; SFRECOLA (2009) utilizou redes neurais para prever preço das ações em um modelo de Jogo da Minoria. Concluiu que os resultados de predição são interessantes para séries temporais pseudoaleatórias, e não muito boas para sequências aleatórias. Afirmou que a hipótese da caminhada aleatória é frequentemente violada por séries temporais financeiras, e que elas apresentam a existência de anomalias a respeito da hipótese do mercado eficiente, sugerindo assim que séries temporais financeiras são pseudo-aleatórias e que existe possibilidade do uso de algoritmos de predição. Com a mesma afirmação de que a hipótese da caminhada aleatória apresenta ineficiências, ELY (2011) verificou que existe previsibilidade no mercado acionário brasileiro. Seu estudo focou na divisão das ações por setor econômico e por tamanho da empresa.

## 3 FUNDAMENTAÇÃO TEÓRICA

### 3.1 Algoritmo de Nikolov

Recentemente, um novo algoritmo para detecção de tendências foi proposto. Sabendo da importância da detecção, classificação e predição de eventos na extração de informação útil de dados temporais, NIKOLOV (2012) estudou o problema de predição em um sistema complexo utilizando uma grande quantidade de dados. Focou especificamente na classificação binária de séries temporais, onde procurava separar “eventos” de “não-eventos” dada uma quantidade suficiente de exemplos históricos. Para isso elaborou um algoritmo de aprendizagem de máquina sem modelo para fazer inferências na série temporal. Na classificação, o sinal observado era comparado com dois conjuntos de sinais de referências, um contendo exemplos positivos de eventos e outro contendo exemplos negativos, de “não-eventos”. Assumindo que todos os sinais observados são uma forma ruidosa de um sinal ou do conjunto positivo ou do conjunto negativo, o sinal observado era comparado com os sinais dos dois conjuntos e era calculada a probabilidade de este sinal pertencer a um conjunto ou a outro. A comparação era feita calculando-se a distância do sinal observado para os sinais das classes de referência, permitindo assim que o sinal fosse classificado de uma forma não-paramétrica diretamente a partir dos dados em si, sem a necessidade de especificar uma estrutura de modelo.

As seções seguintes descrevem com mais detalhes o trabalho de Nikolov. Iniciando com explicação da abordagem do problema, passando pela motivação e o desenvolvimento da técnica.

#### 3.1.1 A abordagem

Simples, modelos paramétricos se provaram ineficientes na modelagem de muitos sistemas complexos do mundo real. Para solucionar isto, foi proposto um *framework* não-paramétrico para realização de inferências em séries temporais. Neste modelo, foi defendida a ideia da existência de um conjunto de séries temporais de *fonte latente*, ou sinais, cada uma correspondendo a um protótipo de um evento de um certo tipo,

e que cada série temporal é uma observação ruidosa de uma das séries temporais latentes.

No caso de classificação, um sinal observado é comparado a dois conjuntos de sinais de referências – um formado por exemplos positivos e outro de exemplos negativos. É pressuposto que o sinal observado pertence à classe positiva se foi gerado pela mesma fonte latente de um dos exemplos positivos. E de forma análoga, um sinal pertence à classe negativa se foi gerado pela mesma fonte latente de um dos exemplos negativos. Para fazer a classificação, é calculada a classe de probabilidades condicionadas na observação. Neste modelo, isto envolve uma computação simples – para ver o quanto uma observação pertence a uma certa classe, são calculadas as distâncias da observação para os sinais de referência daquela classe. Isto permite inferir a classe de uma forma não-paramétrica diretamente dos dados sem especificar qualquer estrutura de modelo.

### 3.1.2 Motivação

Suponha que temos um espaço de objetos  $\Omega$ , um conjunto de nomes de classes  $Z$ , e uma distribuição de probabilidade  $\mu$  em  $\Omega \times Z$ . Por simplicidade, tomamos as classes como sendo  $+$  e  $-$ . Baseados nos objetos  $X$  e nomes  $Y$  dados por  $\mu$ , queremos aprender uma função de classificação que mapeia cada objeto para sua classe correta. Este é o problema padrão de aprendizado supervisionado.

Normalmente, quando deseja-se aprender um modelo de como mapear objetos para rótulos seleciona-se algum espaço de modelo e escolhe-se o melhor modelo de tal espaço de modelo, preferindo um modelo que se encaixa aos dados mas que não seja muito complexo. A escolha do espaço de modelo exige certo conhecimento do problema.

Há muitas maneiras de especificar um modelo e portanto muitos tipos de espaços de modelos. Pode ser especificado por uma função em um espaço de funções, por uma rede neural com uma certa arquitetura, por uma *spline* com um certo número de nós, por uma expressão booleana com um certo número de termos, ou por outros objetos. Cada uma destas configurações possuem formas correspondentes para determinar a complexidade do modelo e para fazer a seleção do modelo.

Com isto em mente foi proposto uma configuração para a seleção e especificação do modelo em aprendizado supervisionado baseado em um *modelo de fonte latente*. Nesta configuração, o modelo é especificado por uma pequena coleção de *fontes latentes* desconhecidas. É pressuposto que os dados são gerados por uma fonte latente de acordo com um modelo estocástico relacionando fontes latentes e observações. Entretanto, ao invés de realizar suposições sobre os dados através da escolha do espaço de modelos e procurar pelo melhor conjunto de fontes latentes no espaço de modelos, confia-se diretamente nos dados como um *proxy* para as fontes latentes

desconhecidas.

Em outras palavras, a estrutura da função de classificação é totalmente desconhecida. Para resolver isto, foi proposto o seguinte modelo não-paramétrico relacionando objetos observados a seus rótulos. Foi pressuposto que existe um número relativamente pequeno de *fontes latentes* distintas em cada classe que descreve todos os objetos observados naquela classe. Tomando  $t_1, \dots, t_n$  para  $+$  e  $q_1, \dots, q_l$  para  $-$ . Cada observação é rotulada como  $+$  é considerada como sendo uma versão ruidosa de uma das fontes latentes  $t_1, \dots, t_n$ . Da mesma forma, cada observação rotulada  $-$  é considerada uma versão ruidosa de uma das fontes latentes  $q_1, \dots, q_l$ . Não é conhecido quais são as fontes latentes nem quantas são. É apenas conhecido o modelo estocástico que relaciona uma observação para sua fonte latente.

### 3.1.3 Modelo Estocástico

O restante da seção é focado nos sinais que variam com o tempo – os principais objetos de estudo do trabalho. Neste contexto, um objeto observado é simplesmente um sinal em uma janela de tempo com um certo comprimento. Um objeto de fonte latente pode ser imaginado como um sinal correspondente a um protótipo de um tipo de evento. Se o mesmo tipo de evento ocorre muitas vezes, é suposto que os sinais observados são versões ruidosas do sinal de fonte latente correspondente àquele tipo de evento.

É dito que uma observação  $s$  de um fluxo infinito  $S_\infty$  é *gerada* por uma fonte latente  $q$  se  $s$  é uma versão ruidosa de  $q$ . Portanto, é proposto o seguinte modelo estocástico relacionando uma fonte latente  $q$  e uma observação  $s$ :

$$\mathbb{P}(s \text{ gerado por } q) \propto \exp(-\gamma d(s, q))$$

onde  $d(s, q)$  é a *distância* entre  $s$  e  $q$ , e  $\gamma$  é um parâmetro de escala. Isto coincide com a noção de que quanto mais próxima a observação está da fonte latente, mais provável é que a observação tenha vindo daquela fonte. Uma escolha de função de distância pode ser

$$d(s, q) = \sum_{i=1}^{N_{obs}} (s_i - q_i)^2$$

para sinais digitais  $s$  e  $q$  de comprimento  $N_{obs}$ . Entretanto, qualquer  $d$  simétrico, positivo e convexo pode ser utilizado.

### 3.1.4 Detecção

#### 3.1.4.1 Probabilidades de classe

Suponha que  $s$  é um sinal observado de comprimento  $N_{obs}$ . É desejado calcular a probabilidade de  $s$  pertencer a cada classe. Estas probabilidades podem ser utilizadas



para estimar a classe de  $s$ . Para calcular a probabilidade que  $s$  tem de pertencer a cada classe, é feito uso de um conjunto de sinais de *referências* para cada classe – um conjunto  $\mathcal{R}_+$  de exemplos de sinais de  $+$  e um conjunto  $\mathcal{R}_-$  de exemplos de sinais de  $-$ . Sinais de referência representam dados históricos sobre atividades passadas de cada classe para as quais pode-se comparar a observação e traçar conclusões sobre qual classe ela pertence. Os sinais de referência terão comprimento  $N_{ref} \geq N_{obs}$ . Será tratado o caso  $N_{ref} = N_{obs}$  inicialmente e generalizado mais adiante.

Neste modelo a observação deve pertencer a  $+$  se possui a mesma fonte latente de um dos sinais de referência de  $\mathcal{R}_+$ , e deve pertencer a  $-$  se possui a mesma fonte latente de um dos sinais de referência de  $\mathcal{R}_-$ . Esta conta envolve as distâncias entre os sinais de referências e as fontes latentes que as originaram e as distâncias entre as observações e as mesmas fontes latentes. Busca-se a menor soma das duas diferenças. Entretanto, a conta leva em consideração as fontes latentes, que são desconhecidas. Para eliminar este fator desconhecido, supõe-se que o espaço de sinais é bem coberto pelas fontes latentes, e que a fonte latente que minimiza as distâncias é simplesmente a média de  $s$  e  $t$ . A Figura 1 ilustra o sinal de referência  $r$ , a observação  $s$ , as fontes latentes  $t_1, \dots, t_n$ , e a fonte latente minimizadora  $t_{j*}$ .

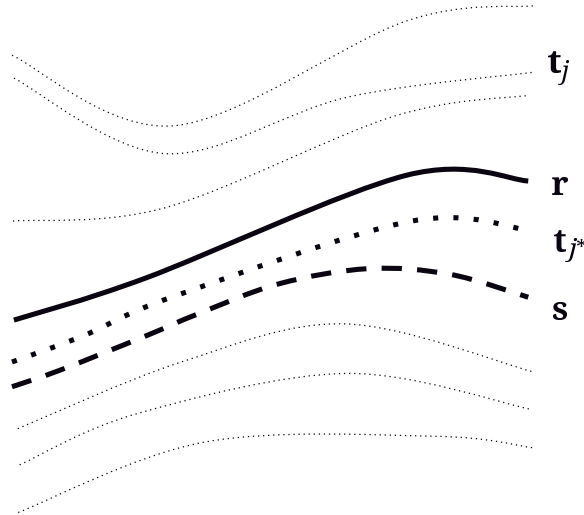


Figura 1: Fonte latente  $t_{j*}$ , média que minimiza a distância entre o sinal de referência  $r$  e a observação  $s$ . Os pontilhados são as fontes latentes  $t_j$ . Retirado de NIKOLOV (2012)

Ao final da construção do cálculo, a probabilidade de que a uma observação pertença a  $+$  é:

$$\mathbb{P}(+ \mid s) \propto \sum_{r \in \mathcal{R}_+} \exp(-\gamma d(s, r))$$

e, de forma similar, a probabilidade de que a uma observação pertença a  $-$  é:

$$\mathbb{P}(- \mid s) \propto \sum_{r \in \mathcal{R}_-} \exp(-\gamma d(s, r))$$

#### 3.1.4.2 Estimador de classe

A estimação da classe é feita de forma simples, atribuindo à observação a classe com a maior probabilidade. É calculada a razão  $R(s)$  entre  $\mathbb{P}(+ \mid s)$  e  $\mathbb{P}(- \mid s)$  e verificada se excede ao limite  $\theta = 1$ . O estimador de classe  $L(s)$  é portanto:

$$L(s) = \begin{cases} +, & \text{se } R(s) > \theta. \\ -, & \text{se } R(s) \leq \theta. \end{cases}$$

Outros valores de  $\theta$  podem ser usados. Por exemplo, se positivos verdadeiros possuem maior importância que falsos positivos, o valor de  $\theta$  pode ser menor que 1, e maior que 1 caso a importância seja invertida.

#### 3.1.4.3 Classificação online

Até então é assumido que os sinais de referência e as observações possuem o mesmo comprimento. Na classificação online é conveniente estender a classificação para uso de sinais de referência de qualquer comprimento. Neste momento, o método compara uma observação – recentes observações sobre uma propriedade de um sistema – a sinais de referências – conjuntos de medições históricas daquela propriedade para cada classe. Medições recentes são julgadas como pertencendo à classe cujos sinais de referências mais se assemelham. Para sinais de referência e observações do mesmo tamanho o cálculo da distância é feito como tratado anteriormente. Entretanto, na prática os sinais de referências são maiores que as observações e estas podem ser representações de qualquer trecho<sup>1</sup> dos sinais de referências, podendo estar deslocadas no tempo. Assim é necessário verificar se a observação se assemelha com qualquer pedaço de mesmo tamanho do sinal de referência.

A Figura 2 ilustra esta verificação. Do lado esquerdo temos exemplos de sinais de referências. As linhas tracejadas são referências de eventos, e as outras linhas são referências de não-eventos. Do lado direito temos a observação atual do sinal. Ele é comparado com todos os sinais de referências, trecho a trecho, como ressaltado pela linha em negrito do lado esquerdo, e sua distância é calculada.

Para adequação a este problema, a função de distância é generalizada. Assumindo que todas as observações são do tamanho  $N_{obs}$  e todos os sinais de referência são do tamanho  $N_{ref} \geq N_{obs}$ , é definida a distância entre  $r$  e  $s$  como a mínima distância entre

<sup>1</sup>Um *trecho* de série (ou de sinal) é uma sub-série contígua da série original. Dada uma série de  $n$  elementos  $s_{1,n}$ , um trecho desta série é uma série  $s'_{i,j} \in s$ , tal que  $1 \leq i \leq j \leq n$ .

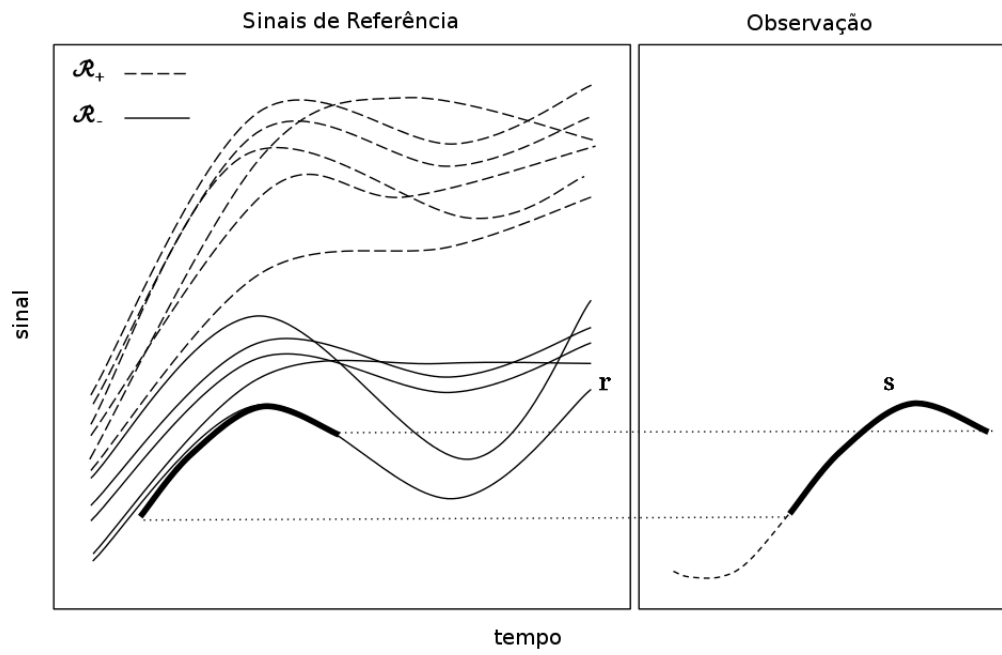


Figura 2: Comparando um pedaço do comprido sinal de referência com uma curta observação. Retirado de NIKOLOV (2012)

$s$  e todos os sub-sinais contíguos de  $r$  com comprimento  $N_{obs}$ :

$$d(s, r) = \min_{k=1, \dots, N_{ref} - N_{obs} + 1} d(r_{k:k+N_{obs}-1}, s)$$

## **4 METODOLOGIA**

Este Capítulo apresenta a metodologia a ser utilizada no projeto, algumas dificuldades e o processo de escolha das decisões. Na Seção 4.1 apresentaremos os objetivos do trabalho. Na Seção 4.2 descrevemos as adaptações do algoritmo, enumerando particularidades dos ambientes de aplicação, dificuldades na adaptação e como se dão as escolhas nos ajustes.

### **4.1 Objetivos**

O objetivo principal deste trabalho é verificar a aplicabilidade do algoritmo de Nikolov como uma ferramenta de operação no mercado de renda variável. Mais especificamente, em operações de compra e venda de ações na Bolsa de Valores de São Paulo – BM&FBovespa. Isso será feito através de simulações de operações de compra e venda de ativos e posterior comparação dos resultados obtidos com uma ferramenta tradicional de análise técnica bastante difundida entre operadores de bolsa de valores. Sendo o algoritmo de Nikolov um algoritmo de aprendizado de máquina, ele será treinado segundo indicações da ferramenta tradicional. Assim ele saberá o que é entendido por tendência e o que não é entendido como tendência baseando-se nas informações desta ferramenta.

Sabendo deste comportamento do algoritmo, espera-se que ele saiba identificar quando realizar uma operação de compra e uma operação de venda de forma semelhante à ferramenta que foi utilizada para treinamento. Além disso, como será treinado também com informações de operações mal-sucedidas, seria interessante que o algoritmo filtrasse estas operações em suas decisões. Realizaria as operações que trouxesse resultado positivo, e evitaria de realizar as operações que trariam algum prejuízo. Desta forma estaria melhorando a performance se comparado aos resultados da ferramenta tradicional.

Como forma de medida, utilizaremos o retorno financeiro final, porcentagem de operações lucrativas, e retorno médio por operação.

## 4.2 Adaptação do algoritmo

O algoritmo de Nikolov foi desenvolvido com o intuito de ser um algoritmo genérico, podendo ser fornecido com dados de qualquer origem. Entretanto ele foi utilizado em um caso de teste específico, detecção de tendências em séries temporais de uma rede social. Tal ambiente não possui o mesmo comportamento que o ambiente planejado para este trabalho, necessitando assim de algumas alterações.

### 4.2.1 Características dos dois ambientes

Explicitamos aqui algumas das diferentes características dos dois ambientes. Estas foram as principais características que levaremos em consideração para realizar as adaptações do algoritmo. Um resumo das mesmas encontra-se na Tabela 1.

#### **Influência dos *players***

O comportamento dos *players* influencia de forma diferente os dois ambientes. Em uma rede social, para definição de tendência é utilizada a informação do volume de mensagens de um determinado tópico. Cada usuário pode enviar uma mensagem por vez, e esta mensagem tem peso um na contagem do volume do tópico específico. Portanto, para uma movimentação significativa na série, é necessário que um grande número de usuários envie mensagens com o tópico específico. Além disso, uma ação do usuário exerce pressão para apenas um sentido na série: para cima. Como não há a possibilidade de um usuário enviar uma mensagem – ou ordem – para diminuir o volume, suas ações influenciam apenas no incremento do volume de um tópico, e não no decremento do mesmo.

Já em séries temporais de ações em bolsa de valores, a definição de tendência geralmente utiliza informações de preço do ativo. E, diferentemente do que ocorre na rede social, um operador – ou *player* – pode enviar ordens – ou mensagens – com pesos diferentes. Uma ordem pequena – envolvendo poucas ações – influencia pouco na série, enquanto que uma única ordem grande – envolvendo muitas ações – pode levar a série a outros patamares. Esta ordem do operador pode exercer pressão em dois sentidos da série, para cima ou para baixo, dependendo do tipo da ordem. Existem ordens de compra e ordens de venda. Uma ordem de compra de tamanho considerável exerce uma pressão tal que pode elevar os preços da série temporal, e de forma análoga a uma ordem de venda de tamanho considerável exerce uma pressão que pode diminuir os preços da série.

## Continuidade

Uma rede social tem um funcionamento contínuo, vinte e quatro horas por dia, sete dias por semana. Não apresenta um horário de início e fim das atividades. Isso faz com que sua série temporal seja contínua. Dados dois momentos quaisquer, com informações de dia, hora e minuto, pode-se seguir uma linha contínua de um ponto ao outro. Isso permite um comportamento suavizado da série, uma vez que as movimentações ocorrem de forma gradual.

Uma bolsa de valores, por outro lado, possui horário de abertura e de fechamento, possibilitando operações apenas durante este intervalo de tempo. Durante o tempo em que a bolsa de valores está fechada não existem negociações. Entretanto a empresa que possui ações continua em funcionamento, e neste período podem ocorrer fatos que alterem o valor desta empresa perante o entendimento dos operadores. Caso isso ocorra, na abertura do pregão seguinte, as ações desta empresa podem sofrer grandes alterações de preço. Assim, a série temporal não é contínua e esta descontinuidade pode apresentar movimentações repentinas na série que, sem aviso suficiente, dificulta identificações.

## Granularidade

Chamaremos de granularidade o período de agrupamento dos dados para geração de um único ponto na série, após realização de uma computação. Esta granularidade pode ser tão *fin*a quanto o menor valor de período disponível nos dados, como um segundo, e tão *gross*a quanto for possível agrupar dados em um período, como em semanas, meses, anos, e assim por diante.

A rede social permite a escolha de qualquer granularidade. Como a intenção de Nikolov em seu caso teste era de identificar e possivelmente prever ocorrências de tendências ele escolheu para sua análise um período curto mas com informações suficientes. Assim, cada ponto em sua série temporal representava a soma do volume de mensagens em um período de dois minutos.

Diferentemente do que é feito nas séries das redes sociais, os agrupamentos nas séries de ações servem como uma espécie de filtro. Ao invés de somar todos os elementos do período para formação de um ponto, neste tipo de série é feita uma replicação do último valor do período. As cotações das bolsas de valores também possibilitam a escolha de praticamente qualquer período para granularidade. Algumas ferramentas permitem formular séries temporais com períodos tão pequenos quanto microssegundos. São informações extremamente precisas mas existem dois motivos que impossibilitam, neste momento, uso de períodos *intraday* (menores do que um dia). O primeiro diz respeito à descontinuidade da série. Quanto menor o período



utilizado para agrupamento dos dados, maior é a influência exercida pela diferença dos preços de abertura e fechamento do mercado. Um gráfico com período de dois minutos, por exemplo, pode sofrer uma variação muito grande entre o fechamento de um pregão e a abertura do pregão seguinte resultando em uma movimentação não treinada e, portanto, inesperada. O segundo motivo diz respeito à disponibilidade dos dados de cotações *intraday*. Tais dados não são facilmente encontrados, como os dados de período diário. Eles precisam ser gerados ou comprados. Como não dispomos de recursos suficientes para tal finalidade, esta não se mostra uma opção viável.

### **Definição de tendência**

O algoritmo faz uso de conjuntos de referências, que são conjuntos de amostras do que queremos identificar e amostras do que não nos interessa. Para que seja feita uma identificação eficiente precisamos que estes conjuntos sejam populados com bons exemplos de eventos – movimentos com tendências – e não-eventos – movimentos sem tendência. Como então identificamos estes exemplos e criamos os conjuntos?

Nikolov, em seu caso teste, tinha estas informações prontas. A própria rede social apresenta uma lista dos assuntos mais comentados – os tópicos em tendência. Para preencher seu conjunto de eventos basta consultar os dados disponíveis, localizar onde a tendência foi identificada no assunto, e inserir o exemplo no conjunto. Para preencher o conjunto de não-eventos ele selecionou os assuntos que não apareciam na lista de assuntos mais comentados e, para cada um dos selecionados, escolheu aleatoriamente um trecho. Este trecho era então inserido no conjunto de não-eventos.

A criação destes conjuntos é menos trivial no ambiente deste trabalho. Não existe simplesmente uma lista das ações que entraram em tendência, ou uma lista de eventos a serem identificados. A própria definição de tendência no mercado de ações não é única e, com frequência, não é determinística. A definição de tendência segundo Dow (RHEA, 1993; SHAN, 2012) é facilmente percebida de forma visual, mas não tão simples de identificar utilizando código. Existem ainda vários rastreadores de tendências que foram criados para facilitar a identificação destas (ROBERTS, 1959; SHARPE; ALEXANDER; BAILEY, 1999; CORREIA, 2008). Apesar de resolverem o problema da falta de determinismo, a grande variedade dificulta a escolha, pois cada ferramenta apresenta um comportamento diferente, e algumas vezes indicações contraditórias.

O preenchimento dos conjuntos de referências com bons exemplos é parte crucial para o funcionamento eficiente do algoritmo. O resultado obtido por um algoritmo que utiliza exemplos não representativos no conjunto de referências pode ser muito diferente do resultado obtido pelo mesmo algoritmo treinado com exemplos representativos. Dentre as diferenças dos dois ambientes, encontrar uma definição eficiente para

tendência, uma que gere bons exemplos para treinamento, é o maior dos desafios.

A Tabela 1 apresenta um resumo destas diferenças.

<b>Característica</b>	<b>Rede social</b>	<b>Bolsa de valores</b>
<b>Peso da ação dos <i>players</i></b>	Peso um	Peso variado
<b>Influência dos <i>players</i></b>	Apenas incremento	Incremento e decremento
<b>Continuidade</b>	Contínua	Descontínua
<b>Granularidade</b>	Fina	Média
<b>Definição de tendência</b>	Pré definida e disponível	Variada e não disponível

Tabela 1: Comparação das características dos dois ambientes.

### 4.3 Seleção das ações e do período histórico

Para a seleção das ações a serem analisadas temos muitas opções. Na Bovespa existem centenas de ações, índices, e derivativos. Pretendemos ter uma seleção que não seja muito grande mas que ao mesmo tempo tenha uma quantidade suficiente de exemplos para representar significativamente o comportamento do mercado. Começamos então a eliminação das opções pelos derivativos, pois possuem seus valores baseados nas ações. Sendo a própria ação mais significativa que o seu derivativo, não há razão para utilizar este. Outra opção é utilização de índices. Eles são ótimos indicadores de mercado, mas ao mesmo tempo filtram muitos movimentos menores por funcionarem como uma média de várias ações. Além disso, poucos índices possuem um volume suficiente para realização de operações que iremos simular. Das três opções ficamos então com as ações.

Ainda é necessário escolher um conjunto menor de ações. Seguindo a ideia de ter um conjunto pequeno mas representativo, utilizamos a separação que a Bovespa faz das ações por setores da economia<sup>1</sup>. Nesta listagem há dez setores. Dentre as ações de cada setor, verificamos seu valor de mercado<sup>2</sup> e escolhemos a de maior valor como representante do setor. Desta forma selecionamos o conjunto de ações significativas para nossa análise. As ações, os códigos de seus ativos e respectivo setor estão listadas na Tabela 2.

Vale uma observação sobre a escolha do ativo LREN3. A empresa com o maior valor de mercado do setor de consumo não cíclico é a Kroton Educacional SA (KROT3). Entretanto, devido ao histórico muito curto de KROT3, se comparado com os outros ativos e com o período analisado, ele foi substituído por LREN3, a empresa com segundo maior valor de mercado do mesmo setor.

Para obtenção de melhores resultados no algoritmo de aprendizado de máquina,

<sup>1</sup><http://www.bmfbovespa.com.br/cias-listadas/empresas-listadas/BuscaEmpresaListada.aspx?idioma=pt-br>

<sup>2</sup><http://exame.abril.com.br/mercados/cotacoes-bovespa/acoes/>

Ativo	Empresa	Setor da economia
CPFE3	CPFL Energia SA	Utilidade pública
TIMP3	TIM Participações SA	Telecomunicações
TOTS3	Totvs SA	Tecnologia da informação
PETR4	Petróleo Brasileiro SA	Petróleo e gás
VALE5	Vale SA	Materiais básicos
ITUB4	Itaú Unibanco Holding SA	Financeiro e outros
BRFS3	BRF SA	Consumo não cíclico
LREN3	Lojas Renner SA	Consumo cíclico
CYRE3	Cyrela Brazil Realty SA	Construção e transporte
EMBR3	Embraer SA	Bens industriais

Tabela 2: Ativos selecionados para o trabalho, referentes às maiores empresas de cada setor da economia.

é interessante que todos os exemplos fornecidos na fase de testes tenham algum correspondente na fase de treinamento. Desda forma o algoritmo tem a capacidade de reconhecer novas ocorrências dos eventos. No nosso ambiente, isso significa fornecer uma boa quantidade de exemplos de tendências. Procurando fornecer ao treinamento do algoritmo exemplos suficientes, escolhemos uma janela de tempo de dez anos para treinamento.

#### 4.4 Obtenção e ajustes das séries temporais

Para obtenção do histórico das cotações das ações utilizamos o Yahoo Finance<sup>3</sup>. A ferramenta disponibiliza gratuitamente um longo histórico em período diário, suficiente para a janela que selecionamos para análise. Além disso os dados apresentam as cotações com ajustes e poucos erros, sendo uma fonte aceitável.

Das ações cujo histórico é longo o suficiente, toda a janela é preenchida. Das outras, todo o histórico foi obtido, e foi aceito mesmo sendo menor que a janela. No passo seguinte removemos dos históricos os dias onde não houveram negociações dos ativos. Estes dias acontecem por dois motivos principais. O primeiro, pelos dados obtidos seguirem o calendário Norte Americano e a Bovespa o de São Paulo, se dá pela diferença de datas comemorativas entre os dois lugares. Os dias em que a Bovespa não abria para negociações eram contabilizados pela ferramenta como dias onde a Bovespa abriu mas não houve nenhuma negociação, deixando o volume em zero e os preços dos ativos inalterados. O segundo se dá pela própria bolsa de valores brasileira. Em algumas situações específicas, ela interrompe as negociações de um ou de um grupo de ações por um período determinado. Essa interrupção pode durar um dia inteiro, e este dia também apresenta volume zero de negociações.

Outro ajuste necessário nos dados é a remoção de parte dos pontos iniciais de

<sup>3</sup><http://finance.yahoo.com/>

algumas ações. Alguns dos ativos apresentam alterações no preço, de um dia para o outro, com um fator alto de multiplicação ou de divisão. Isto pode ser resultado de operações planejadas, normais neste mercado, mas não relevante para o tipo de aplicação que abordamos neste trabalho.

Por exemplo, um dos ativos analisados apresentava uma cotação na casa dos trinta reais em um dia, e no dia seguinte sua cotação fechou abaixo de dez reais. Neste caso específico, houve uma operação de *split*. Neste tipo de operação o valor de um ativo é dividido propositadamente por um determinado fator, e a quantidade de ações disponíveis é multiplicada pelo mesmo fator, deixando inalterado o valor de mercado da companhia. Supondo que a cotação de um ativo esteja a R\$30,00 e um operador tenha em sua carteira 100 ações desta companhia, uma operação *split* que leve a cotação do ativo a R\$10,00 (dividindo a cotação por 3) faria com que o operador tivesse agora em sua posse 300 ações (quantidade multiplicada por 3), sem prejuízo algum ao seu investimento inicial de R\$3.000,00.

Esta diferença, apesar de não influenciar em nada operações como no caso acima, pode causar anomalias no comportamento do algoritmo, que não está preparado ainda para realizar este tipo de correções na série temporal original. Isso justifica a remoção de todo o histórico deste ativo desde seu início até o momento seguinte ao evento.

## 4.5 Divisão do histórico em 3 conjuntos

Tendo o histórico recebido os ajustes anteriormente descritos, precisamos realizar a separação do mesmo. Sendo o algoritmo de Nikolov um algoritmo de aprendizado de máquina, é necessário que ele seja treinado e sua eficiência avaliada. Para isso é preciso realizar a escolha de exemplos para formar um conjunto de treinamento, e outros exemplos para formar um conjunto de validação e de teste (PARDO, 1992).

Dependendo do ambiente no qual se aplica o aprendizado de máquina, uma série temporal pode ser dividida em exemplos independentes. Após esta divisão a separação dos exemplos em grupos e posterior utilização deles se torna trivial. Entretanto, no nosso ambiente, vamos mensurar também o retorno financeiro do uso do algoritmo em negociações simuladas. E, para medir a evolução de uma negociação, é necessário ter acesso às informações do ponto da série imediatamente posterior ao ponto atual. Dividir a série em blocos faria com que informações das negociações fossem perdidas, alterando o resultado final. Por este motivo, as séries temporais não poderão ser divididas em blocos de tamanhos iguais.

Assim sendo, resolvemos utilizar a série histórica das ações dividida em três grandes partes, uma para cada fase do algoritmo. Olhando a série de trás pra frente – do dado mais recente para o mais antigo – realizamos a primeira divisão separando os seis meses mais recentes para a fase de testes. Estes dados serão utilizados apenas

uma vez, para verificar a eficiência do algoritmo. Do restante da série, separamos novamente os últimos seis meses para a fase de validação. Estes dados serão utilizados para buscar qual a configuração<sup>4</sup> das variáveis do algoritmo que oferece os melhores resultados, segundo critérios estabelecidos. O restante final da série é utilizada em sua totalidade na fase de treinamento do algoritmo. A equação 1 oferece uma visão mais direta desta divisão.

$$Serie = SerieTreinamento \oplus SerieValidacao \oplus SerieTeste \quad (1)$$

---

<sup>4</sup>Dentro da exploração dos valores de conjuntos de variáveis uma configuração é uma das possibilidades de atribuição de valores a estas variáveis. Por exemplo, dadas as variáveis  $(x, y, z)$ , que aceitam valores inteiros de 0 a 9, duas configurações possíveis seriam  $(1, 4, 5)$  e  $(1, 4, 8)$ .

## 5 ABORDAGENS

Este capítulo trata das diferentes abordagens de adaptação que faremos no algoritmo de Nikolov. As duas primeiras abordagens utilizam ideias originais, enquanto que as abordagens sucessivas apresentam melhorias incrementais às suas antecessoras.

Nas três primeiras abordagens todo o processo é realizado por ativo. Para o primeiro ativo, a abordagem constrói os conjuntos de referências e realiza a execução simulando operações de compra e venda. Para os ativos seguintes, os conjuntos de referências são descartados e recriados, e então uma nova execução é realizada contendo apenas informações do ativo atual. A última abordagem, entretanto, difere das iniciais pois ela primeiro gera todos os conjuntos de referências, agrupa-os em um só – um conjunto global de referências – e então realiza a execução para cada ativo.

A dificuldade principal de todo o processo de adaptação está na geração dos conjuntos de referência para o algoritmo. Mais precisamente, na definição do evento a ser identificado. No nosso caso, uma tendência. Como não existe uma definição precisa única, global, temos que escolher uma das definições existentes ou gerar alguma definição própria para uma tendência. E a qualidade dos resultados está intimamente ligadas à qualidade desta definição. Aqui apresentamos algumas destas definições. Para trabalhos futuros é muito interessante que outras definições sejam testadas e comparadas.

### 5.1 Abordagem 1

Abordagem que utiliza um limite de valorização dentro de um espaço temporal como definição de tendência.

#### **Definição de tendência**

Em nossa primeira abordagem utilizamos uma definição simplificada de tendência. Assumimos que, em uma tendência altista, o preço de um ativo aumenta com o tempo. Aceitamos a existência de alguma breve diminuição em determinados pontos, mas, de modo geral, se escolhermos dois pontos diferentes da série com um determinado



tempo de diferença entre eles, o ponto mais recente apresenta um valor maior que o ponto mais antigo. Se esta diferença de valores for maior que um limite determinado, dizemos que o ativo apresenta uma tendência de alta.

Então, nesta abordagem, dados dois pontos em uma série, um limite de valorização, e um tempo, identificamos uma tendência se a diferença entre os valores dos dois pontos for igual ou maior este limite. Para uma série  $S$ , com índice  $i$ , um intervalo de tempo  $t$  e um limite de valorização  $l$ , a Equação 2 indica a ocorrência de uma tendência.

$$\text{Tendência} = \begin{cases} 1 & \text{se } (S[i+t] - S[i]) > l \\ 0 & \text{caso contrário} \end{cases} \quad (2)$$

A Figura 3 exemplifica uma detecção de tendência segundo esta definição. A valorização do ativo entre os instantes  $S_i$  e  $S_{i+t}$  é maior que o limite  $l$  definido, caracterizando portanto uma tendência. Desta forma, o trecho de série anterior ao início da tendência, ou seja, anterior ao instante  $S_i$ , é inserido no conjunto de referências.

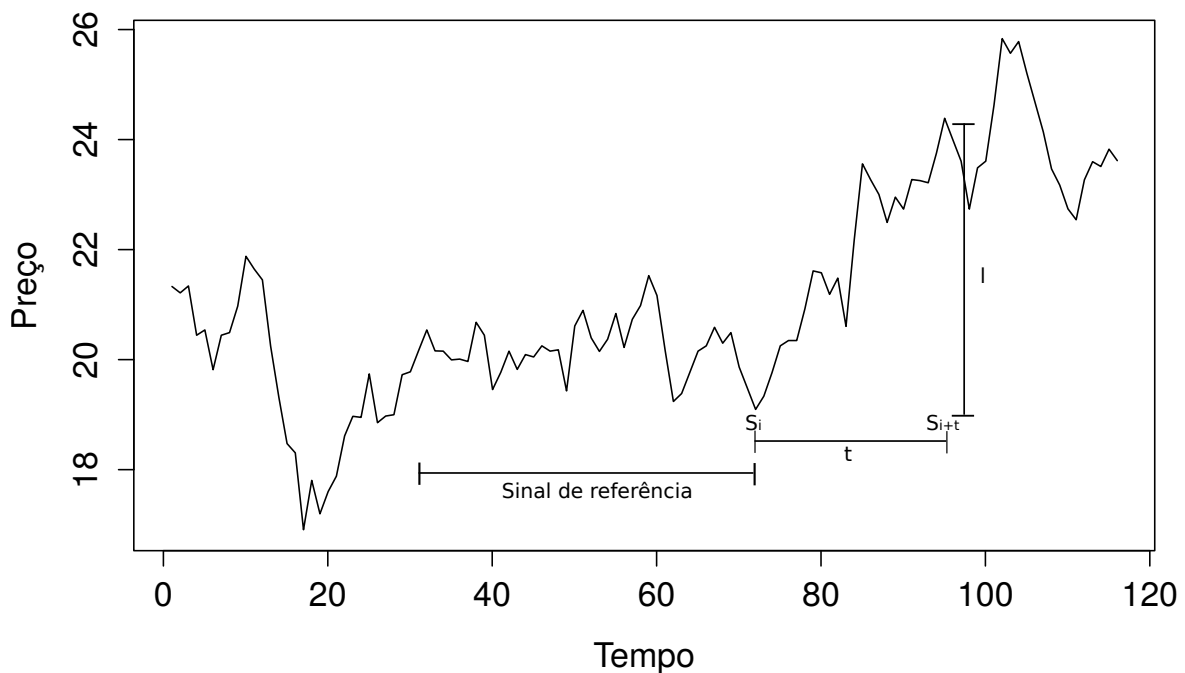


Figura 3: Definição de tendência da Abordagem 1.

### Conjuntos de referência

Toda tendência identificada pela Equação 2 é inserida no conjunto de exemplos

positivos. Para a criação do conjunto de exemplos negativos, que não são tendências, são escolhidos trechos aleatórios da série. Cada trecho da série que não é identificado como tendência tem uma probabilidade de ser inserido no conjunto. Para evitar que este segundo conjunto fique demasiadamente grande, limitamos seu tamanho ao tamanho do primeiro conjunto. Assim, os conjuntos positivos e negativos teriam praticamente a mesma quantidade de elementos. Note que existe uma aleatoriedade na construção do conjunto de exemplos negativos.

### **Execução**

Durante a execução do algoritmo, em tempo real, para cada novo valor da série temporal apresentada verificamos se aquele trecho da série assemelha-se com os trechos no conjunto positivo. Caso se assemelhe, é indicado o início de uma tendência e uma operação de compra é iniciada pelo simulador. A cada valor seguinte, verifica-se novamente o trecho, e caso este não se assemelhe com os trechos no conjunto positivo, uma venda é realizada dando fim à operação de compra. A partir deste momento, estaremos sempre em uma posição de compra ou em uma posição de venda. Assim, no mesmo instante que a operação de compra for encerrada, uma operação de venda tem início. Esta alternância entre operações de compra e operações de venda continua até o final da série.

## **5.2 Abordagem 2**

Abordagem que inicia uso de médias móveis para definição de tendências.

### **Definição de tendência**

Nesta segunda abordagem alteramos a definição de tendência. Ao invés de utilizarmos a definição gerada na Abordagem 1, utilizamos um indicador técnico conhecido no mercado de ações para nos auxiliar nesta tarefa, as Médias Móveis.

Uma das formas de utilizar as médias móveis é fazendo uso de seus cruzamentos. Utiliza-se duas médias móveis, cada uma com um período  $n$  diferente. Calcula-se então a média dos últimos  $n$  períodos da série e adiciona-se esta informação ao gráfico. Como as médias são de valores diferentes, elas apresentarão curvas diferentes e eventualmente se cruzarão. Estes cruzamentos indicam início de uma tendência e, conseqüentemente, final de outra. Quando a média com período menor cruza a média com período maior de baixo para cima temos uma indicação de tendência altista – e conseqüente fim da tendência baixista. Quando a média com período menor cruza a média com período maior de cima para baixo temos uma indicação de tendência baixista – e conseqüente fim da tendência altista.

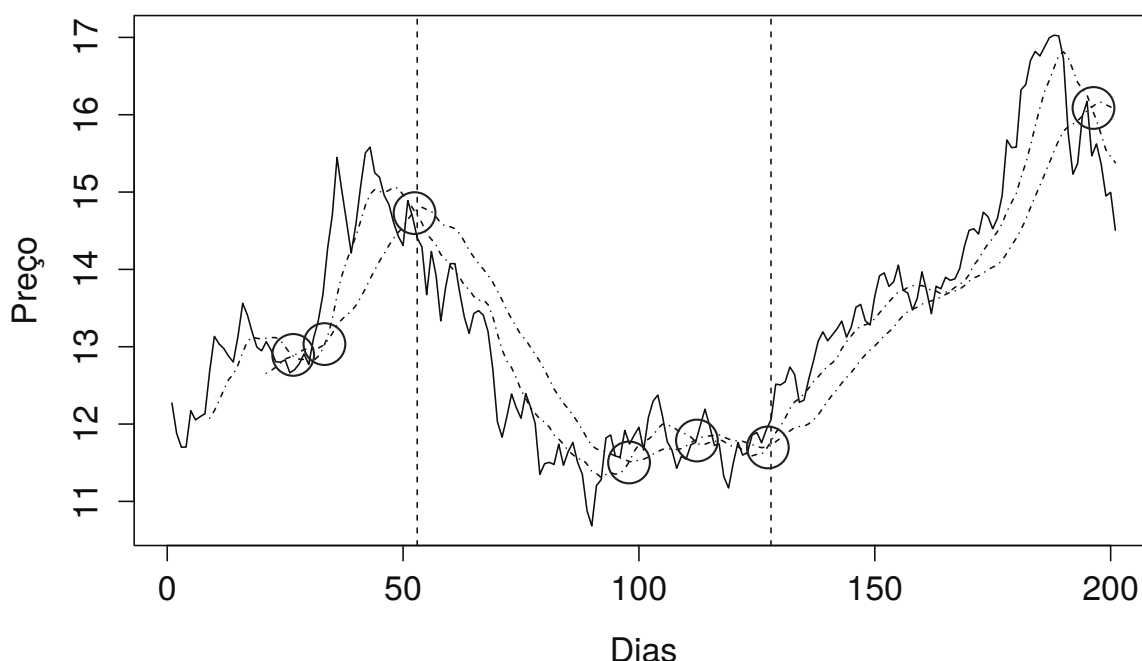


Figura 4: Cruzamento de médias móveis de períodos 9 e 21

Na Figura 4 temos uma amostra deste comportamento. Nela vemos a oscilação do preço de um ativo durante duzentos dias, indicado pela linha mais forte. Em linhas tracejadas, suavizadas e que seguem a linha do preço, estão as duas médias móveis, uma de 9 e outra de 21 períodos. Os círculos indicam momentos onde as linhas das médias móveis se cruzam. As linhas verticais tracejadas enfatizam os momentos dos dois cruzamentos das médias móveis que melhor exemplificam a identificação de tendências realizada por esta técnica. O primeiro cruzamento, na posição 53, indica o cruzamento para baixo – média móvel mais curta cruzando de cima para baixo a média móvel mais longa – simbolizando o final de uma tendência de alta e início de uma tendência de baixa. O segundo, na posição 128, indica um cruzamento para cima, simbolizando final de uma tendência de baixa e início de uma tendência de alta.

A fim de encontrar o melhor par de períodos para as médias móveis, para cada ativo foi realizada uma exploração de todos os pares possíveis de  $n$ , com  $1 \leq n \leq 50, n \in \mathbb{Z}$ . Verificamos o retorno obtido que cada par traria caso fosse utilizado com a estratégia de cruzamentos ao longo do período de treinamento. O par que mostrasse o maior retorno seria utilizado como parâmetro para a criação dos conjuntos de referências.

### Conjuntos de referência

Os conjuntos de referências são formados pelos cruzamentos das médias móveis encontrados no passo anterior. Todo cruzamento para cima sinaliza uma tendência

altista, indicando uma possível entrada em uma operação de compra. Logo, este trecho é inserido no conjunto de exemplos positivos. De forma análoga, todo cruzamento para baixo sinaliza início de uma tendência baixista, e seu trecho da série é inserido no conjunto de exemplos negativos. Nesta abordagem não temos aleatoriedade na criação dos conjuntos de referência.

### **Execução**

A execução do algoritmo se dá da mesma forma que na Abordagem 1. Verificamos os trechos da série até que o algoritmo acuse identificação de tendência altista, e iniciamos uma operação simulada de compra. Permanecemos na operação enquanto a indicação de tendência continuar. A finalizamos quando o algoritmo acusar que o trecho atual não se assemelha ao conjunto de exemplos positivos, iniciando no mesmo momento uma operação de venda.

## **5.3 Abordagem 3**

Abordagem com adição de um terceiro conjunto de referência. Utilizamos dois algoritmos de Nikolov.

### **Definição de tendência**

A definição de tendência segue igual à da Abordagem 2. Utilizamos o cruzamento de médias móveis como parâmetro.

### **Conjuntos de referência**

Verificamos que, até a abordagem anterior, inserimos no conjunto de referências positivas ( $\mathcal{R}_+$ ) as indicações de tendências altistas dadas pelos cruzamentos das médias móveis, e inserimos no conjunto de referências negativas ( $\mathcal{R}_-$ ) as indicações de tendências baixistas dadas pelos mesmos cruzamentos. Porém, analisando o funcionamento do algoritmo, percebemos que existem duas situações ignoradas por esta sequência.

A primeira é sobre a *qualidade* das tendências indicadas. Ao inserir um exemplo em um conjunto de referência, simplesmente olhamos o cruzamento das médias móveis e inserimos o trecho indicado no conjunto de exemplos positivos ou negativos. Entretanto, esta técnica – cruzamento de médias móveis – não possui cem por cento de acertos em suas indicações. Ela falha com frequência, indicando uma tendência altista que não se confirma. Caso alguém realize operações de compra e venda seguindo esta técnica, esta pessoa terá alguns resultados negativos, com prejuízo. E, até o momento, estas indicações de tendências estão sendo inseridas no conjunto

positivo,  $\mathcal{R}_+$ , independente se for um falso positivo ou não.

Com o intuito de evitar treinar o algoritmo com exemplos falhos, criamos um novo conjunto de referência,  $\mathcal{R}_0$ . Para cada indicação de tendência das médias móveis, verificamos se esta é uma indicação verdadeira ou não. Para isso, verificamos a diferença entre o preço no cruzamento atual e o preço no próximo cruzamento. Caso esta operação simulada apresente lucro, o exemplo apontado pelo primeiro cruzamento é direcionado para seu respectivo conjunto,  $\mathcal{R}_+$  ou  $\mathcal{R}_-$ . Caso apresente prejuízo, uma falsa indicação, o exemplo apontado pelo cruzamento é direcionado para o conjunto neutro,  $\mathcal{R}_0$ . Com esta separação temos a certeza de que todos os exemplos em  $\mathcal{R}_+$  e  $\mathcal{R}_-$  são bons exemplos, e os exemplos em  $\mathcal{R}_0$  não foram boas indicações.

A segunda situação é sobre o funcionamento do algoritmo. Ele foi desenvolvido para identificar ocorrências de uma, e apenas uma, classe de eventos. Ele compara uma observação atual com um conjunto de exemplos de eventos e com um conjunto de exemplos de não-eventos. Seria perfeito para o nosso caso se estivéssemos procurando identificar somente um tipo de tendência, a altista por exemplo. Entretanto, precisamos identificar tendências baixistas também, pois o início de uma tendência baixista é, segundo a definição utilizada, o final de uma tendência altista. Desta forma, temos que identificar dois tipos de eventos, e não apenas um. Justificamos assim, mais uma vez, a adição de um terceiro conjunto de referências.

## Execução

Com a inserção de um terceiro conjunto de referência temos que alterar a execução do algoritmo, que é preparado para uso de apenas dois conjuntos. Vamos adaptar a parte de detecção do algoritmo para que ele simule a utilização de dois algoritmos originais.

Quando estamos em uma operação de compra, procuramos uma indicação de tendência baixista para finalizar a operação. Assim, utilizamos os conjuntos  $\mathcal{R}_0$  e  $\mathcal{R}_-$  na detecção. Cada observação é comparada com os exemplos dos dois conjuntos, e assim que ela for indicada como pertencente a  $\mathcal{R}_-$ , sinalizando final da tendência de alta, a operação de compra é finalizada.

De forma análoga, quando estamos em uma operação de venda, procuramos uma indicação de tendência altista para finalizar a operação. Utilizamos então os conjuntos  $\mathcal{R}_0$  e  $\mathcal{R}_+$  na detecção. Cada observação é comparada com os exemplos dos dois conjuntos, e assim que ela for indicada como pertencente a  $\mathcal{R}_+$ , a operação de venda é finalizada.

Assim temos uma simulação de dois algoritmos de Nikolov, um para identificar início de tendências de baixa e outro para identificar início de tendências de alta.

## 5.4 Abordagem 4

Abordagem com três conjuntos de referência, agora globais. Seguimos utilizando dois algoritmos de Nikolov.

### Definição de tendência

A definição de tendência segue igual à da Abordagem 2. Utilizamos o cruzamento de médias móveis como parâmetro.

### Conjuntos de referência

A quantidade de conjuntos de referência e a forma com que são preenchidos é igual à da Abordagem 3. A diferença agora é que esta abordagem utiliza conjuntos globais.

Até a abordagem anterior, cada ativo tinha seus próprios conjuntos de referência,  $\mathcal{R}_-$ ,  $\mathcal{R}_+$  e  $\mathcal{R}_0$ , gerados por exemplos da série temporal deste ativo. Na fase da detecção, cada ativo utilizava seus próprios conjuntos para identificar inícios de tendências.

Supondo que ativos diferentes podem apresentar comportamentos semelhantes, podemos tentar aproveitar exemplos de um ativo para identificar tendências em outro. Isso aumentaria os conjuntos de referências, dando mais possibilidades de identificações. Então, agrupamos os conjuntos de referências de todos os ativos em um único conjunto. Temos assim conjuntos  $\mathcal{R}_-$ ,  $\mathcal{R}_+$  e  $\mathcal{R}_0$  globais. Os mesmos conjuntos de referência são utilizados na detecção de tendências de todos os ativos.

### Execução

A execução do algoritmo segue igual à da Abordagem 3.

## 6 RESULTADOS

Os resultados estão divididos em duas partes principais. Na primeira descrevemos os resultados obtidos na Abordagem 1. Na segunda descrevemos os resultados obtidos pelas Abordagens 2, 3 e 4. Esta separação se dá por a primeira abordagem utilizar de aleatoriedade e portanto precisar de uma forma diferente de análise.

### 6.1 Abordagem 1

Esta abordagem possui uma variável aleatória na criação do conjunto de referência de não-eventos,  $\mathcal{R}_-$ . Dada uma observação da série temporal, caso ela não seja um exemplo de evento, ela possui uma probabilidade de ser inserida em  $\mathcal{R}_-$ . Esta probabilidade faz com que duas execuções do algoritmo, utilizando os mesmo parâmetros de entrada, possam apresentar resultados diferentes. Sendo assim, para uma mesma configuração de valores das variáveis de entrada foram executados vários testes.

Ao testar todas as configurações possíveis de variáveis, ordenamos os resultados em ordem decrescente de acordo com o retorno financeiro. Desta lista, escolhemos sete configurações para reexecução e verificação da confiabilidade dos resultados. As configurações selecionadas foram as que apresentaram os três maiores retornos (nomeadas respectivamente de A, B e C nas Figuras 5 e 6), duas das configurações que tiveram retorno igual a -100% (D e E), escolhidas aleatoriamente, e duas das configurações que tiveram retorno igual a -50% (F e G), também escolhidas aleatoriamente.

Para cada configuração, foram feitas mais mil execuções do algoritmo, utilizando os mesmos parâmetros de entrada e definição de tendência. Esta é a fase de validação. A Figura 5a mostra a distribuição dos resultados destas mil execuções. A Figura 5b traz os mesmos resultados, de forma ampliada, exibindo apenas resultados com retorno menor que mil, para facilitar a visualização.

Percebemos retornos muito altos para as primeiras três configurações. A configuração A, em especial, tem como mediana um retorno de aproximadamente 300% do capital inicial, e como resultado mais alto um retorno maior que 6.000%.

As mesmas configurações foram utilizadas na fase de testes, com um conjunto de

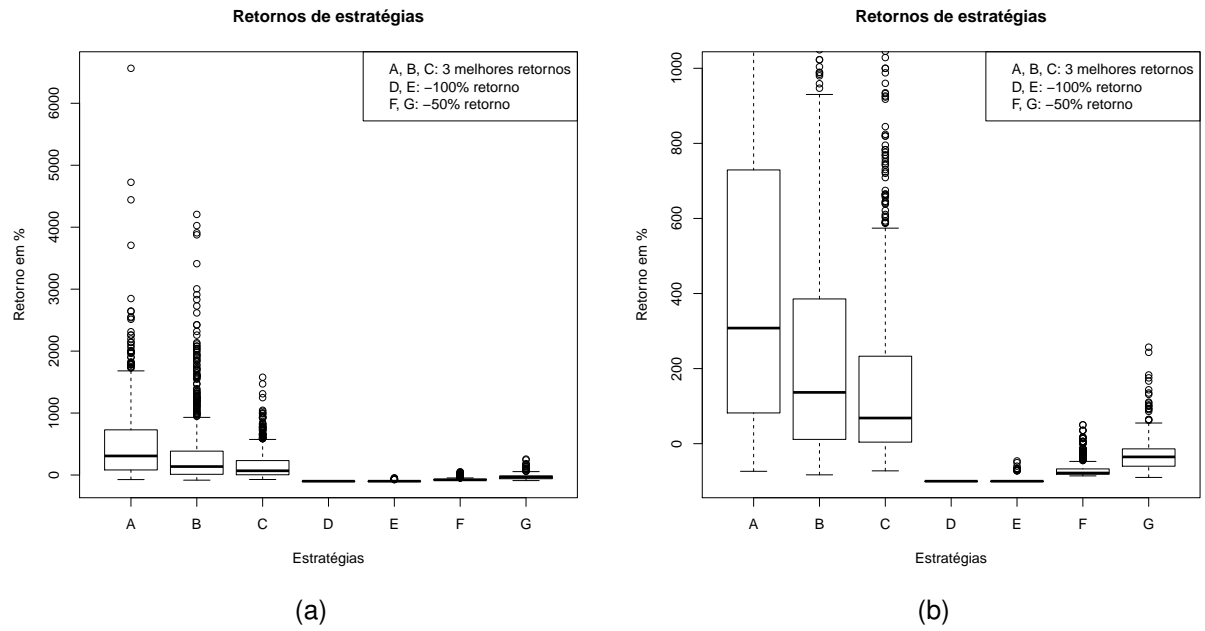


Figura 5: (a) Distribuição dos resultados de mil execuções de sete configurações da Abordagem 1, na fase de validação. (b) Ampliação do mesmo gráfico.

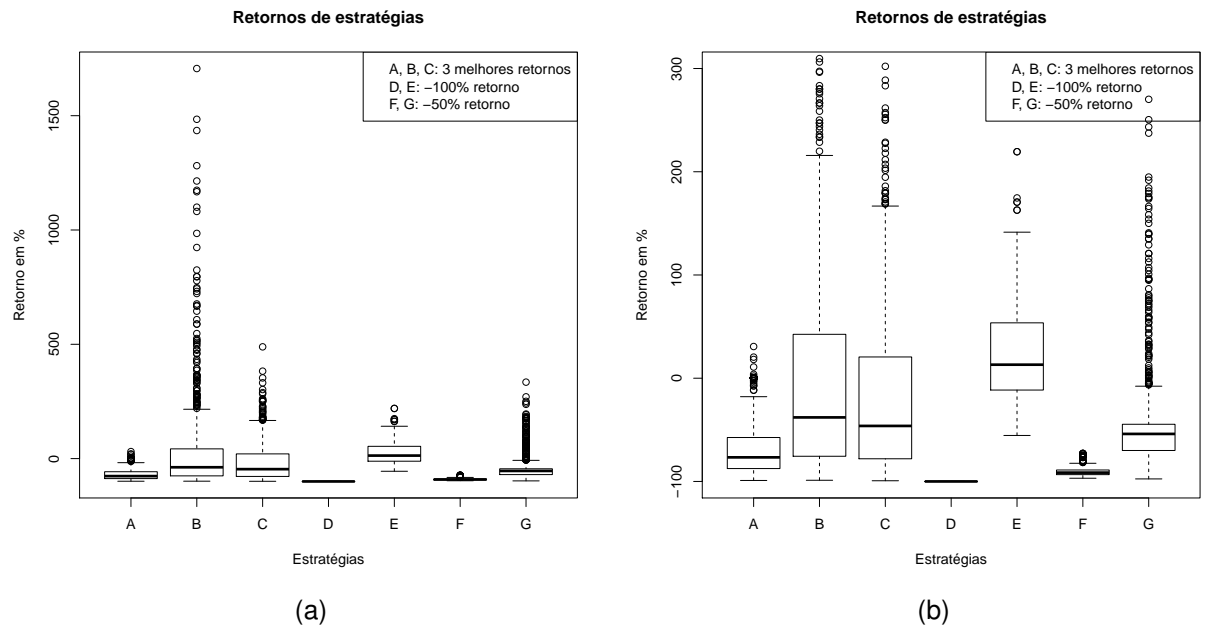


Figura 6: (a) Distribuição dos resultados de mil execuções de sete configurações da Abordagem 1, na fase de testes. (b) Ampliação do mesmo gráfico.



dados diferente do utilizado na validação. Novamente foram realizadas mil execuções para cada configuração. A distribuição dos resultados é apresentada na Figura 6a, e a Figura 6b traz uma ampliação dos resultados, excluindo da exibição os maiores que trezentos.

Desta vez temos resultados bem diferentes dos vistos na fase de validação. As três melhores estratégias (A, B e C) agora apresentam mediana abaixo de zero. Isso significa que na maioria das vezes estas estratégias trariam prejuízo ao operador. A estratégia D permaneceu com o pior resultado, e a estratégia G mostrou uma leve piora, tendo agora todo o corpo de seu gráfico abaixo da linha zero, na zona de prejuízo. A única estratégia que apresentou melhora foi a E, tendo sua mediana acima da linha zero. Ela era a segunda pior na validação e agora é a de melhor resultado na fase de teste.

## 6.2 Abordagens 2, 3 e 4

Estas abordagens não apresentam mais aleatoriedade no algoritmo. As tendências são definidas pelos cruzamentos das médias móveis e seus resultados são facilmente replicados. Assim não necessitamos de várias execuções para cada configuração de variáveis.

### 6.2.1 Validação

Na fase de validação, executamos o algoritmo uma vez para cada configuração de variável. Criamos boxplots com os resultados das execuções, separados por variável de entrada e por ativo. Criamos também tabelas com as médias, medianas e desvios padrão das variáveis, separadas também por ativo. No Apêndice A estão as tabelas da Abordagem 2 e os boxplot desta abordagem estão no Apêndice B. Da Abordagem 3, estão no Apêndice C e Apêndice D. Da Abordagem 4, estão no Apêndice E e Apêndice F.

Em algumas execuções, o algoritmo não identificou nenhuma tendência, e não foi realizada nenhuma operação de compra ou venda. O resultado final nestes casos ficou igual a 100% do capital inicial, recebendo portanto o valor 1 no boxplot. Para facilitar a visualização, os valores destas execuções onde não houveram operações foram removidos das figuras. Ou seja, nas figuras dos apêndices estão somente os resultados das configurações que resultaram em pelo menos uma operação de compra ou venda.

Foi feito também um boxplot de uma carteira teórica contendo os 10 ativos. Separamos os retornos por variável. Assim podemos ver a influência de cada variável no retorno final da carteira.

A Figura 7 mostra os boxplots da Abordagem 2. Em geral, a alteração dos valores

das variáveis não exerce muita influência no retorno da carteira. As duas exceções são as variáveis  $\theta$  e  $N_{smooth}$ . A primeira mostra um melhor retorno para seu maior valor, 3. Já a segunda variável mostra valores pouco melhores para seu menor valor, 1. Nas outras variáveis percebe-se alterações apenas nos valores *outliers*.

As Abordagens 3 e 4 apresentam resultados semelhantes entre si, como visto nas Figuras 8 e 9, respectivamente. Apenas duas variáveis parecem não influenciar muito nos resultados,  $\theta$  e  $detectionsLimit$ . As variáveis  $\gamma$  e  $N_{smooth}$  mostram retornos mais elevados para os valores menores, 0.1 para  $\gamma$ , e 1 e 5 para  $N_{smooth}$ . A variável  $N_{obs}$  melhora o retorno ao passo que aumenta seu valor, até o maior deles, 80. A variável  $N_{ref}$  é a única que diverge entre as duas abordagens. Enquanto na Abordagem 4 a melhor figura é vista com leve vantagem no valor 80, na Abordagem 3 é mais clara a vantagem com o menor valor, 10.

### 6.2.2 Teste

A configuração com o melhor retorno de cada ativo foi escolhida para a fase de teste. Nesta fase, o algoritmo foi executado apenas uma vez por ativo, com dados desconhecidos pela aplicação. Os resultados encontram-se na Tabela 3.

Ativo	A2	A3	A4	MM
BRFS3	0.8026	0.9077	1.017	0.8046
CPFE3	0.7016	1.0963	0.7691	0.9466
CYRE3	0.5321	0.5361	0.8565	1.0636
EMBR3	1.2007	1.0312	1.0938	0.8918
ITUB4	0.8853	1.1403	1.1197	0.8068
LREN3	0.6719	0.6431	1.1323	1.2232
PETR4	0.8317	0.6774	1.0824	1.109
TIMP3	0.7028	1.022	1.0871	0.9601
TOTS3	0.9382	0.8987	1.1399	0.772
VALE5	1.7472	0.9023	0.9971	0.4153
Media	0.9014	0.8855	1.0295	0.8993
Desvio	0.3301	0.1935	0.1184	0.2121

Tabela 3: Retornos das Abordagens 2, 3 e 4 (A2, A3 e A4), do cruzamento das médias móveis (MM), média e desvio padrão de cada estratégia na fase de teste.

Tomando a estratégia de operação de cruzamento de médias móveis (MM) como parâmetro, podemos tecer comparações sobre retornos de uma carteira hipotética contendo os dez ativos listados.

A Abordagem 3 apresentou o menor resultado dentre as quatro técnicas, sendo a única das abordagens a apresentar resultado inferior às médias móveis. Esta, por sua vez, teve um retorno pouco inferior ao da Abordagem 2, a que apresentou o segundo melhor resultado. Já a Abordagem 4 foi a única técnica a apresentar retorno positivo, resultando em lucro sobre o investimento inicial. Foi esta a abordagem também que

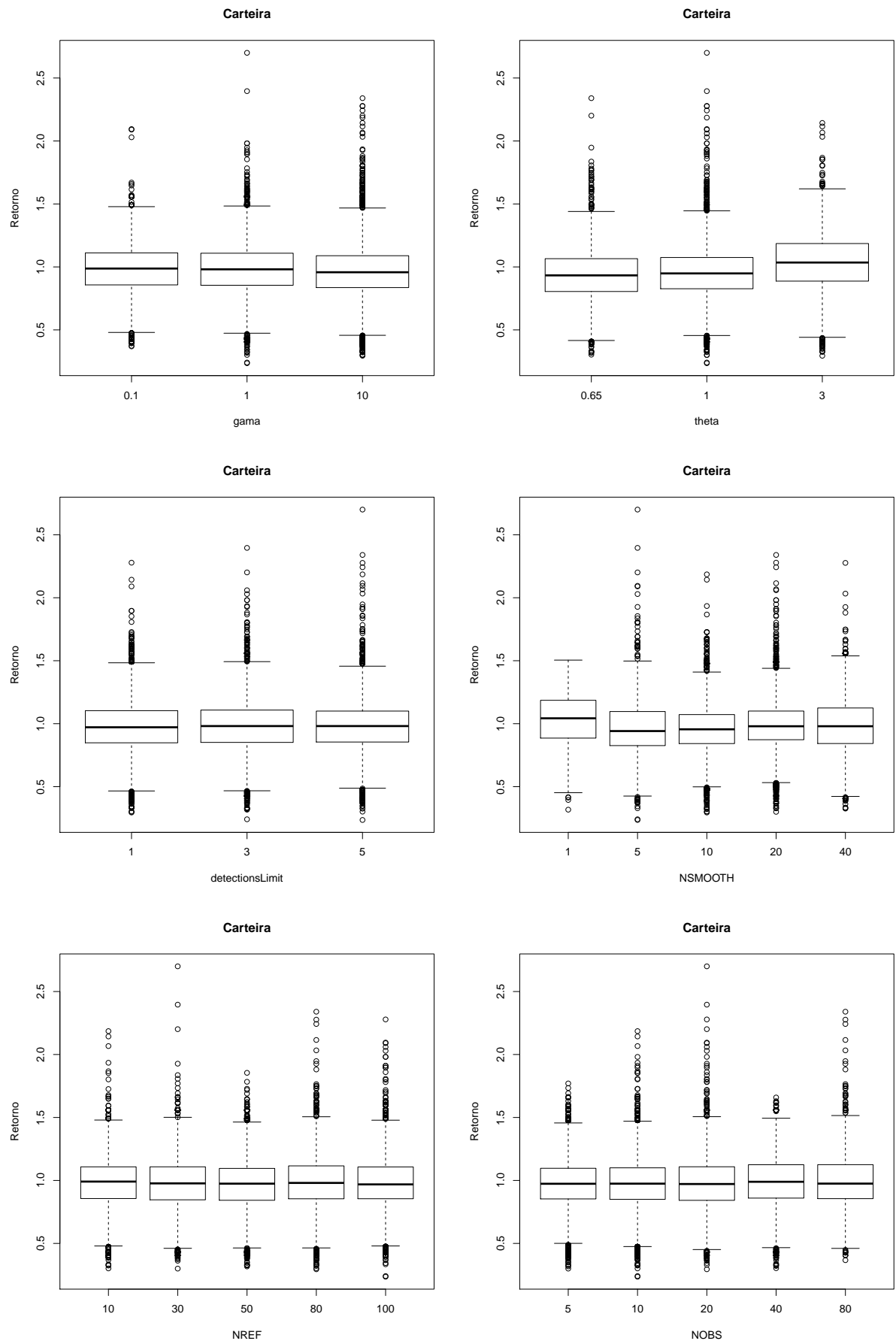


Figura 7: Boxplots dos retornos, separados por variável, da carteira teórica com os 10 ativos para a Abordagem 2.

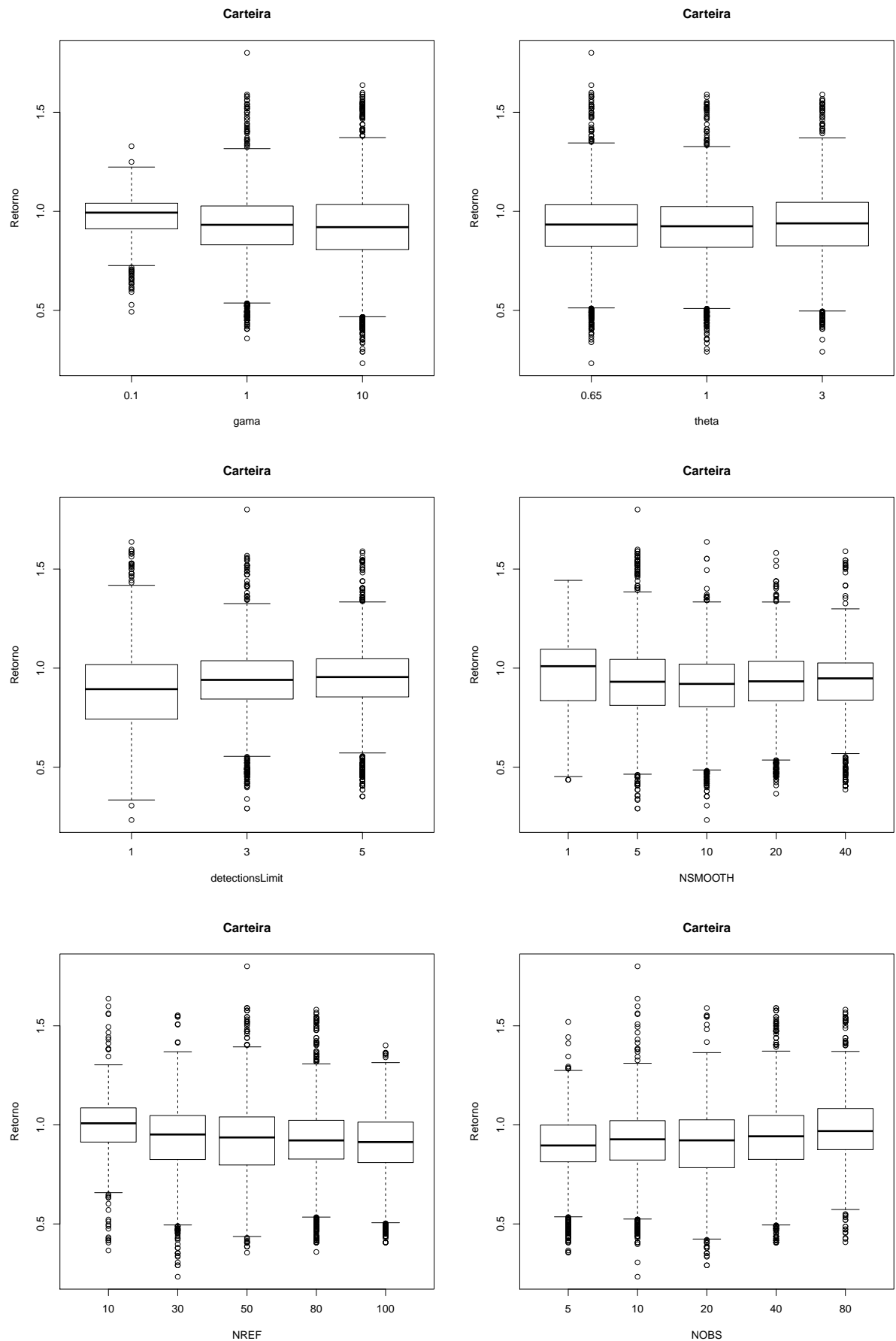


Figura 8: Boxplots dos retornos, separados por variável, da carteira teórica com os 10 ativos para a Abordagem 3.

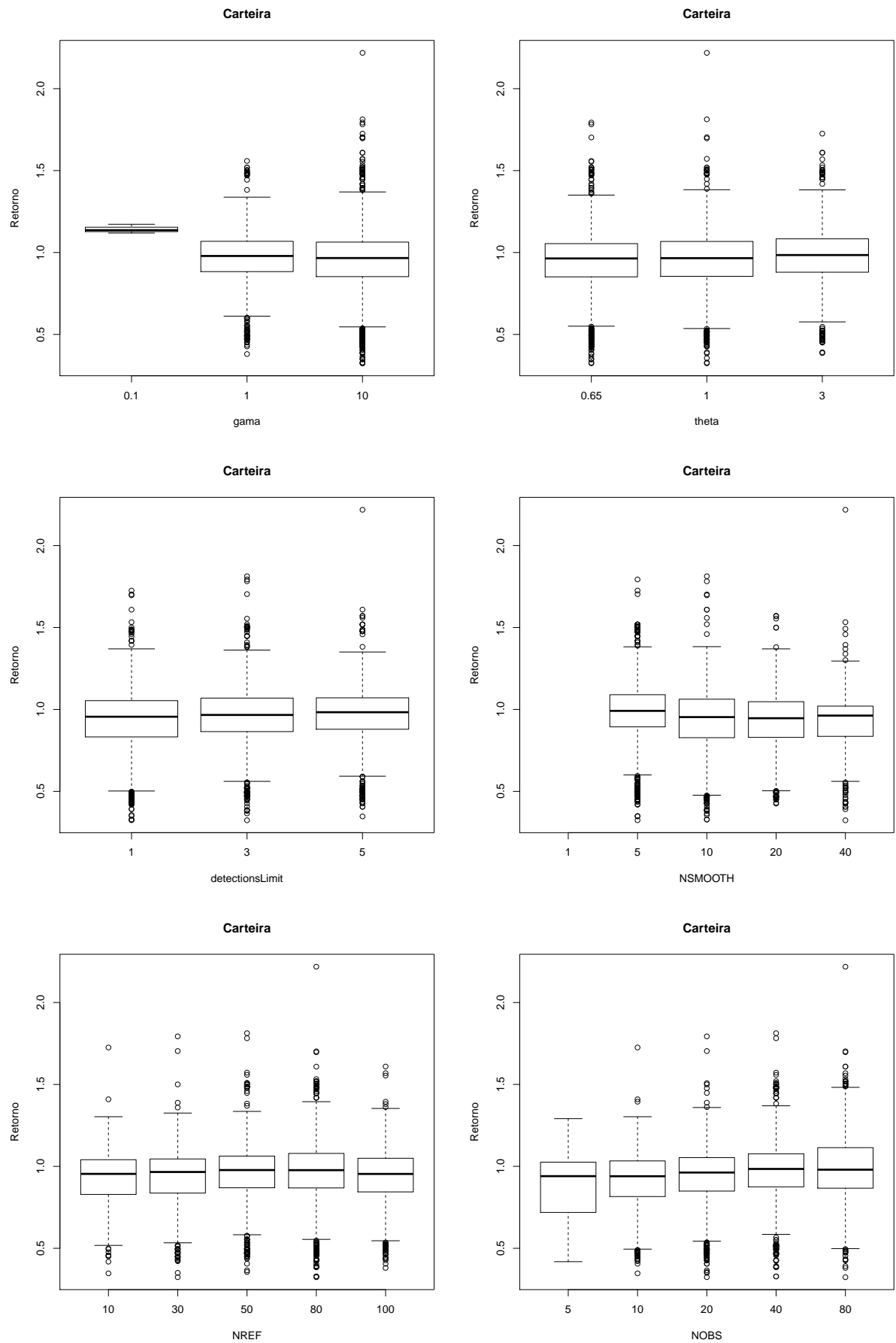


Figura 9: Boxplots dos retornos, separados por variável, da carteira teórica com os 10 ativos para a Abordagem 4.

obteve o menor desvio padrão, mostrando maior regularidade.

O maior retorno individual foi da o Abordagem 2 na VALE5, quase 75% de lucro no período. Este resultado foi o principal responsável por elevar a média desta técnica, uma vez que ela lucrou em apenas 2 das 10 ações da carteira.

Curiosamente o menor retorno individual foi obtido com o mesmo ativo, VALE5, na técnica das médias móveis, quase 59% de prejuízo. Isto demonstra, segundo as características desta técnica, que o período em questão apresentou poucas tendências favoráveis às médias móveis utilizadas.

A maior quantidade de retornos positivos pertence à Abordagem 4, onde 7 ativos dos 10 presentes na carteira foram lucrativos.

Foi realizado também um teste estatístico nos resultados da Tabela 3. O teste utilizado foi o ANOVA (*Analysis Of Variance*), que analisa a diferença entre as médias dos grupos. Para um grau de liberdade de 5% este teste teve como resultado os valores  $F=0,797$  e  $Pr(>F)=0,504$ , indicando que não há diferença estatística entre os valores.

## 7 DISCUSSÕES E TRABALHOS FUTUROS

Segundo os resultados apresentados no Capítulo 6, pudemos perceber a grande influência que os conjuntos de referências exercem sobre o resultado final. Na Abordagem 1, o uso de aleatoriedade na construção dos conjuntos acarretou em dificuldades de replicação dos resultados. De uma execução para outra havia muita diferença, o que trazia insegurança quanto a confiabilidade dos resultados.

Esta aleatoriedade foi retirada na Abordagem 2. Os resultados agora eram facilmente replicados. Entretanto a utilização do algoritmo ainda não estava sendo feita da forma com a qual foi planejado. Havia apenas dois conjuntos de referências, um para eventos e outro para não-eventos, e tentávamos identificar dois tipos de eventos: tendências de alta e tendências de baixa. Além disso, não verificávamos se uma indicação de tendência dada pelo cruzamento das médias móveis se confirmava ou não. Com isso utilizávamos também exemplos falsos nos conjuntos, prejudicando a identificação realizada pelo algoritmo.

Na Abordagem 3 resolvemos estes dois problemas inserindo um novo conjunto de referências. Os três conjuntos, individualmente, eram formados por exemplos de tendências de alta, exemplos de tendências de baixa, e não-exemplos – que neste caso eram falsas indicações de tendências informadas pelas médias móveis. Esta mudança trouxe melhorias, mas fez com que os exemplos fossem divididos em três conjuntos de referências, e não em dois. Isso diminuiu o número de exemplos por conjunto, tornando um pouco mais impreciso o processo de comparação de sinais e identificação de tendências.

A Abordagem 4, ao utilizar conjuntos de referências globais, aumenta o número de exemplos por conjunto. Cada observação agora seria comparada com mais exemplos, e de origens diferentes. Como vimos nos resultados, este aumento nos conjuntos de referências trouxe significativas melhorias nos resultados. Sendo esta a única alteração entre a Abordagem 3 e a Abordagem 4, percebemos a importância da qualidade dos conjuntos de referências ao constatarmos que ela trouxe a média dos retornos da pior para a melhor posição dentre as técnicas abordadas.

Sobre sua usabilidade, podemos concluir que o algoritmo de Nikolov pode ser uma

ferramenta útil em operações na bolsa de valores. A Abordagem 4 apresentou o melhor dos retornos dentre todas as técnicas testadas neste trabalho, sendo também o único retorno positivo. Seu resultado é 13% maior que a técnica de cruzamentos de médias móveis, se comparada a valorização do montante financeiro inicial.

Nos resultados individuais, por ação, esta abordagem apresentou maior número de retornos positivos. De 10 possíveis, 7 tiveram lucro, contra 3 das médias móveis. Dentre os retornos negativos, o seu pior resultado ficou ainda 35% acima do pior resultado apresentado pelas médias móveis. Ainda, o menor desvio padrão indica uma maior regularidade nos resultados. Isso pode resultar em maior sentimento de segurança para o operador, pois ele pode reduzir as aplicações com grandes perdas

Do ponto de vista de um investidor, ainda não se justifica o uso do algoritmo na forma que ele se encontra. O retorno obtido na simulação foi, na melhor abordagem, de 2,95% em 6 meses. O valor é pouco acima do rendimento de uma aplicação na nova poupança no mesmo período, 2,36%<sup>1</sup>. Se considerar abatimentos do imposto de renda, isento para poupança e incidente para operações em bolsa de valores, esta diferença fica menor ainda.

Entretanto, caso alguns ajustes sejam feitos e novos testes realizados, esta abordagem pode vir a ser uma opção rentável para o investidor. A própria técnica de cruzamento de médias móveis, por exemplo, utilizada como base no algoritmo, pode ter resultados melhorados se aplicada com valores mais adequados para as médias. BERNARDES (2015) realizou um trabalho de descoberta dinâmica dos melhores pares de médias móveis para um determinado ativo em um período especificado, obtendo resultados interessantes. A utilização de indicações de tendências de um trabalho como este poderá trazer melhores resultados que a utilização de um par fixo de médias móveis por um período de dez anos.

Existem ainda outras métricas que podem ser utilizadas pelo algoritmo. Porcentagem de indicações corretas, retorno médio por operação, prejuízo médio por operação, máximo *drawdown*<sup>2</sup> e volatilidade, para citar algumas. Pode-se treinar o algoritmo visando otimizar uma destas métricas, por exemplo, treinar para que se diminua a quantidade de operações com prejuízo. Inclusive combinar seu uso com outras ferramentas tradicionais da análise técnica, para entrar ou sair de uma operação. SACHETIM (2015), por exemplo, mostrou que algumas outras ferramentas de análise técnica realizam melhores identificações de tendências que as médias móveis.

É interessante também realizar outros experimentos com tamanhos de históricos diferentes. Procurar, para treinamento, validação e testes, tamanhos de históricos mais adequados. Um histórico menor que dez anos para treinamento, como utilizado neste trabalho, pode trazer melhores resultados. Outro possível trabalho pode ser

<sup>1</sup><https://www3.bcb.gov.br/CALCIDADAO/>

<sup>2</sup>Diferença percentual entre o máximo do retorno e um mínimo seguinte em um investimento.



utilizar o algoritmo na escala *intraday*. No caso de teste de seu trabalho, Nikolov utilizou período de dois minutos para seus dados. Tal período é possível de ser utilizado também com algumas ações, e podem representar com mais agilidade alguns comportamentos do mercado.

Pode-se também explorar mais o universo dos parâmetros de entrada do algoritmo. Aumentar ou diminuir a amplitude e granularidade dos valores conforme os resultados obtidos, a fim de buscar uma configuração mais adequada aos objetivos. Para isso pode ser feito uso de algoritmos de otimização, como por exemplo algoritmos genéticos.

O teste ANOVA indicou que não existe diferença significativa entre os resultados verificados neste trabalho. Isso pode ser reflexo do uso da pequena quantidade de ações analisadas (10 ativos). Um novo trabalho pode ser feito utilizando um número maior de ações para constatar se realmente existe ou não alguma diferença nos retornos encontrados.

## REFERÊNCIAS

- ADRIÃO, M. C. **Um Estudo de Caso de Previsão de Tendência em uma Série Temporal Financeira Utilizando Análise Técnica**. 2009. Dissertação de Mestrado — COPPE/UFRJ, Rio de Janeiro/RJ.
- ARTHUR, W. B.; HOLLAND, J. H.; LEBARON, B.; PALMER, R. G.; TAYLER, P. Asset pricing under endogenous expectations in an artificial stock market. **Available at SSRN 2252**, [S.I.], 1996.
- BERNARDES, S. d. C. **Uma Ferramenta de Suporte à Decisão para o Mercado de Ações via a Parametrização de Indicadores Técnicos**. 2015. Monografia de Graduação — Escola de Engenharia Elétrica, Mecânica e Computação/Universidade Federal de Goiás, Goiânia/GO.
- BRITO, O. S. de. **Mercado financeiro**. [S.I.]: Saraiva, 2005.
- CAMARGO, R. S. de. **Minority Games com Trocas de Informações Através de Estrutura de Redes Complexas**. 2006. Dissertação de Mestrado — Universidade Católica de Brasília, Brasília/DF.
- CHALLET, D.; ZHANG, Y.-C. Emergence of Cooperation and Organization in an Evolutionary Game. **Physica A: Statistical Mechanics and its Applications**, [S.I.], v.246, n.3, p.407–418, 1997.
- CORREIA, J. S. **Operando na Bolsa de Valores Utilizando Análise Técnica**. [S.I.]: Novatec Editora, 2008.
- ELDER, A. **Como se transformar em um operador e investidor de sucesso**. [S.I.]: Rio de Janeiro: Elsevier, 2004.
- ELY, R. A. Returns predictability and stock market efficiency in Brazil. **Brazilian Review of Finance**, [S.I.], v.9, n.4, p.571–584, 2011.
- FERREIRA, F. F.; FRANCISCO, G.; MACHADO, B. S.; MURUGANANDAM, P. Time Series Analysis for Minority Game Simulations of Financial Markets. **Physica A: Statistical Mechanics And Its Applications**, [S.I.], v.321, n.3, p.619–632, 2003.

GRILLI, L.; SFRECOLA, A. A Neural Networks Approach to Minority Game. **Neural Computing and Applications**, [S.l.], v.18, n.2, p.109–113, 2009.

JOHANN, F. J.; VIDAL, J. P.; TÁRTARO, M. **Utilização dos Indicadores Médias Móveis para a Análise Gráfica De Ações**. 2010. Dissertação (Mestrado em Ciência da Computação) — Faculdade Assis Gurgacz.

JÚNIOR, A. B. L.; RIGO, C. M.; CHEROBIM, A. P. M. S. **Administração financeira: princípios, fundamentos e práticas brasileiras**. [S.l.]: Campus, 2002.

MALHEIROS, R. S. **Operando com Trading Systems na Bolsa de Valores**. [S.l.]: Editora Évora, 2011.

NIKOLOV, S. **Trend or no trend: a novel nonparametric method for classifying time series**. 2012. Dissertação (Mestrado em Ciência da Computação) — M. Eng., Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

NORONHA, M. **Análise técnica: teorias, ferramentas e estratégias**. [S.l.]: Timing, 2009.

KAUFMAN, P. J. (Ed.). **Design, Testing, and Optimization of Trading Systems**. New York, NY, USA: John Wiley & Sons, Inc., 1992.

PINHEIRO, J. L. **Mercado de capitais: fundamentos e técnicas**. [S.l.]: Atlas, 2005.

RHEA, R. **The Dow theory: An explanation of its development and an attempt to define its usefulness as an aid in speculation**. [S.l.]: Fraser Publishing Company, 1993.

RIBEIRO, F. L. **Pogramação Genética, Redes Neurais e o Jogo da Minoria**. 2005. Dissertação de Mestrado — Instituto de Física/Universidade de São Paulo, São Paulo/SP.

ROBERTS, H. V. Stock-Market ?Patterns? And Financial Analysis: Methodological Suggestions. **The Journal of Finance**, [S.l.], v.14, n.1, p.1–10, 1959.

SACHETIM, H. M. **Análise técnica: estudo da confiabilidade dos principais indicadores de análise técnica, aplicados as ações mais negociadas na Bovespa no período de 1995 a 2005**. 2015. Dissertação de Mestrado — Universidade Federal do Paraná, Curitiba/PR.

SHAN, K. The Dow Theory. **Aweshkar Research Journal**, [S.l.], v.13, n.1, 2012.

SHARPE, W. F.; ALEXANDER, G. J.; BAILEY, J. V. **Investments**. [S.l.]: Prentice-Hall Upper Saddle River, NJ, 1999. v.6.

## APÊNDICE A TABELAS DA VALIDAÇÃO DA ABORDAGEM 2

### A.1 gama

BRFS3			
gama	Media	Mediana	Desvio
0.1	0.979	0.979	0.1376
1	0.9539	0.9539	0.142
10	0.9285	0.9285	0.1399
CYRE3			
gama	Media	Mediana	Desvio
0.1	1.0044	1.0044	0.1835
1	1.0192	1.0192	0.1917
10	1.0299	1.0299	0.1993
ITUB4			
gama	Media	Mediana	Desvio
0.1	0.9508	0.9508	0.1063
1	0.9405	0.9405	0.159
10	0.9209	0.9209	0.1826
PETR4			
gama	Media	Mediana	Desvio
0.1	0.9812	0.9812	0.4843
1	1.0469	1.0469	0.4626
10	0.9986	0.9986	0.4577
TOTS3			
gama	Media	Mediana	Desvio
0.1	1.0065	1.0065	0.1054
1	1.0089	1.0089	0.1138
10	1.0049	1.0049	0.1238

CPFE3			
gama	Media	Mediana	Desvio
0.1	0.9788	0.9788	0.1001
1	0.9816	0.9816	0.1239
10	0.9717	0.9717	0.1324
EMBR3			
gama	Media	Mediana	Desvio
0.1	0.9778	0.9778	0.1606
1	0.948	0.948	0.1503
10	0.9246	0.9246	0.1446
LREN3			
gama	Media	Mediana	Desvio
0.1	0.9819	0.9819	0.089
1	0.9683	0.9683	0.0988
10	0.947	0.947	0.1163
TIMP3			
gama	Media	Mediana	Desvio
0.1	1.0036	1.0036	0.0902
1	0.9777	0.9777	0.1377
10	0.9592	0.9592	0.1918
VALE5			
gama	Media	Mediana	Desvio
0.1	1.0155	1.0155	0.2647
1	1.0251	1.0251	0.256
10	1.0378	1.0378	0.2534

## A.2 theta

BRFS3			
theta	Media	Mediana	Desvio
0.65	1.01	1.01	0.1431
1	0.9816	0.9816	0.1512
3	0.8698	0.8698	0.0748

CPFE3			
theta	Media	Mediana	Desvio
0.65	0.9351	0.9351	0.0949
1	0.948	0.948	0.1254
3	1.0491	1.0491	0.1021

CYRE3			
theta	Media	Mediana	Desvio
0.65	0.9012	0.9012	0.1713
1	0.9757	0.9757	0.1915
3	1.1767	1.1767	0.0618

EMBR3			
theta	Media	Mediana	Desvio
0.65	1.0212	1.0212	0.1556
1	0.9778	0.9778	0.1589
3	0.8514	0.8514	0.0754

ITUB4			
theta	Media	Mediana	Desvio
0.65	0.9353	0.9353	0.146
1	0.9078	0.9078	0.1759
3	0.9691	0.9691	0.1275

LREN3			
theta	Media	Mediana	Desvio
0.65	0.9903	0.9903	0.1017
1	0.9731	0.9731	0.1256
3	0.9339	0.9339	0.0636

PETR4			
theta	Media	Mediana	Desvio
0.65	0.8134	0.8134	0.4317
1	0.8903	0.8903	0.4585
3	1.3231	1.3231	0.3355

TIMP3			
theta	Media	Mediana	Desvio
0.65	0.9644	0.9644	0.1393
1	0.9624	0.9624	0.1873
3	1.0137	1.0137	0.0933

TOTS3			
theta	Media	Mediana	Desvio
0.65	0.95	0.95	0.1038
1	0.9987	0.9987	0.1203
3	1.0717	1.0717	0.0814

VALE5			
theta	Media	Mediana	Desvio
0.65	0.8598	0.8598	0.2304
1	0.9623	0.9623	0.2302
3	1.2562	1.2562	0.096

### A.3 detectionsLimit

BRFS3			
dLimit	Media	Mediana	Desvio
1	0.9604	0.9604	0.1453
3	0.9533	0.9533	0.1423
5	0.9477	0.9477	0.1361

CPFE3			
dLimit	Media	Mediana	Desvio
1	0.9761	0.9761	0.1208
3	0.9753	0.9753	0.1213
5	0.9807	0.9807	0.1168

CYRE3			
dLimit	Media	Mediana	Desvio
1	1.007	1.007	0.1908
3	1.0214	1.0214	0.196
5	1.0252	1.0252	0.1883

EMBR3			
dLimit	Media	Mediana	Desvio
1	0.9539	0.9539	0.157
3	0.9499	0.9499	0.1521
5	0.9467	0.9467	0.1514

ITUB4			
dLimit	Media	Mediana	Desvio
1	0.9383	0.9383	0.1564
3	0.9403	0.9403	0.1565
5	0.9336	0.9336	0.1462

LREN3			
dLimit	Media	Mediana	Desvio
1	0.9658	0.9658	0.106
3	0.9665	0.9665	0.105
5	0.9649	0.9649	0.0978

PETR4			
dLimit	Media	Mediana	Desvio
1	0.9949	0.9949	0.458
3	1.011	1.011	0.4695
5	1.0208	1.0208	0.4794

TIMP3			
dLimit	Media	Mediana	Desvio
1	0.9783	0.9783	0.1495
3	0.9726	0.9726	0.1519
5	0.9896	0.9896	0.139

TOTS3			
dLimit	Media	Mediana	Desvio
1	0.9987	0.9987	0.1096
3	1.01	1.01	0.1134
5	1.0116	1.0116	0.1202

VALE5			
dLimit	Media	Mediana	Desvio
1	1.0044	1.0044	0.2668
3	1.0272	1.0272	0.254
5	1.0469	1.0469	0.2518

## A.4 NSMOOTH

BRFS3			
smooth	Media	Mediana	Desvio
1	0.8127	0.8127	0.0627
5	0.9702	0.9702	0.1586
10	0.9727	0.9727	0.1472
20	1.012	1.012	0.1299
40	1.0014	1.0014	0.0784

CPFE3			
smooth	Media	Mediana	Desvio
1	1.032	1.032	0
5	0.9324	0.9324	0.1304
10	0.9677	0.9677	0.1021
20	1.0031	1.0031	0.1134
40	0.9516	0.9516	0.1579

CYRE3			
smooth	Media	Mediana	Desvio
1	1.1857	1.1857	0
5	0.9614	0.9614	0.2103
10	0.9913	0.9913	0.2015
20	0.981	0.981	0.1638
40	0.9699	0.9699	0.1919

EMBR3			
smooth	Media	Mediana	Desvio
1	0.7941	0.7941	0
5	0.9535	0.9535	0.1572
10	0.9902	0.9902	0.1515
20	1.0086	1.0086	0.1463
40	1.0043	1.0043	0.1287

ITUB4			
smooth	Media	Mediana	Desvio
1	0.8905	0.8905	0.0661
5	0.9307	0.9307	0.1975
10	0.9347	0.9347	0.1528
20	0.9826	0.9826	0.1489
40	0.9484	0.9484	0.1551

LREN3			
smooth	Media	Mediana	Desvio
1	0.885	0.885	0.0556
5	0.9671	0.9671	0.1018
10	0.9755	0.9755	0.1217
20	1.0071	1.0071	0.1028
40	0.9939	0.9939	0.0714

PETR4			
smooth	Media	Mediana	Desvio
1	1.3828	1.3828	0.1533
5	0.9995	0.9995	0.4551
10	0.8265	0.8265	0.4383
20	0.9332	0.9332	0.4784
40	0.9026	0.9026	0.5092

TIMP3			
smooth	Media	Mediana	Desvio
1	1.0405	1.0405	0.0335
5	0.9503	0.9503	0.1426
10	0.9536	0.9536	0.183
20	0.9793	0.9793	0.1903
40	0.977	0.977	0.1088

TOTS3			
smooth	Media	Mediana	Desvio
1	1.0655	1.0655	0
5	1.005	1.005	0.1328
10	1.0031	1.0031	0.0945
20	0.9641	0.9641	0.1077
40	0.9963	0.9963	0.1487

VALE5			
smooth	Media	Mediana	Desvio
1	1.2588	1.2588	0
5	0.9616	0.9616	0.2509
10	0.9739	0.9739	0.2709
20	0.9662	0.9662	0.2668
40	0.9703	0.9703	0.2414

## A.5 NREF

BRFS3			
NREF	Media	Mediana	Desvio
10	0.9506	0.9506	0.1424
30	0.9595	0.9595	0.1462
50	0.96	0.96	0.141
80	0.9589	0.9589	0.1481
100	0.9416	0.9416	0.1304

CPFE3			
NREF	Media	Mediana	Desvio
10	0.9743	0.9743	0.1231
30	0.9727	0.9727	0.1245
50	0.9715	0.9715	0.1336
80	0.9953	0.9953	0.1218
100	0.9682	0.9682	0.0976

CYRE3			
NREF	Media	Mediana	Desvio
10	1.0166	1.0166	0.1759
30	0.9984	0.9984	0.19
50	1.0032	1.0032	0.1912
80	1.0477	1.0477	0.2064
100	1.012	1.012	0.1812

EMBR3			
NREF	Media	Mediana	Desvio
10	0.943	0.943	0.1433
30	0.9609	0.9609	0.1542
50	0.9539	0.9539	0.1514
80	0.9442	0.9442	0.1594
100	0.9495	0.9495	0.1527

ITUB4			
NREF	Media	Mediana	Desvio
10	0.9403	0.9403	0.1464
30	0.9405	0.9405	0.1325
50	0.9372	0.9372	0.1568
80	0.9368	0.9368	0.1495
100	0.9351	0.9351	0.1675

LREN3			
NREF	Media	Mediana	Desvio
10	0.9728	0.9728	0.1251
30	0.9641	0.9641	0.1125
50	0.9772	0.9772	0.0818
80	0.958	0.958	0.1072
100	0.9624	0.9624	0.0972

PETR4			
NREF	Media	Mediana	Desvio
10	0.9974	0.9974	0.4921
30	1.0369	1.0369	0.4704
50	0.9131	0.9131	0.4496
80	1.0349	1.0349	0.4669
100	1.0475	1.0475	0.4666

TIMP3			
NREF	Media	Mediana	Desvio
10	0.9921	0.9921	0.1422
30	0.9834	0.9834	0.1375
50	0.9753	0.9753	0.1573
80	0.9895	0.9895	0.1272
100	0.9679	0.9679	0.1629

TOTS3			
NREF	Media	Mediana	Desvio
10	1.0238	1.0238	0.1219
30	0.9859	0.9859	0.121
50	1.0101	1.0101	0.1125
80	1.0091	1.0091	0.116
100	1.0075	1.0075	0.1061

VALE5			
NREF	Media	Mediana	Desvio
10	1.0609	1.0609	0.2448
30	1.0121	1.0121	0.2637
50	0.9998	0.9998	0.2663
80	1.0224	1.0224	0.2596
100	1.0455	1.0455	0.2495



## A.6 NOBS

BRFS3			
NOBS	Media	Mediana	Desvio
5	0.9553	0.9553	0.1361
10	0.9564	0.9564	0.1402
20	0.9572	0.9572	0.1485
40	0.9524	0.9524	0.1491
80	0.939	0.939	0.1297

CPFE3			
NOBS	Media	Mediana	Desvio
5	0.9754	0.9754	0.1069
10	0.9796	0.9796	0.1177
20	0.9647	0.9647	0.127
40	0.9876	0.9876	0.1121
80	0.9865	0.9865	0.1456

CYRE3			
NOBS	Media	Mediana	Desvio
5	1.0023	1.0023	0.1862
10	1.0006	1.0006	0.184
20	0.9956	0.9956	0.1948
40	1.0374	1.0374	0.1804
80	1.1152	1.1152	0.2042

EMBR3			
NOBS	Media	Mediana	Desvio
5	0.9688	0.9688	0.154
10	0.9547	0.9547	0.1516
20	0.9338	0.9338	0.1476
40	0.9527	0.9527	0.1704
80	0.921	0.921	0.1338

ITUB4			
NOBS	Media	Mediana	Desvio
5	0.9466	0.9466	0.1393
10	0.9411	0.9411	0.1304
20	0.9354	0.9354	0.179
40	0.933	0.933	0.161
80	0.9158	0.9158	0.1684

LREN3			
NOBS	Media	Mediana	Desvio
5	0.9659	0.9659	0.1022
10	0.9697	0.9697	0.1013
20	0.9896	0.9896	0.1059
40	0.969	0.969	0.0934
80	0.9028	0.9028	0.0918

PETR4			
NOBS	Media	Mediana	Desvio
5	0.926	0.926	0.47
10	0.9672	0.9672	0.4859
20	1.0278	1.0278	0.4764
40	1.0714	1.0714	0.4256
80	1.189	1.189	0.4046

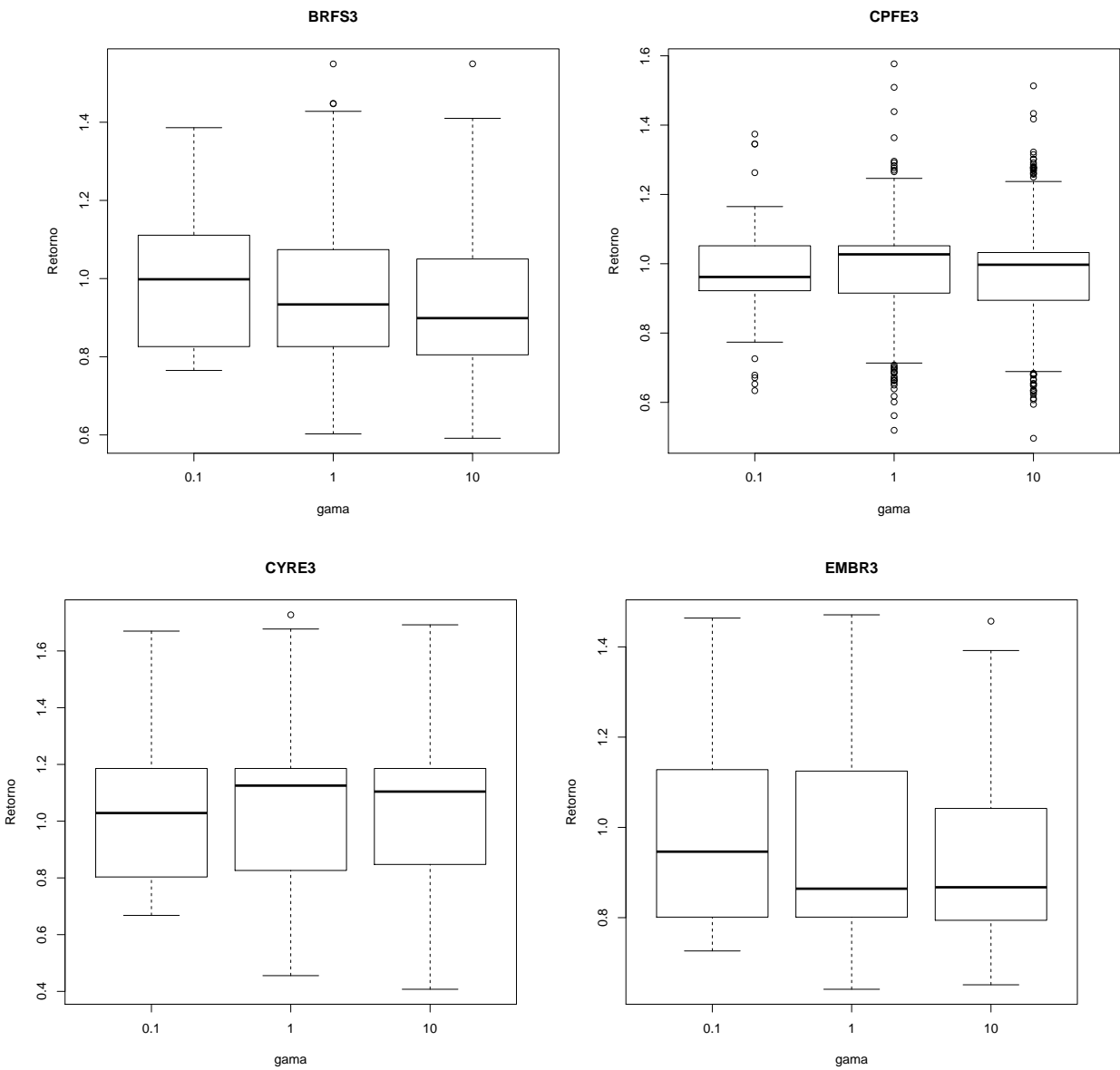
TIMP3			
NOBS	Media	Mediana	Desvio
5	0.9845	0.9845	0.1045
10	0.9913	0.9913	0.127
20	0.9949	0.9949	0.1716
40	0.9772	0.9772	0.1831
80	0.9163	0.9163	0.1551

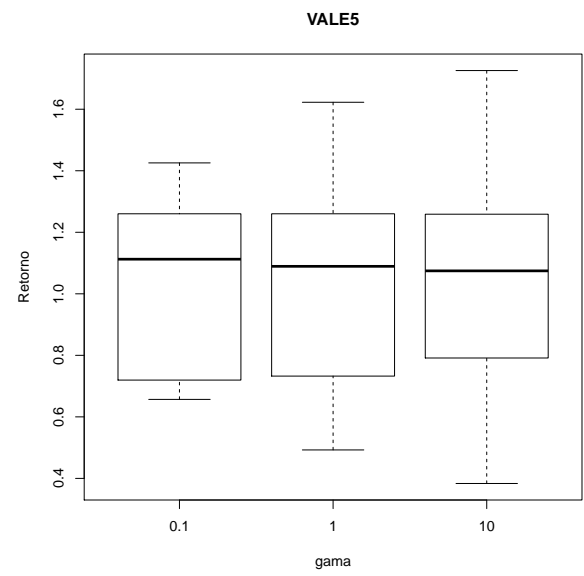
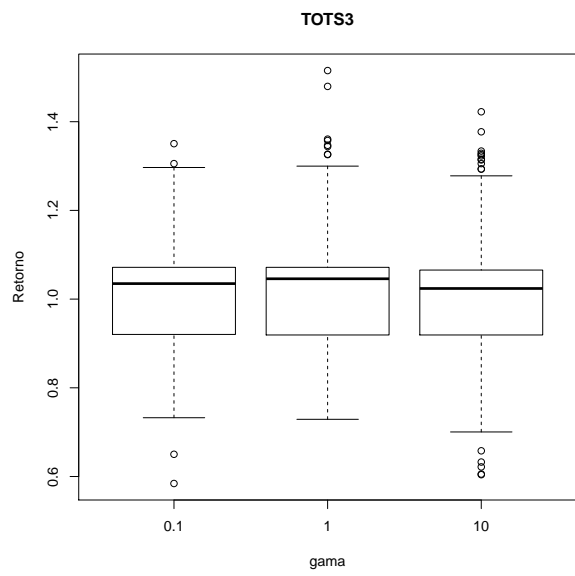
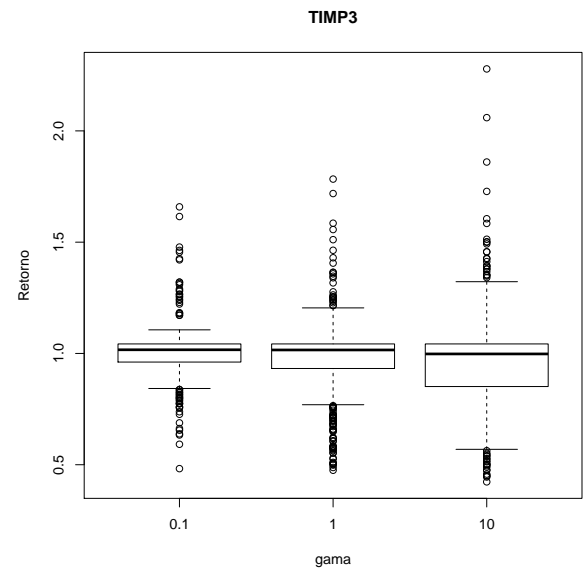
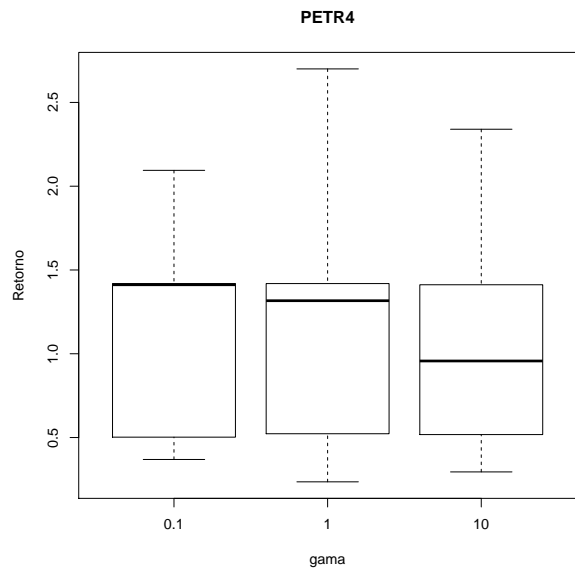
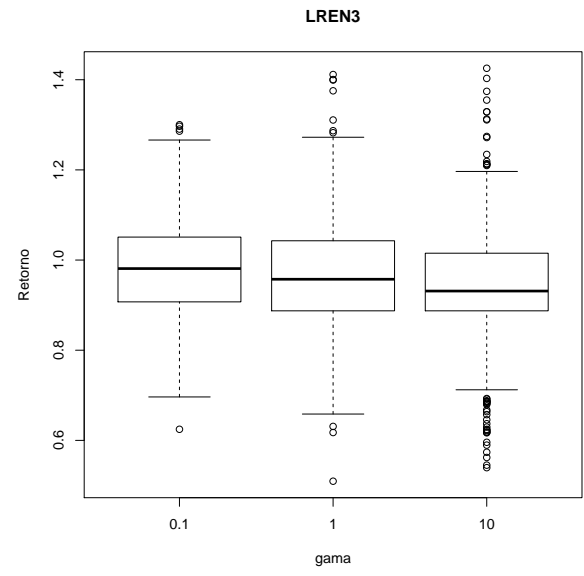
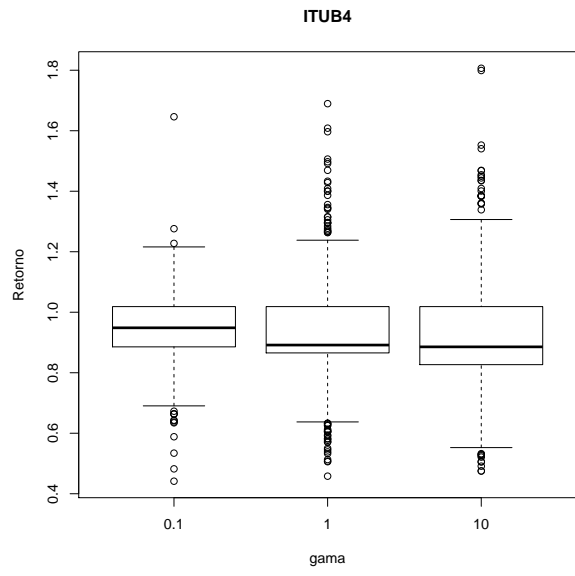
TOTS3			
NOBS	Media	Mediana	Desvio
5	0.9919	0.9919	0.1093
10	1.0007	1.0007	0.1077
20	0.9979	0.9979	0.1193
40	1.0156	1.0156	0.1097
80	1.0639	1.0639	0.1237

VALE5			
NOBS	Media	Mediana	Desvio
5	1.0134	1.0134	0.263
10	1.0108	1.0108	0.2711
20	1.0238	1.0238	0.2607
40	1.0529	1.0529	0.2387
80	1.0611	1.0611	0.2294

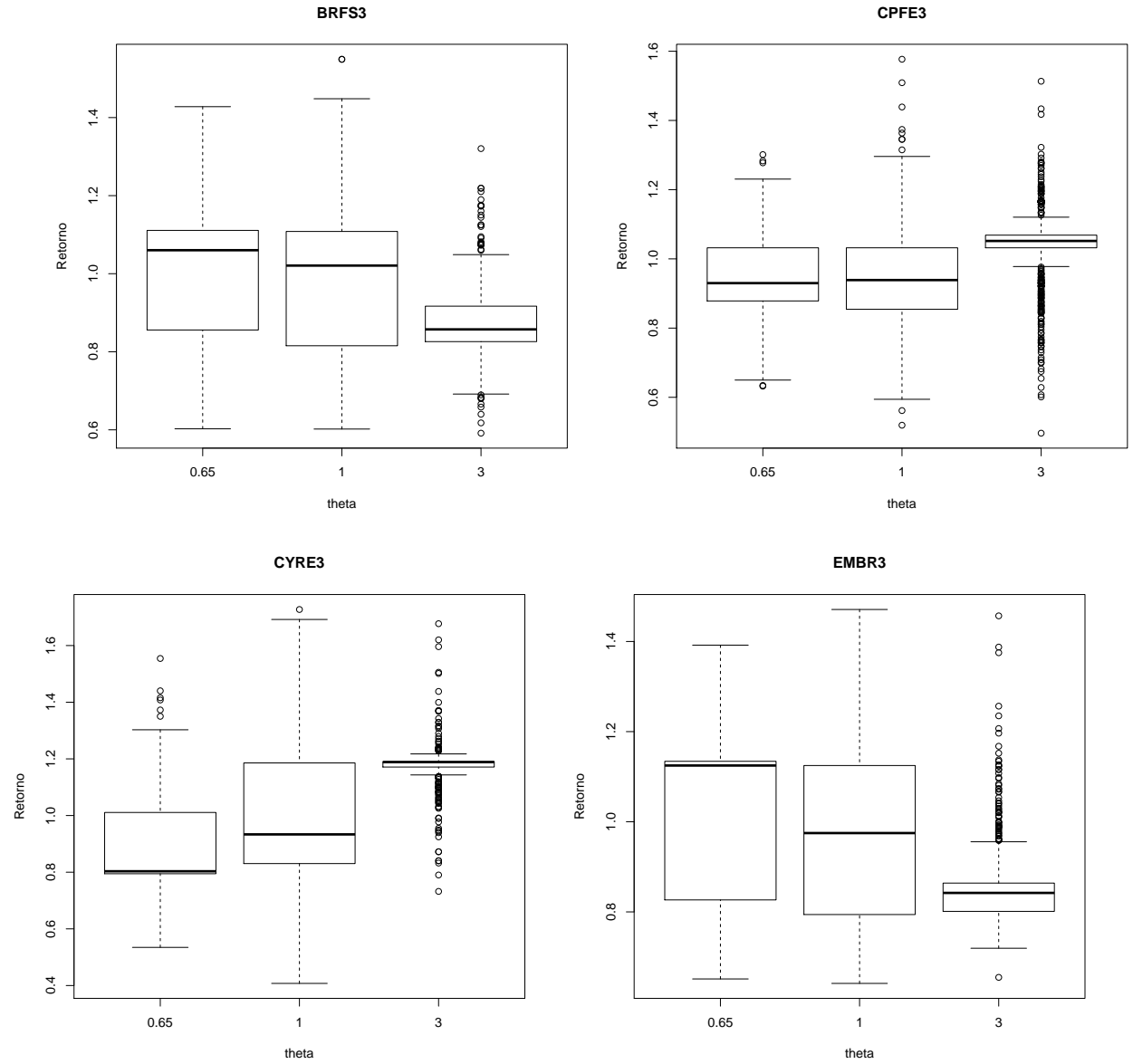
# APÊNDICE B BOXPLOTS DA VALIDAÇÃO DA ABORDAGEM 2

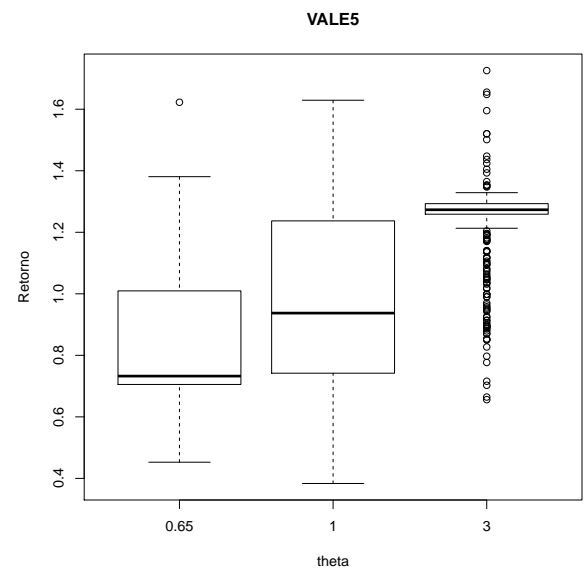
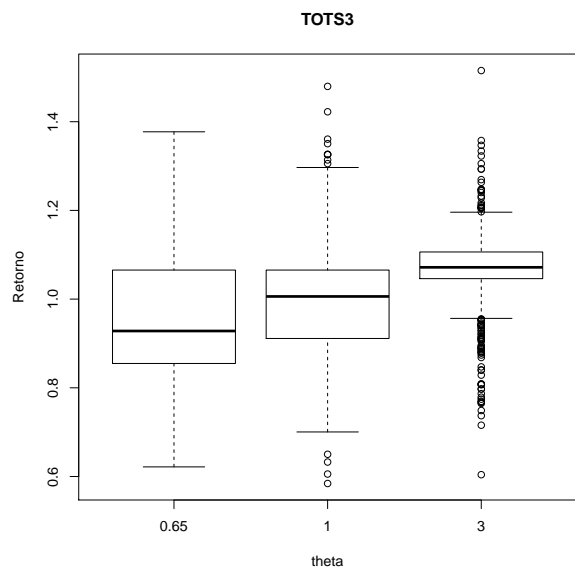
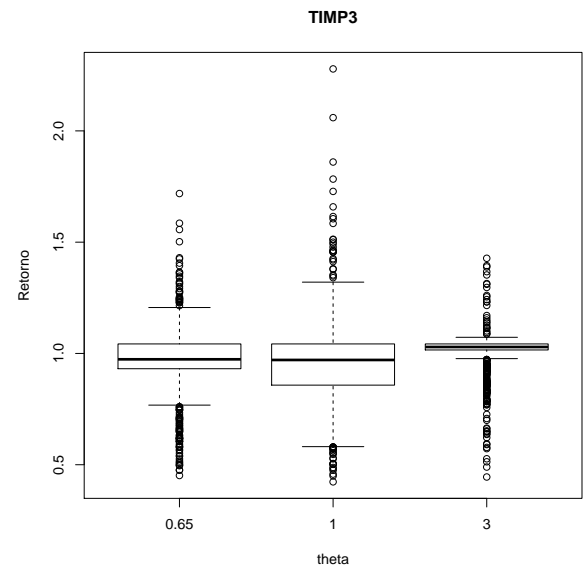
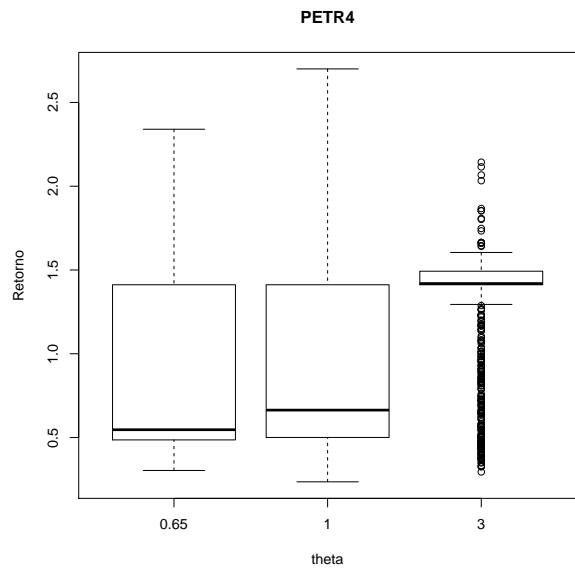
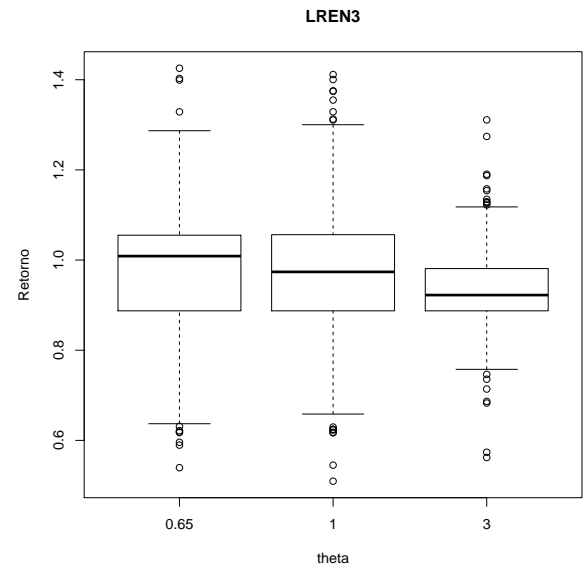
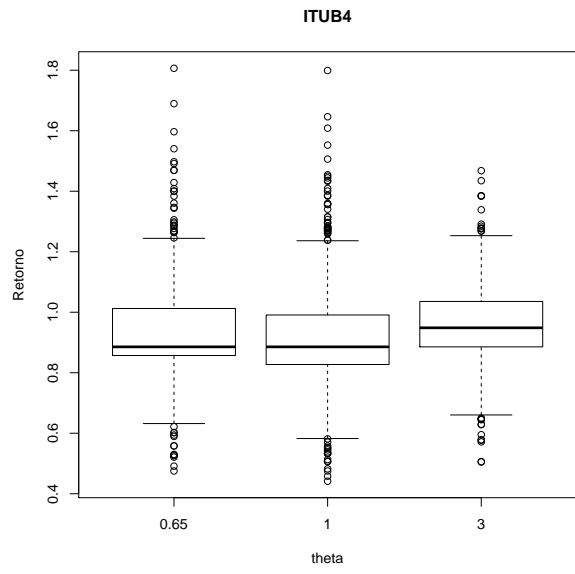
## B.1 gama



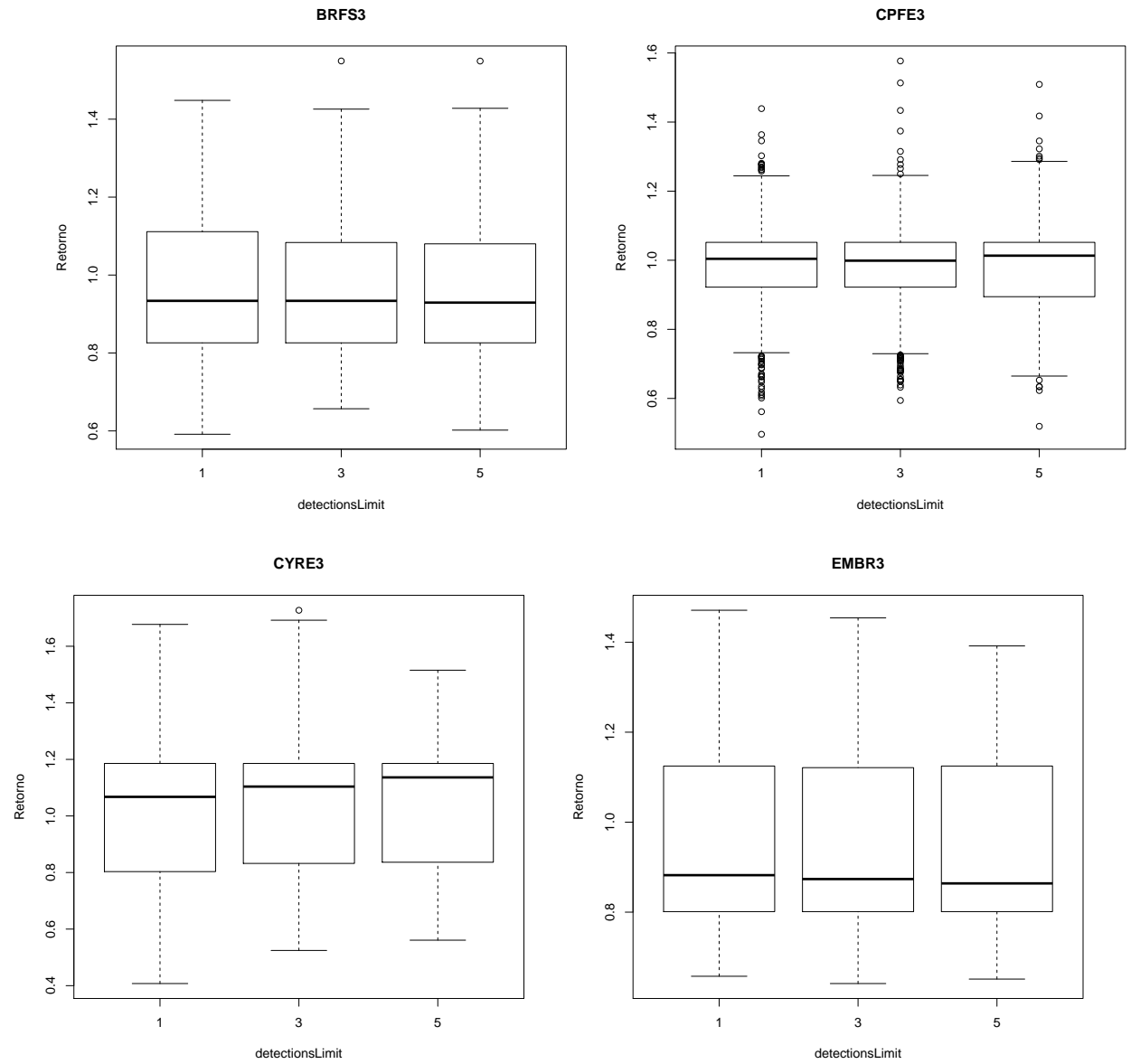


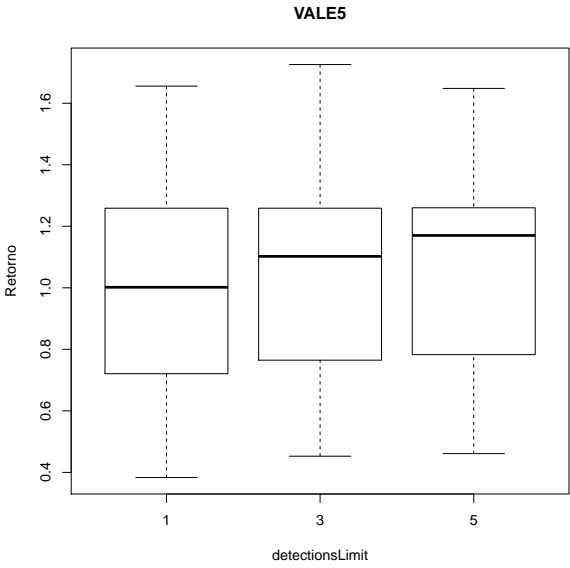
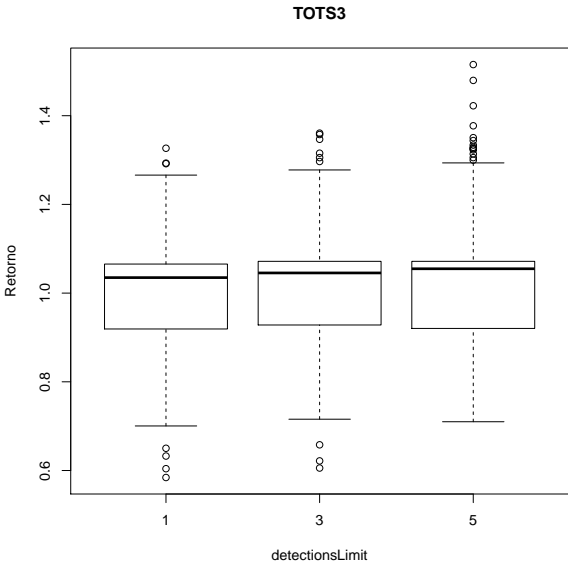
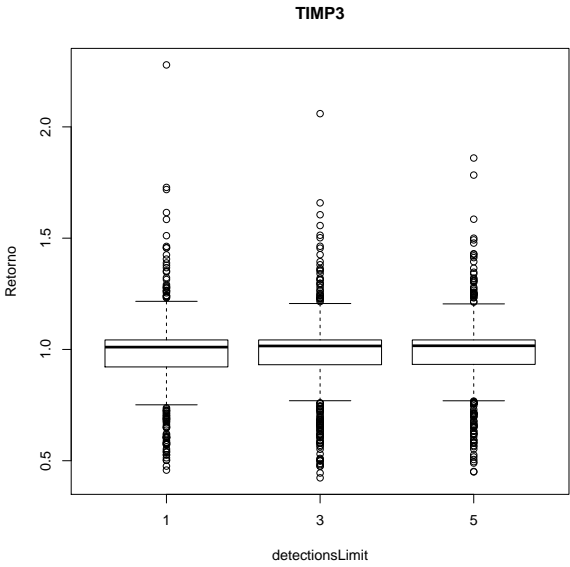
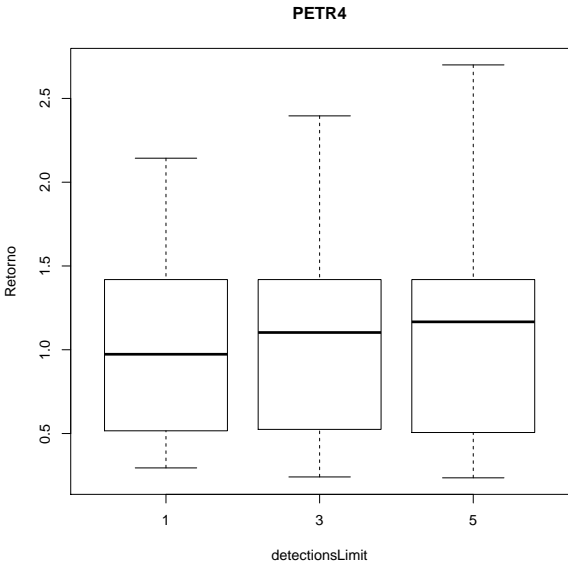
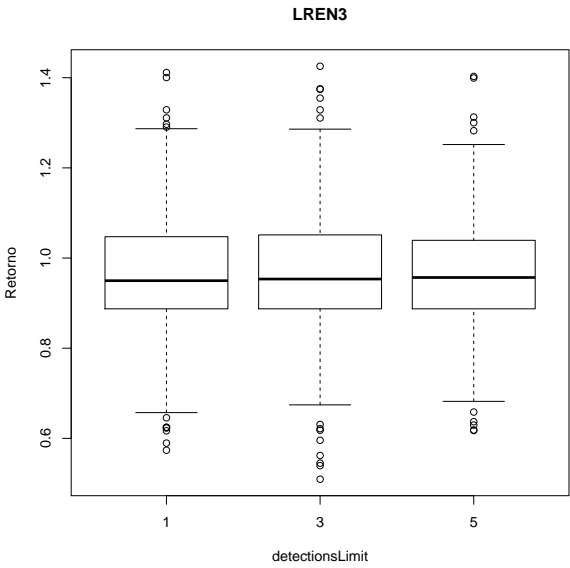
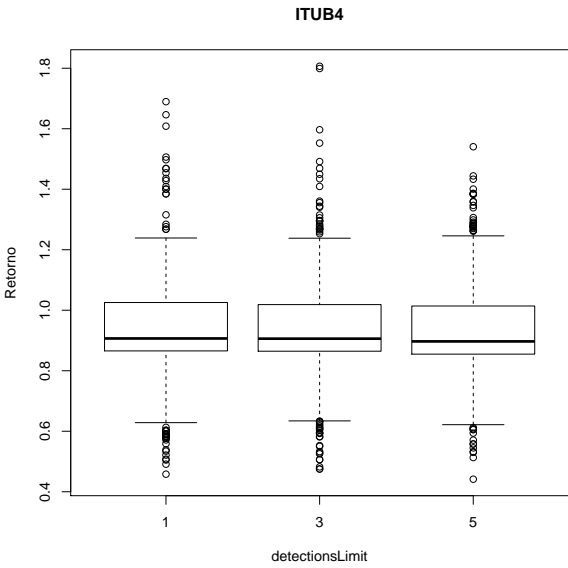
## B.2 theta



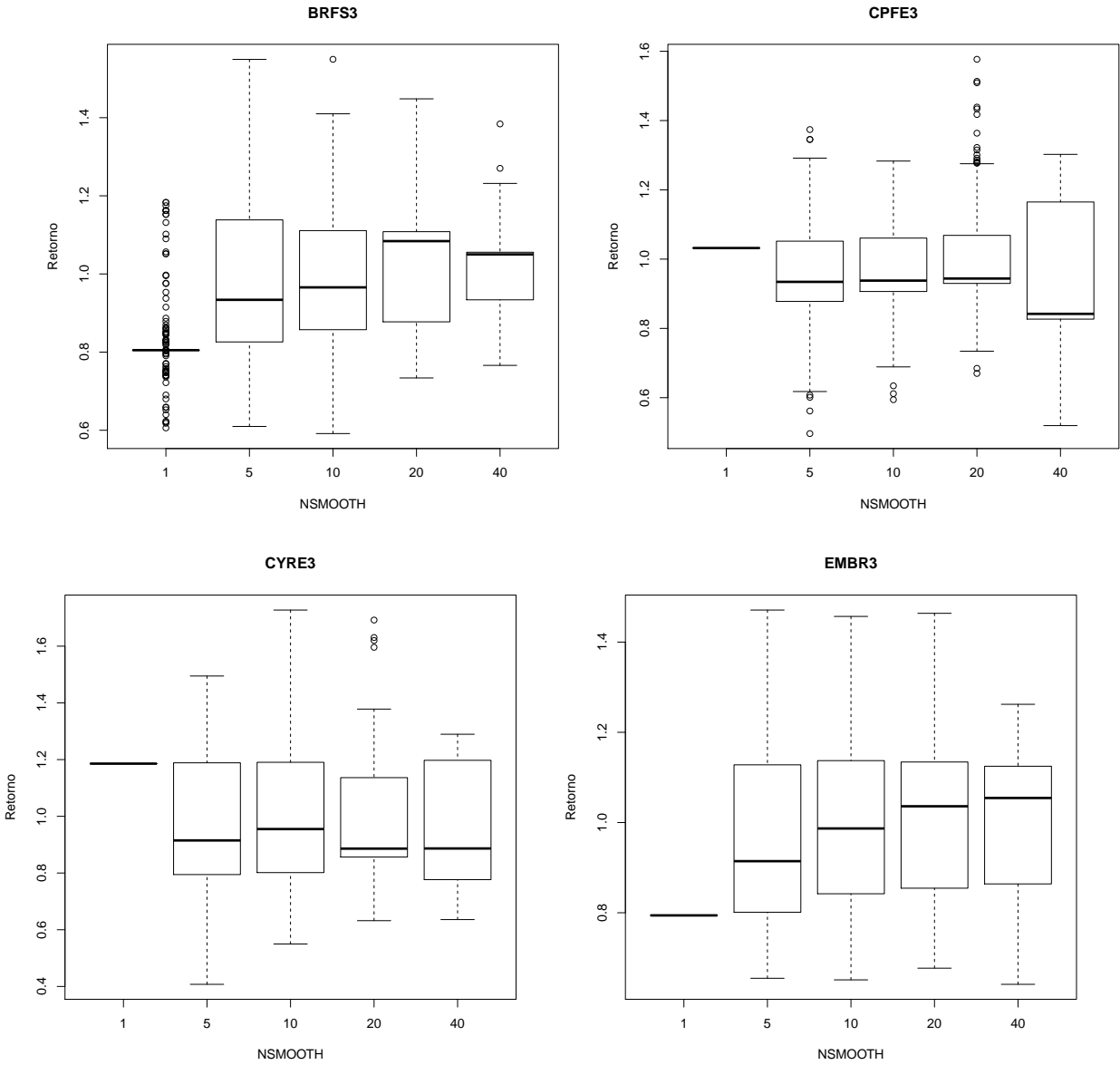


## B.3 detectionsLimit

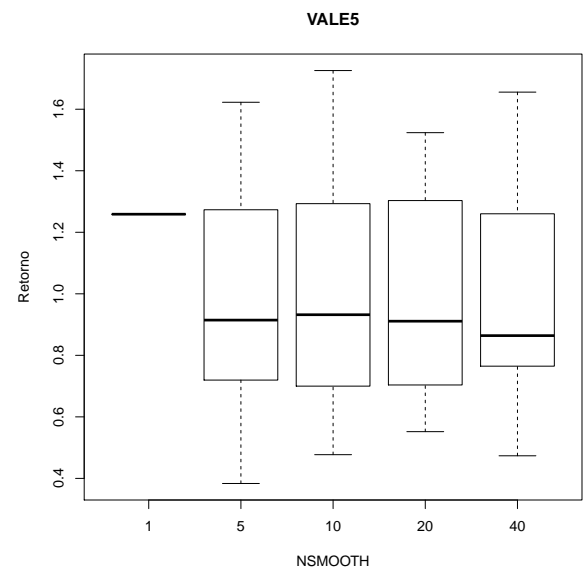
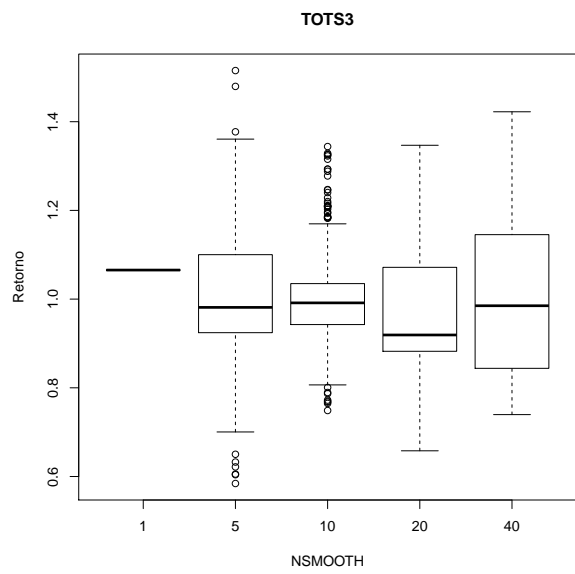
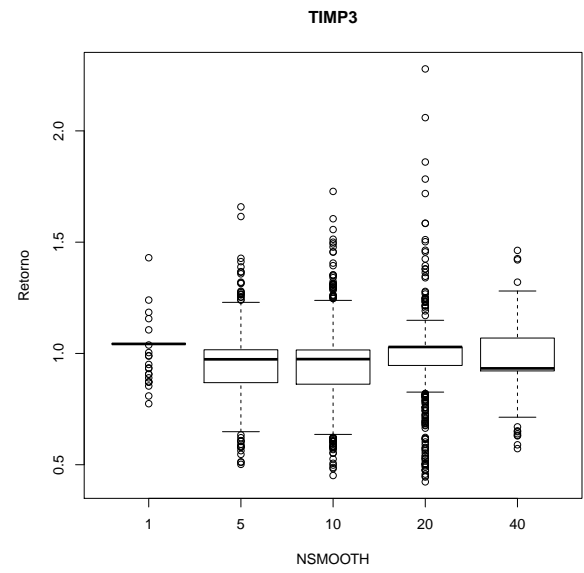
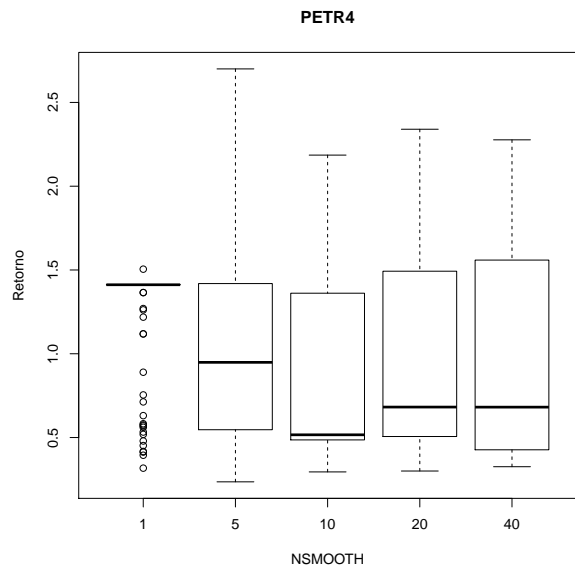
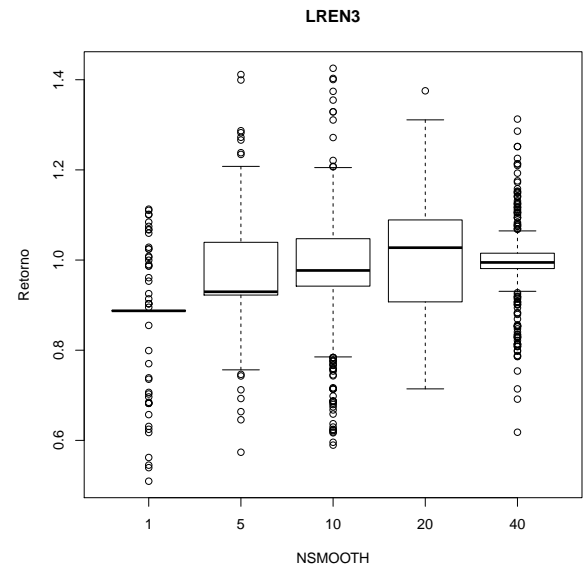
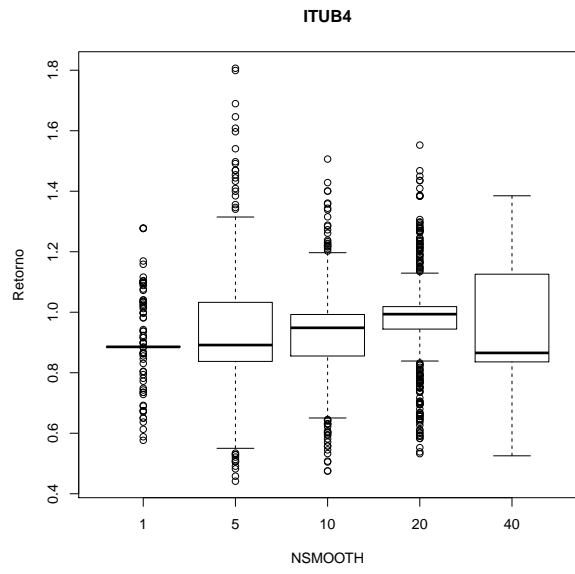




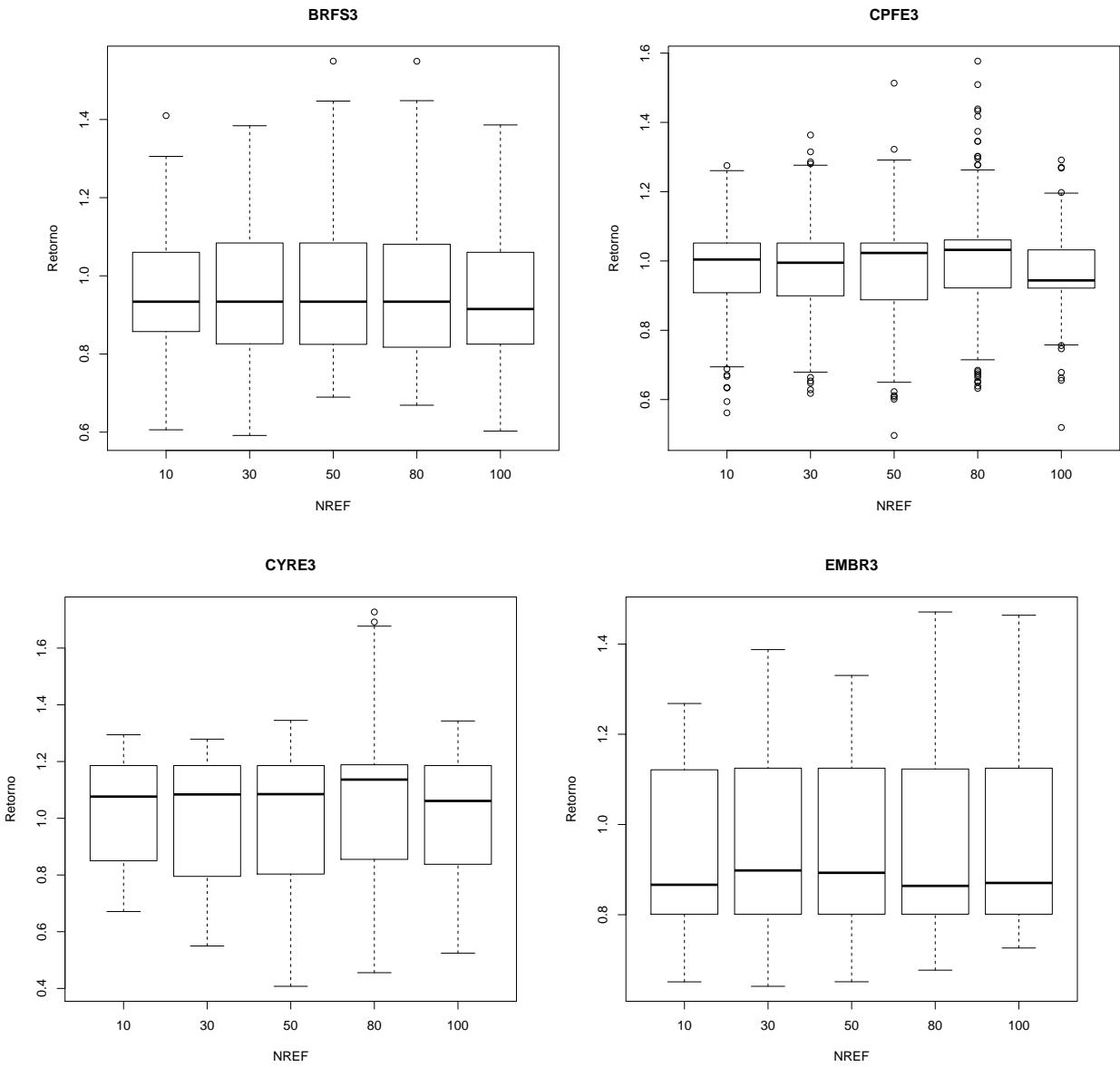
B.4 NSMOOTH

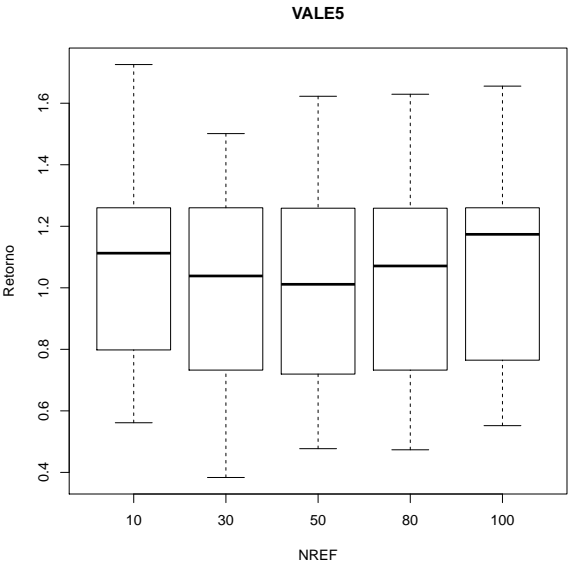
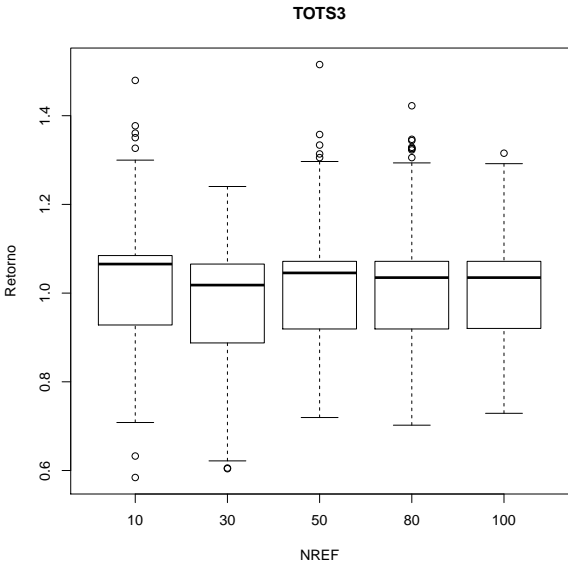
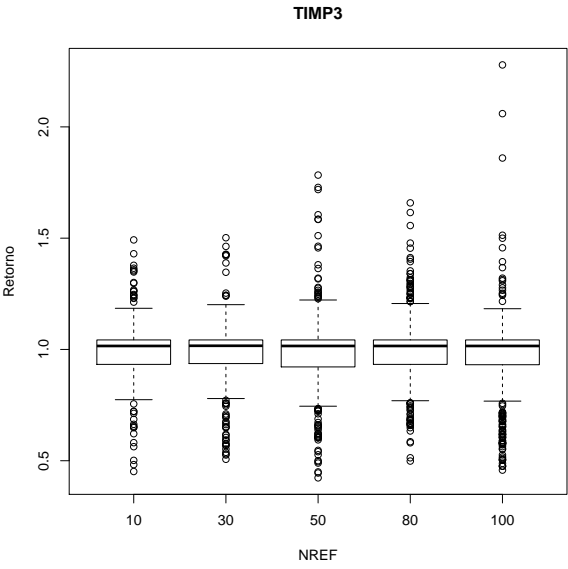
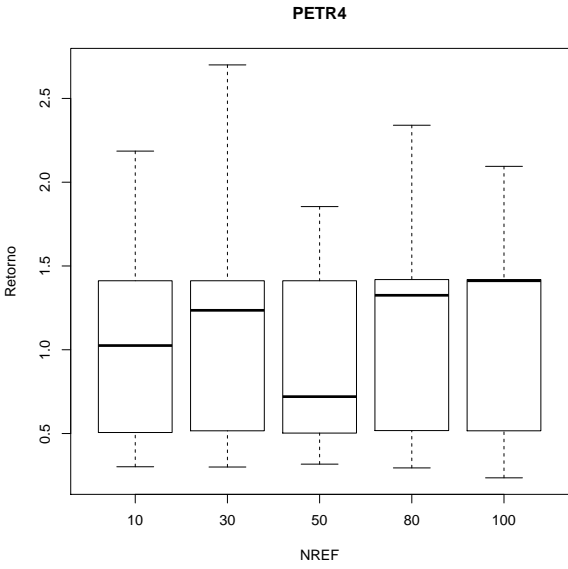
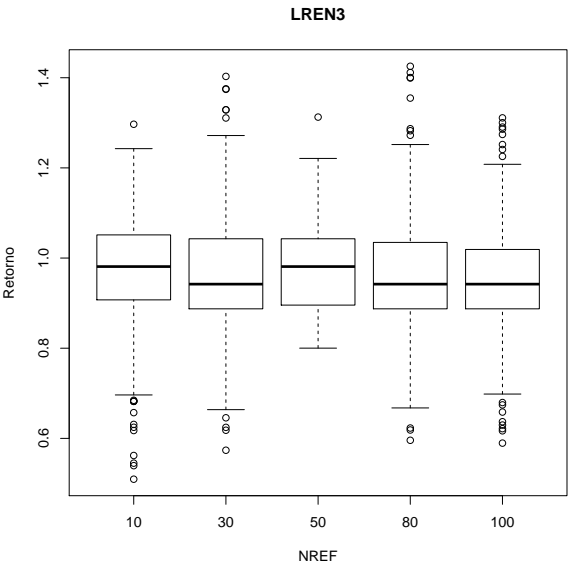
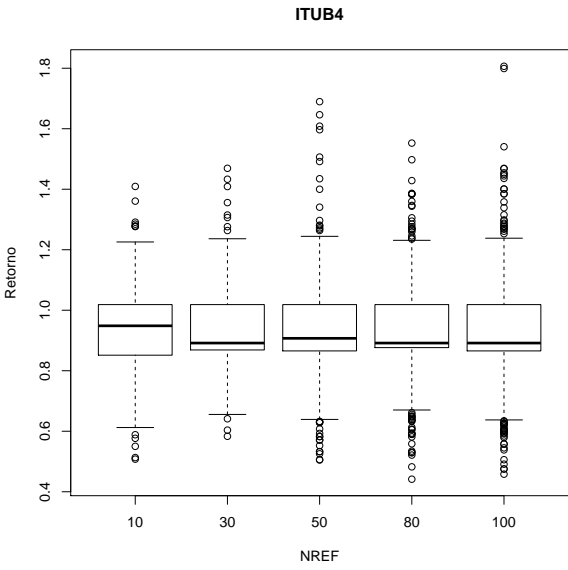




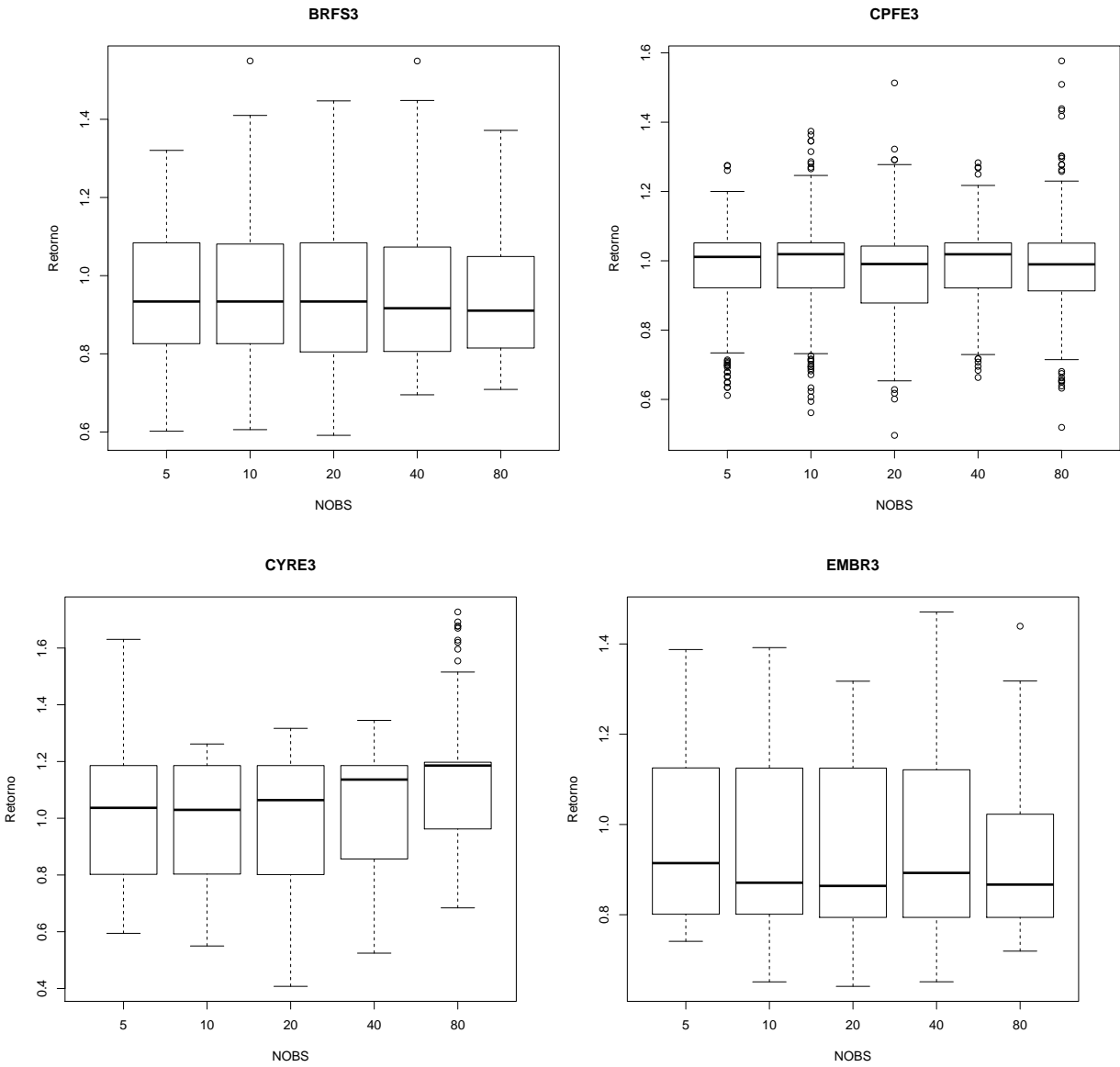


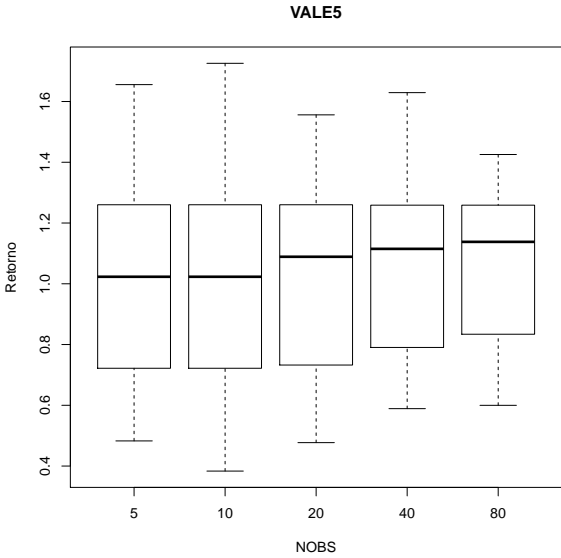
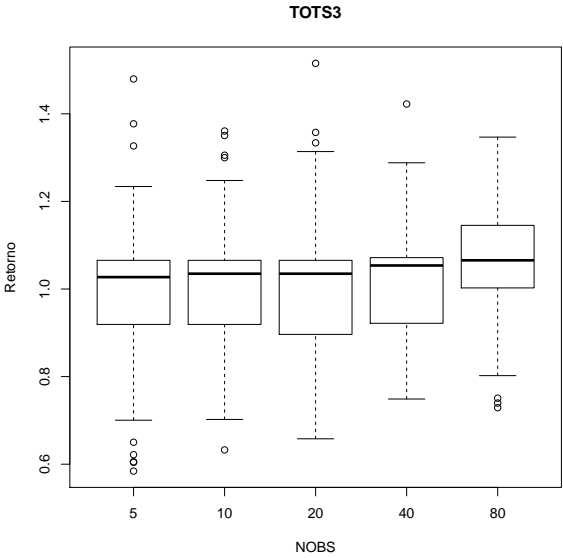
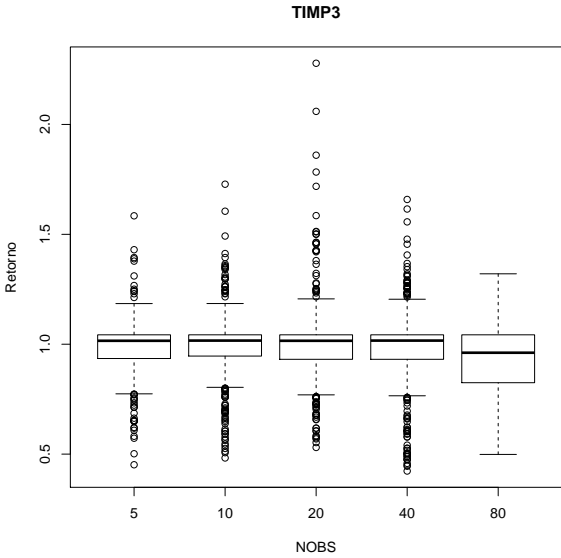
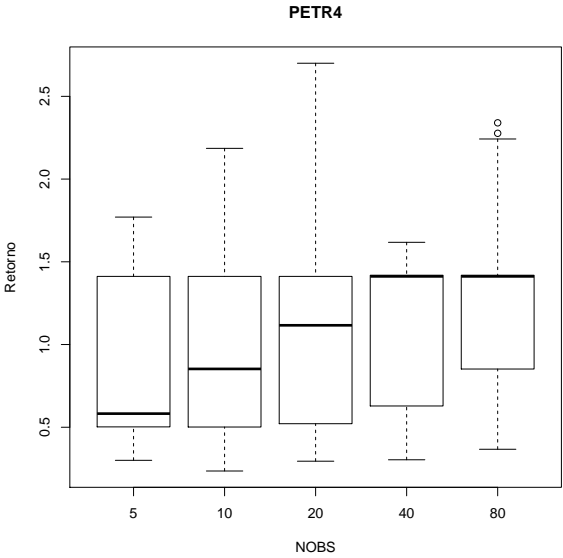
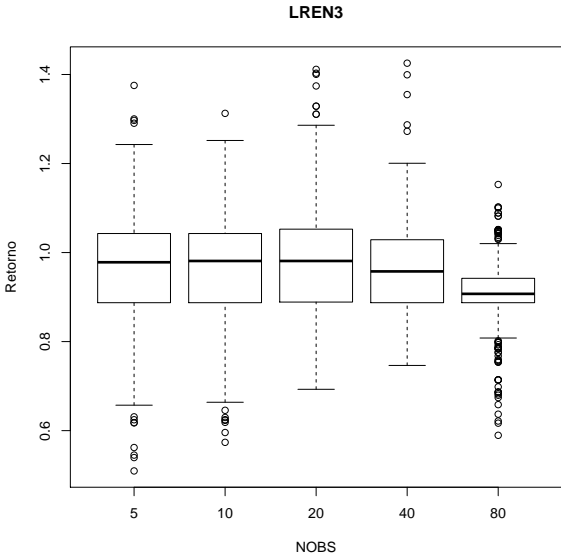
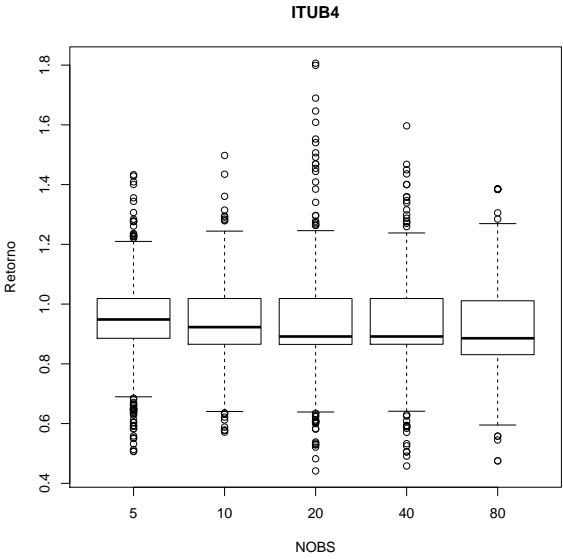
B.5 NREF





B.6 NOBS





## APÊNDICE C TABELAS DA VALIDAÇÃO DA ABORDAGEM

### 3

#### C.1 gama

BRFS3			
gama	Media	Mediana	Desvio
0.1	1.0003	1.0003	0.012
1	0.9944	0.9944	0.062
10	0.9873	0.9873	0.0989
CYRE3			
gama	Media	Mediana	Desvio
0.1	1.0011	1.0011	0.0158
1	0.9868	0.9868	0.0774
10	0.954	0.954	0.114
ITUB4			
gama	Media	Mediana	Desvio
0.1	0.9958	0.9958	0.0405
1	0.9819	0.9819	0.0945
10	0.9675	0.9675	0.1302
PETR4			
gama	Media	Mediana	Desvio
0.1	0.9986	0.9986	0.0215
1	0.9299	0.9299	0.2109
10	0.8376	0.8376	0.2937
TOTS3			
gama	Media	Mediana	Desvio
0.1	0.9999	0.9999	0.0085
1	0.9826	0.9826	0.0605
10	0.9662	0.9662	0.0817

CPFE3			
gama	Media	Mediana	Desvio
0.1	0.9857	0.9857	0.0654
1	0.8983	0.8983	0.1655
10	0.8679	0.8679	0.1787
EMBR3			
gama	Media	Mediana	Desvio
0.1	0.9973	0.9973	0.052
1	0.9467	0.9467	0.1323
10	0.9223	0.9223	0.1474
LREN3			
gama	Media	Mediana	Desvio
0.1	1.001	1.001	0.0134
1	0.9995	0.9995	0.0474
10	0.9985	0.9985	0.0783
TIMP3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	0.992	0.992	0.0534
10	0.9867	0.9867	0.0994
VALE5			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	1.0053	1.0053	0.0833
10	0.9696	0.9696	0.1242

## C.2 theta

BRFS3			
theta	Media	Mediana	Desvio
0.65	0.9932	0.9932	0.0768
1	0.9926	0.9926	0.0678
3	0.9961	0.9961	0.058

CPFE3			
theta	Media	Mediana	Desvio
0.65	0.904	0.904	0.1612
1	0.9157	0.9157	0.1559
3	0.9322	0.9322	0.1428

CYRE3			
theta	Media	Mediana	Desvio
0.65	0.9645	0.9645	0.1007
1	0.9834	0.9834	0.0829
3	0.9941	0.9941	0.0542

EMBR3			
theta	Media	Mediana	Desvio
0.65	0.9352	0.9352	0.1418
1	0.9439	0.9439	0.1274
3	0.9872	0.9872	0.0836

ITUB4			
theta	Media	Mediana	Desvio
0.65	0.972	0.972	0.1094
1	0.9797	0.9797	0.0956
3	0.9934	0.9934	0.0811

LREN3			
theta	Media	Mediana	Desvio
0.65	1.0016	1.0016	0.061
1	0.9988	0.9988	0.0556
3	0.9985	0.9985	0.0417

PETR4			
theta	Media	Mediana	Desvio
0.65	0.8953	0.8953	0.2415
1	0.9218	0.9218	0.2205
3	0.9491	0.9491	0.1894

TIMP3			
theta	Media	Mediana	Desvio
0.65	0.9915	0.9915	0.0753
1	0.9917	0.9917	0.0655
3	0.9955	0.9955	0.0534

TOTS3			
theta	Media	Mediana	Desvio
0.65	0.9764	0.9764	0.0698
1	0.9819	0.9819	0.0593
3	0.9904	0.9904	0.05

VALE5			
theta	Media	Mediana	Desvio
0.65	0.9892	0.9892	0.1006
1	0.9901	0.9901	0.0876
3	0.9956	0.9956	0.0727

### C.3 detectionsLimit

BRFS3			
dLimit	Media	Mediana	Desvio
1	0.9875	0.9875	0.0781
3	0.997	0.997	0.0674
5	0.9975	0.9975	0.0562

CPFE3			
dLimit	Media	Mediana	Desvio
1	0.8589	0.8589	0.187
3	0.9439	0.9439	0.1227
5	0.949	0.949	0.1263

CYRE3			
dLimit	Media	Mediana	Desvio
1	0.9719	0.9719	0.0963
3	0.9818	0.9818	0.0769
5	0.9883	0.9883	0.0713

EMBR3			
dLimit	Media	Mediana	Desvio
1	0.9177	0.9177	0.1581
3	0.9681	0.9681	0.1046
5	0.9805	0.9805	0.0821

ITUB4			
dLimit	Media	Mediana	Desvio
1	0.9705	0.9705	0.1175
3	0.9845	0.9845	0.0882
5	0.9902	0.9902	0.0781

LREN3			
dLimit	Media	Mediana	Desvio
1	0.995	0.995	0.0603
3	1.0002	1.0002	0.0513
5	1.0038	1.0038	0.0474

PETR4			
dLimit	Media	Mediana	Desvio
1	0.9086	0.9086	0.2248
3	0.9238	0.9238	0.2186
5	0.9337	0.9337	0.2136

TIMP3			
dLimit	Media	Mediana	Desvio
1	0.9892	0.9892	0.0767
3	0.9926	0.9926	0.059
5	0.9969	0.9969	0.0585

TOTS3			
dLimit	Media	Mediana	Desvio
1	0.9804	0.9804	0.0641
3	0.9845	0.9845	0.061
5	0.9838	0.9838	0.0561

VALE5			
dLimit	Media	Mediana	Desvio
1	0.984	0.984	0.0953
3	0.9935	0.9935	0.0867
5	0.9975	0.9975	0.0801



## C.4 NSMOOTH

BRFS3			
smooth	Media	Mediana	Desvio
1	0.9984	0.9984	0.0477
5	0.9829	0.9829	0.0867
10	0.9809	0.9809	0.079
20	1.0012	1.0012	0.0736
40	1.0066	1.0066	0.0336

CPFE3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.8847	0.8847	0.1913
10	0.8677	0.8677	0.1704
20	0.9153	0.9153	0.1446
40	0.9188	0.9188	0.1473

CYRE3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.956	0.956	0.1113
10	0.9639	0.9639	0.095
20	0.9839	0.9839	0.0874
40	0.9993	0.9993	0.0578

EMBR3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.9171	0.9171	0.1586
10	0.9341	0.9341	0.1441
20	0.952	0.952	0.128
40	0.9741	0.9741	0.0911

ITUB4			
smooth	Media	Mediana	Desvio
1	1.0064	1.0064	0.0492
5	0.9951	0.9951	0.1086
10	0.9724	0.9724	0.1039
20	0.9727	0.9727	0.1001
40	0.962	0.962	0.1009

LREN3			
smooth	Media	Mediana	Desvio
1	0.9981	0.9981	0.0342
5	0.9898	0.9898	0.0755
10	1.0059	1.0059	0.0578
20	1.0062	1.0062	0.0548
40	0.9982	0.9982	0.0296

PETR4			
smooth	Media	Mediana	Desvio
1	0.9862	0.9862	0.0796
5	0.9471	0.9471	0.2727
10	0.8578	0.8578	0.2533
20	0.8899	0.8899	0.2176
40	0.9292	0.9292	0.196

TIMP3			
smooth	Media	Mediana	Desvio
1	0.9992	0.9992	0.0189
5	0.9953	0.9953	0.071
10	0.9865	0.9865	0.0822
20	0.9911	0.9911	0.077
40	0.9924	0.9924	0.0568

TOTS3			
smooth	Media	Mediana	Desvio
1	1.0008	1.0008	0.0082
5	0.9752	0.9752	0.0714
10	0.9789	0.9789	0.0672
20	0.968	0.968	0.0692
40	0.9915	0.9915	0.0562

VALE5			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.9626	0.9626	0.111
10	0.9924	0.9924	0.1089
20	1.0034	1.0034	0.1062
40	0.9999	0.9999	0.0442

## C.5 NREF

BRFS3			
NREF	Media	Mediana	Desvio
10	1.0102	1.0102	0.054
30	1.0039	1.0039	0.0435
50	1.0002	1.0002	0.0626
80	0.9821	0.9821	0.0731
100	0.9886	0.9886	0.0801

CPFE3			
NREF	Media	Mediana	Desvio
10	0.985	0.985	0.0845
30	0.9598	0.9598	0.1119
50	0.9304	0.9304	0.1577
80	0.8927	0.8927	0.1642
100	0.8788	0.8788	0.1666

CYRE3			
NREF	Media	Mediana	Desvio
10	1.0064	1.0064	0.0567
30	0.9946	0.9946	0.0638
50	0.9722	0.9722	0.0861
80	0.9784	0.9784	0.0891
100	0.971	0.971	0.0878

EMBR3			
NREF	Media	Mediana	Desvio
10	1.0018	1.0018	0.0676
30	0.9782	0.9782	0.0995
50	0.964	0.964	0.1241
80	0.9467	0.9467	0.1196
100	0.925	0.925	0.1423

ITUB4			
NREF	Media	Mediana	Desvio
10	1.0108	1.0108	0.09
30	0.9819	0.9819	0.0763
50	0.9809	0.9809	0.0931
80	0.9741	0.9741	0.1154
100	0.9782	0.9782	0.0893

LREN3			
NREF	Media	Mediana	Desvio
10	0.9982	0.9982	0.0596
30	0.9981	0.9981	0.0497
50	1.0006	1.0006	0.0656
80	0.9978	0.9978	0.0449
100	1.0023	1.0023	0.0497

PETR4			
NREF	Media	Mediana	Desvio
10	0.9642	0.9642	0.1476
30	0.8896	0.8896	0.225
50	0.8772	0.8772	0.2741
80	0.963	0.963	0.2165
100	0.9195	0.9195	0.1792

TIMP3			
NREF	Media	Mediana	Desvio
10	1.0042	1.0042	0.0554
30	0.9976	0.9976	0.052
50	0.9952	0.9952	0.0684
80	0.9935	0.9935	0.0765
100	0.9831	0.9831	0.0604

TOTS3			
NREF	Media	Mediana	Desvio
10	0.9988	0.9988	0.038
30	0.9966	0.9966	0.0472
50	0.9852	0.9852	0.0646
80	0.9736	0.9736	0.0655
100	0.9758	0.9758	0.0635

VALE5			
NREF	Media	Mediana	Desvio
10	1.0054	1.0054	0.0694
30	0.9917	0.9917	0.0664
50	0.9944	0.9944	0.0701
80	0.9964	0.9964	0.102
100	0.9791	0.9791	0.1006

## C.6 NOBS

BRFS3			
NOBS	Media	Mediana	Desvio
5	0.9822	0.9822	0.0609
10	0.9914	0.9914	0.0605
20	1.0005	1.0005	0.0609
40	1.0114	1.0114	0.0847
80	0.991	0.991	0.08

CPFE3			
NOBS	Media	Mediana	Desvio
5	0.9251	0.9251	0.1437
10	0.9245	0.9245	0.1582
20	0.8915	0.8915	0.1611
40	0.9038	0.9038	0.1601
80	0.9516	0.9516	0.1331

CYRE3			
NOBS	Media	Mediana	Desvio
5	0.984	0.984	0.0512
10	0.9811	0.9811	0.0674
20	0.9703	0.9703	0.0867
40	0.9781	0.9781	0.1002
80	0.9957	0.9957	0.1273

EMBR3			
NOBS	Media	Mediana	Desvio
5	0.9588	0.9588	0.119
10	0.9433	0.9433	0.1221
20	0.9438	0.9438	0.137
40	0.9585	0.9585	0.1247
80	0.9962	0.9962	0.0796

ITUB4			
NOBS	Media	Mediana	Desvio
5	0.9986	0.9986	0.0496
10	0.9921	0.9921	0.0777
20	0.9757	0.9757	0.0846
40	0.9667	0.9667	0.1204
80	0.9483	0.9483	0.1694

LREN3			
NOBS	Media	Mediana	Desvio
5	1.0033	1.0033	0.0233
10	1.0057	1.0057	0.0433
20	1.0097	1.0097	0.0543
40	0.9894	0.9894	0.086
80	0.9708	0.9708	0.0525

PETR4			
NOBS	Media	Mediana	Desvio
5	0.9165	0.9165	0.1846
10	0.9039	0.9039	0.203
20	0.8717	0.8717	0.2349
40	0.951	0.951	0.2595
80	1.0383	1.0383	0.1919

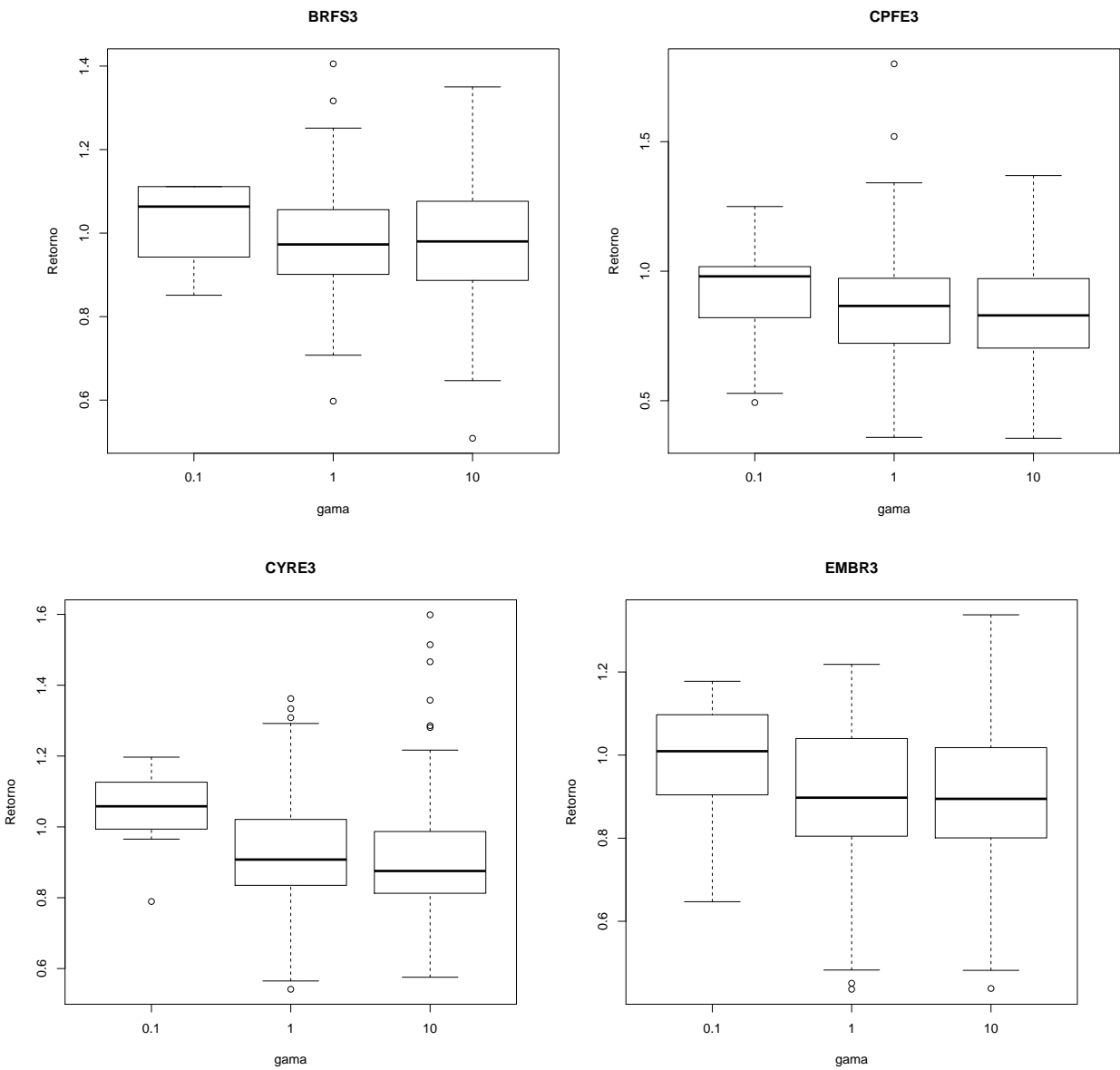
TIMP3			
NOBS	Media	Mediana	Desvio
5	0.9981	0.9981	0.0223
10	1.0014	1.0014	0.039
20	0.9987	0.9987	0.0722
40	0.979	0.979	0.0983
80	0.968	0.968	0.0974

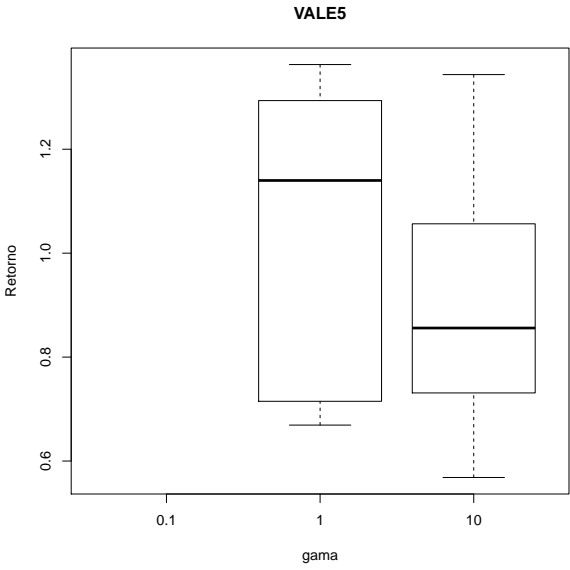
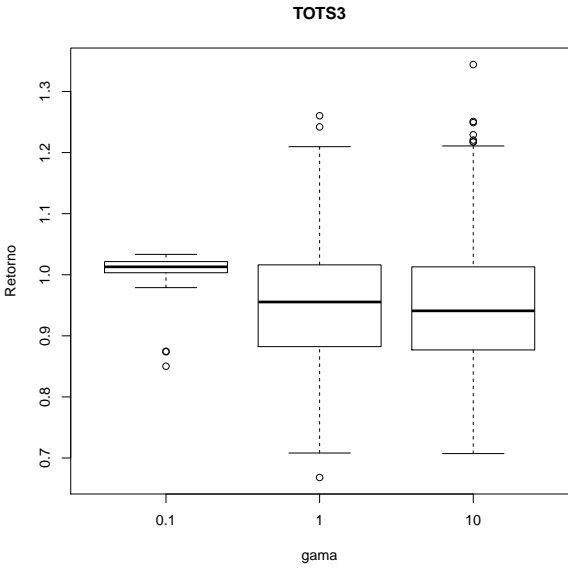
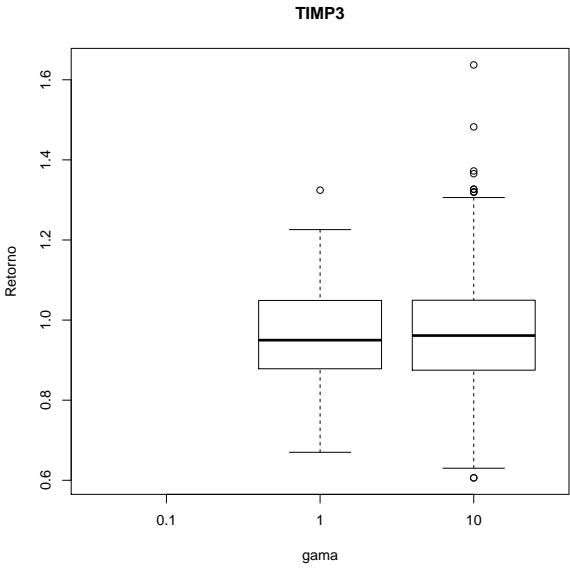
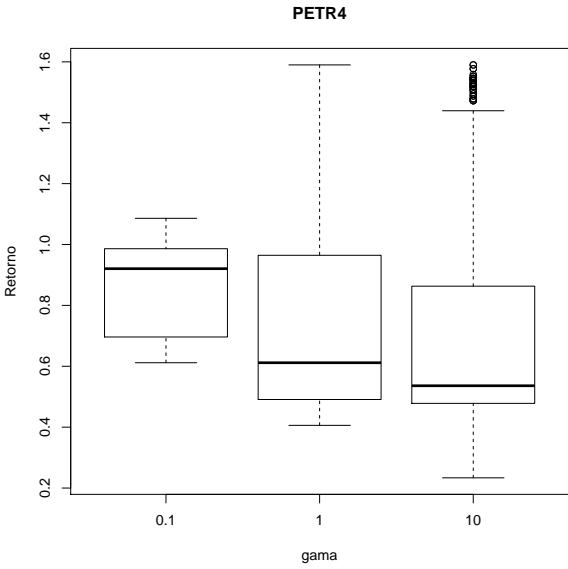
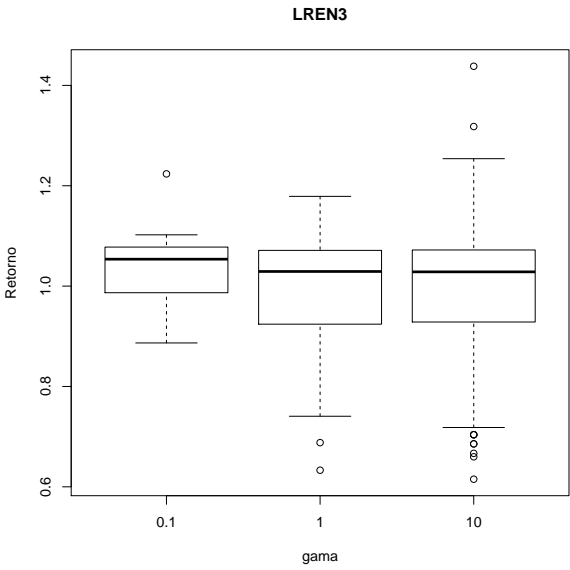
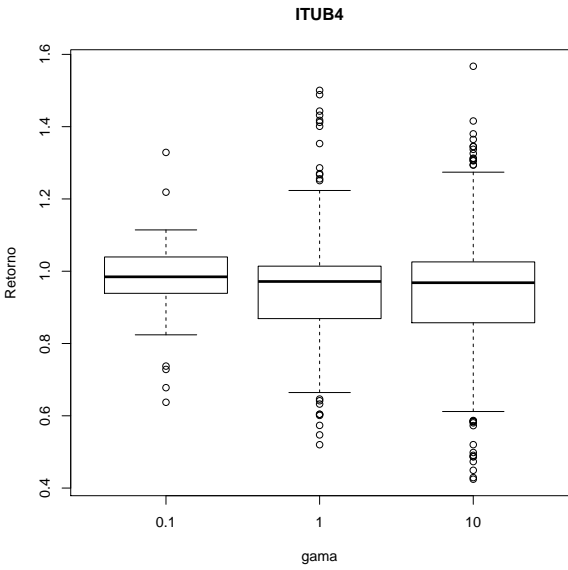
TOTS3			
NOBS	Media	Mediana	Desvio
5	0.9841	0.9841	0.0466
10	0.9804	0.9804	0.0549
20	0.9731	0.9731	0.064
40	0.9771	0.9771	0.0696
80	1.0142	1.0142	0.0718

VALE5			
NOBS	Media	Mediana	Desvio
5	1.0014	1.0014	0.0208
10	0.9964	0.9964	0.0545
20	0.9708	0.9708	0.0915
40	0.9663	0.9663	0.1105
80	1.0352	1.0352	0.1602

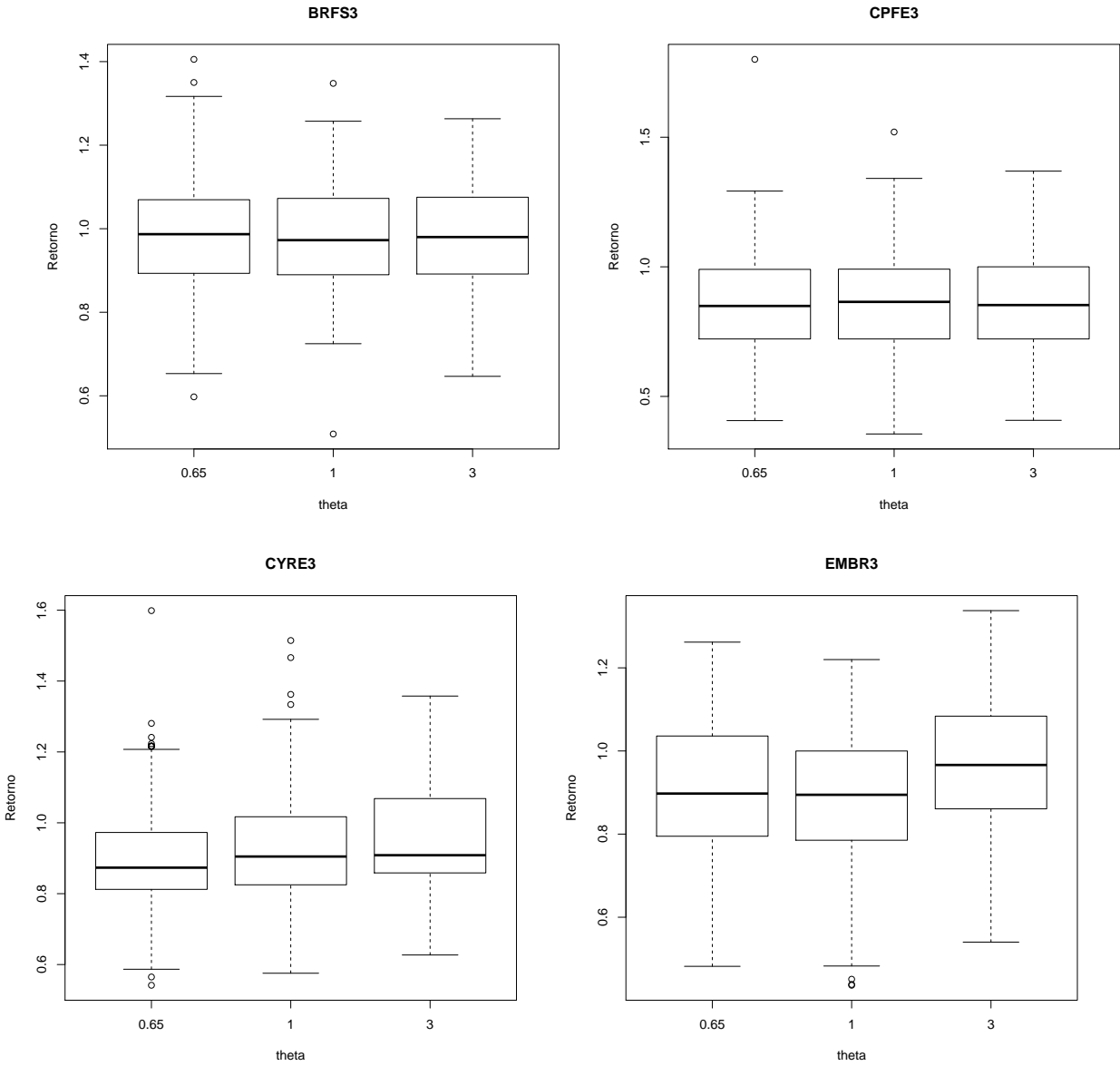
# APÊNDICE D BOXPLOTS DA VALIDAÇÃO DA ABORDA- GEM 3

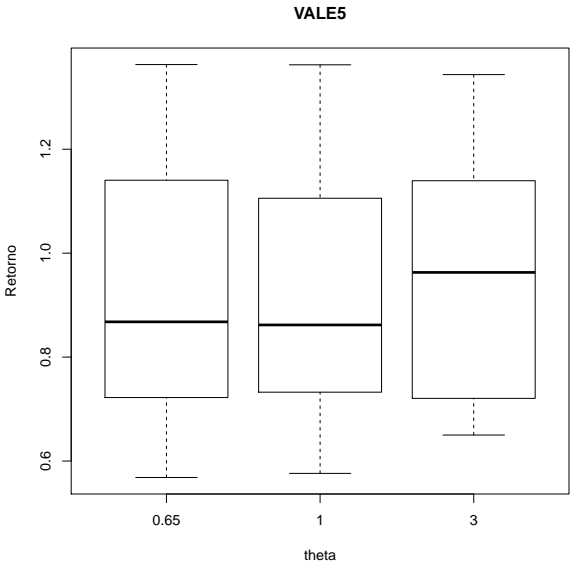
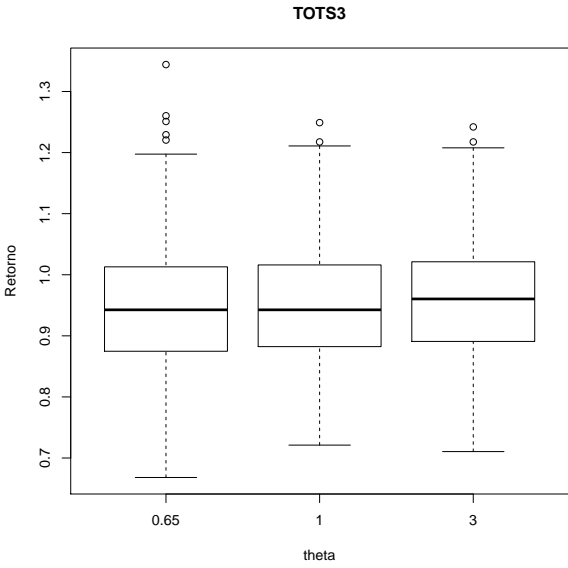
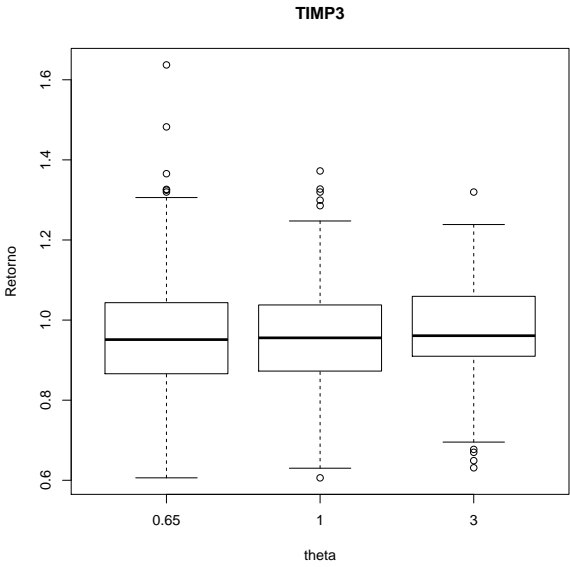
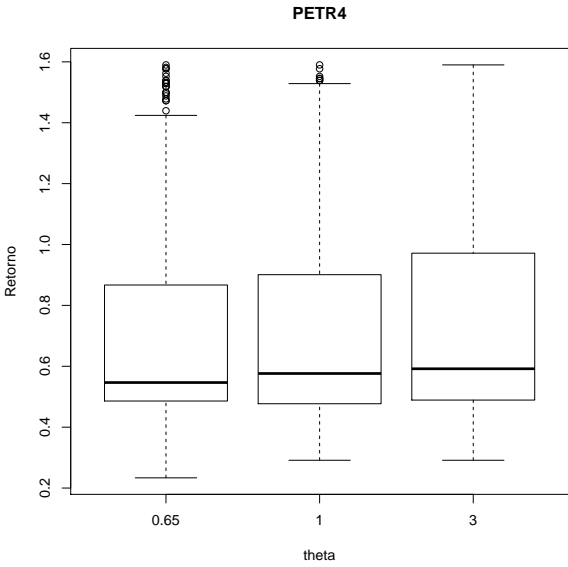
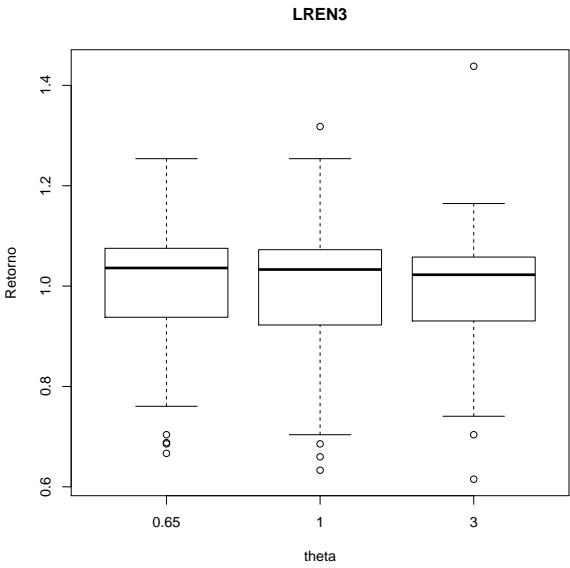
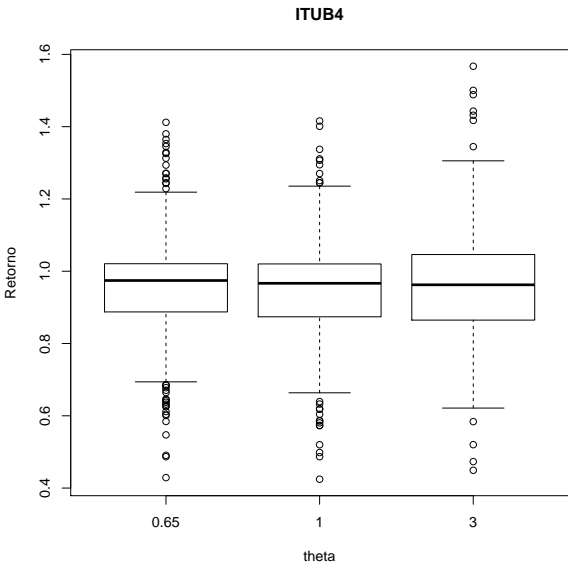
## D.1 gama



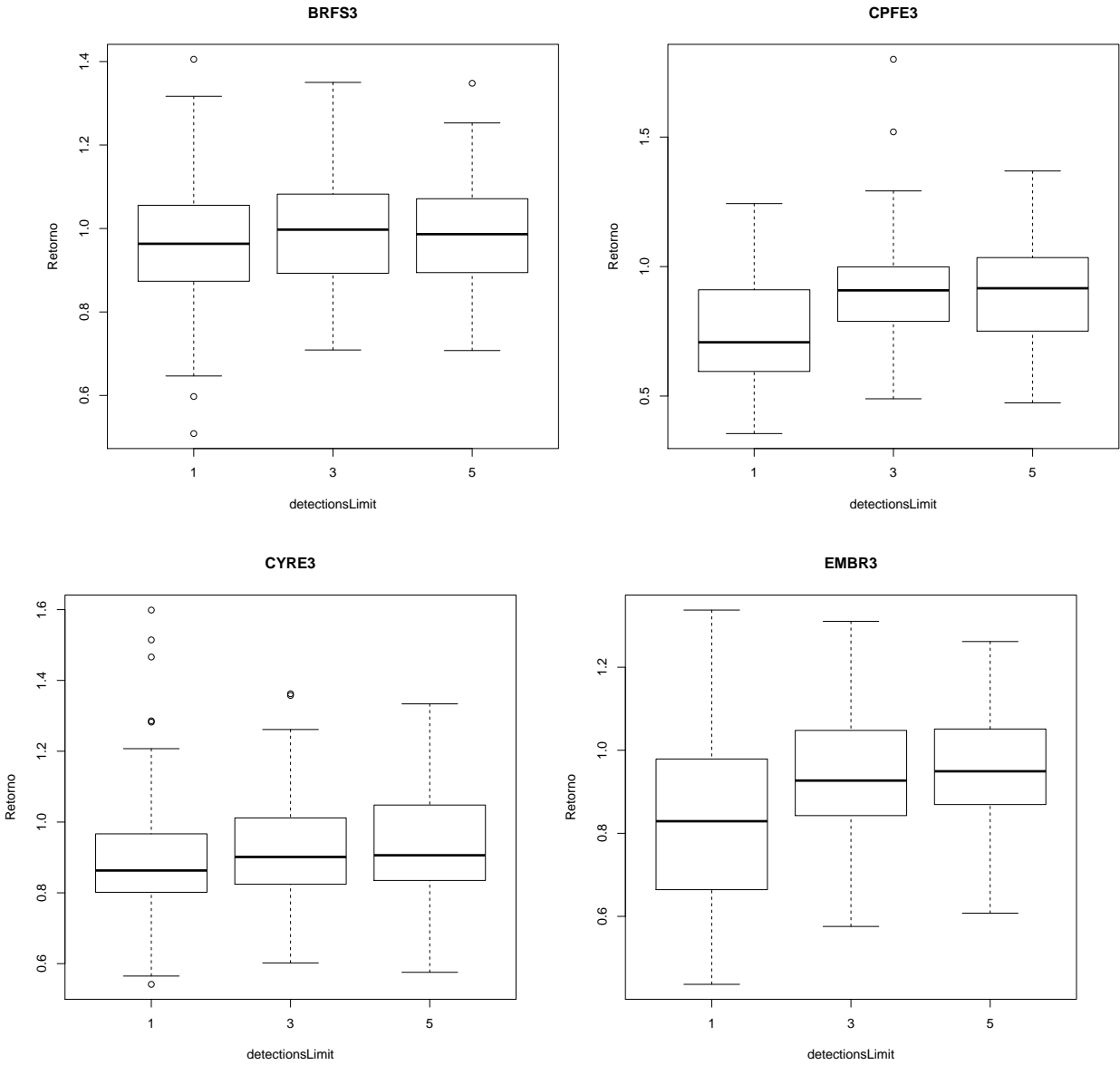


D.2 theta

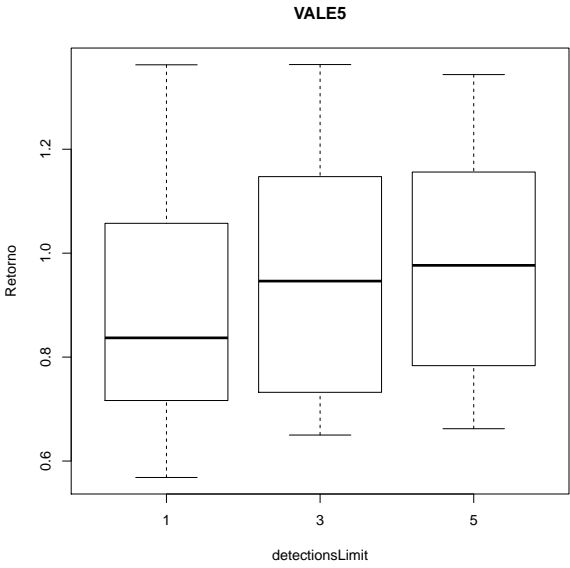
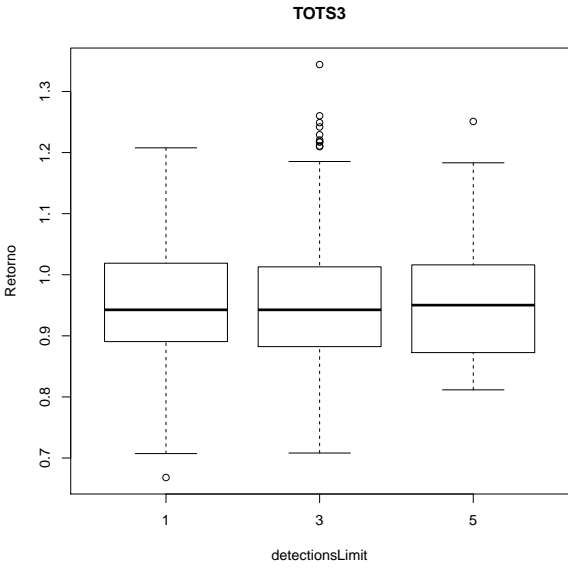
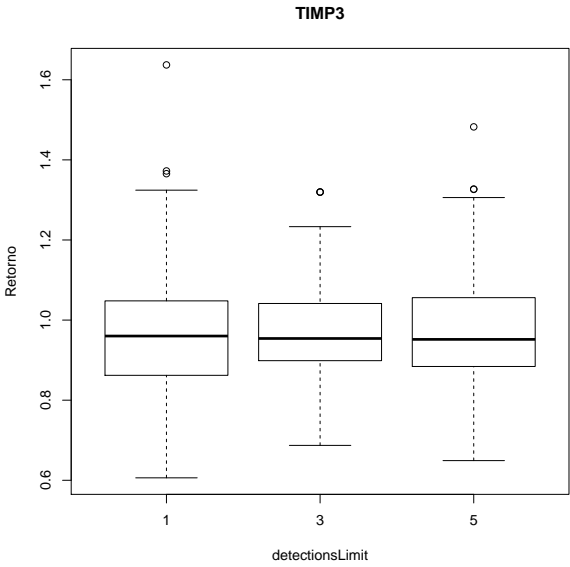
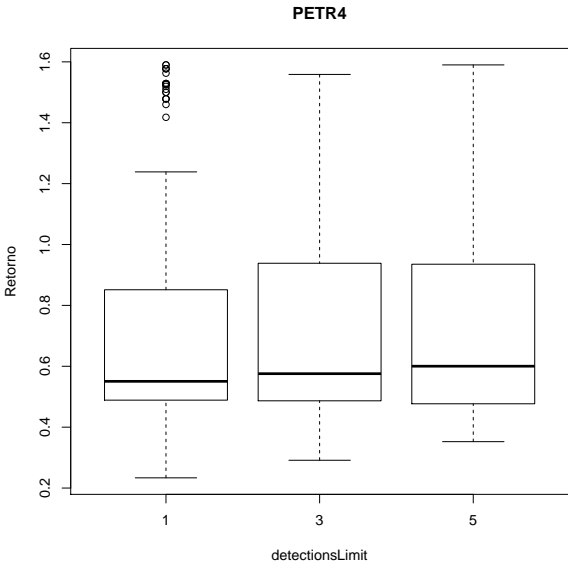
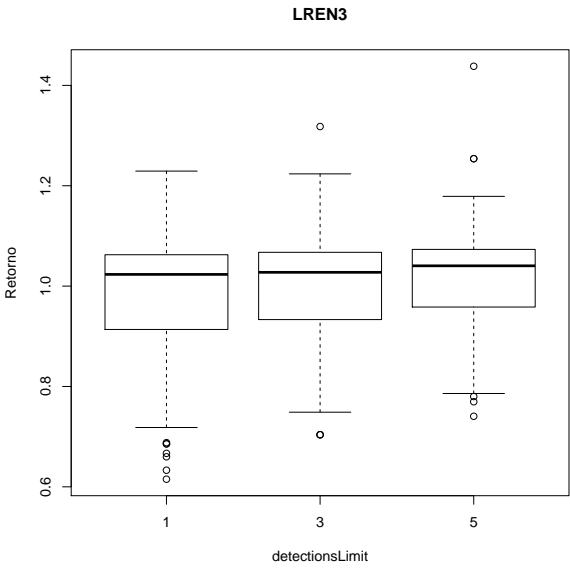
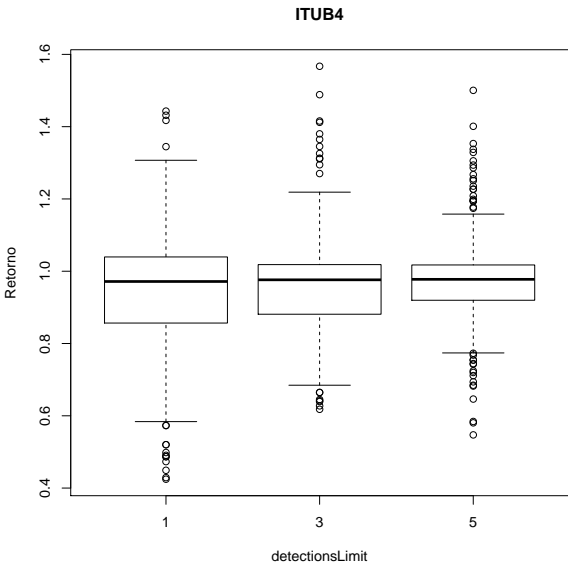




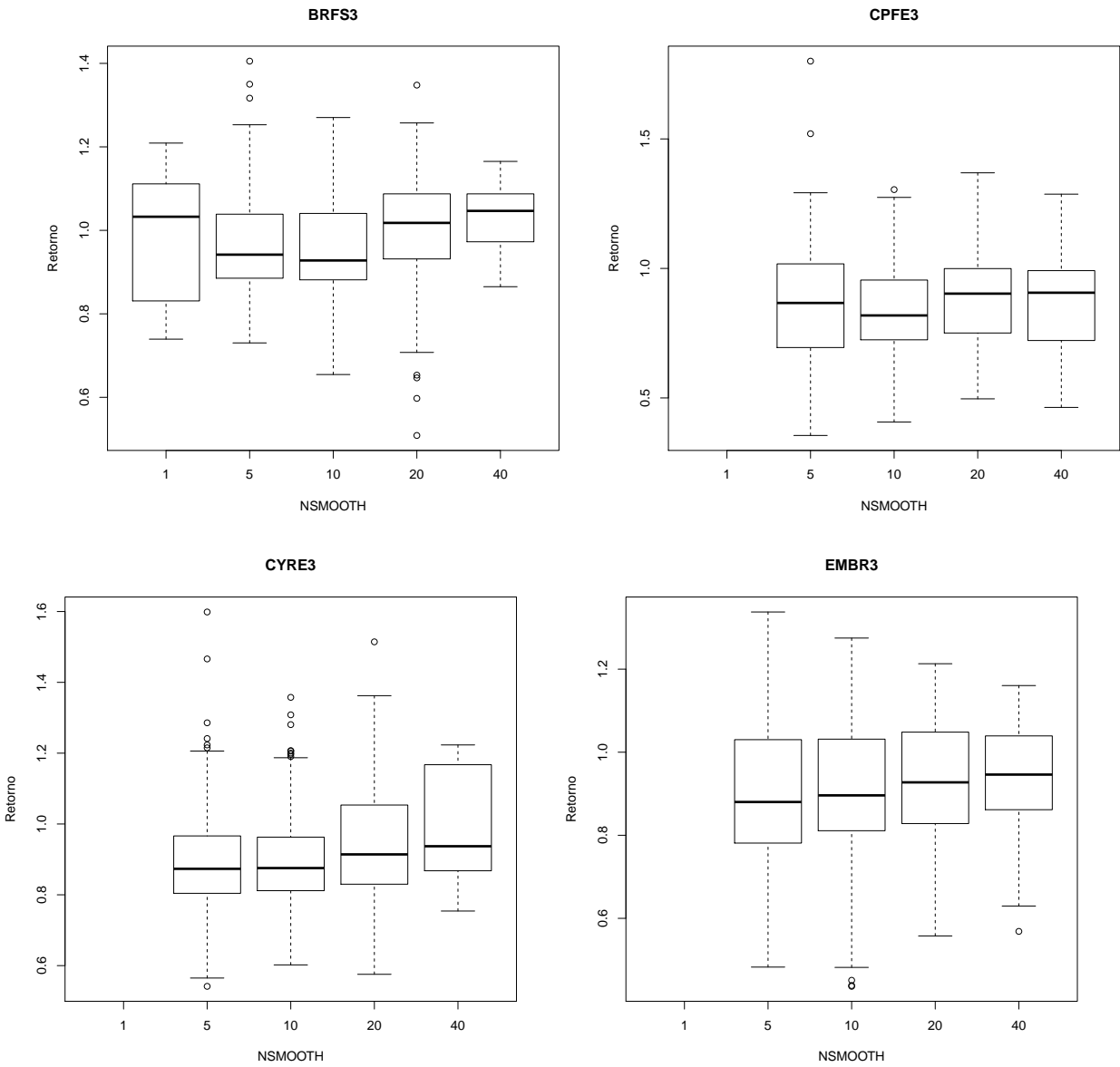
D.3 detectionsLimit

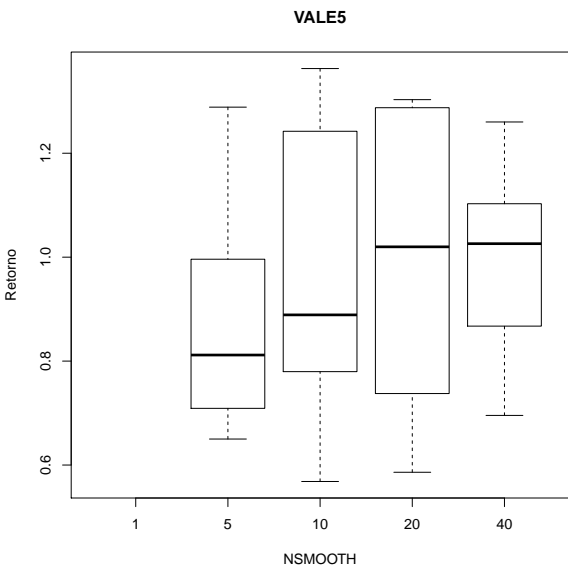
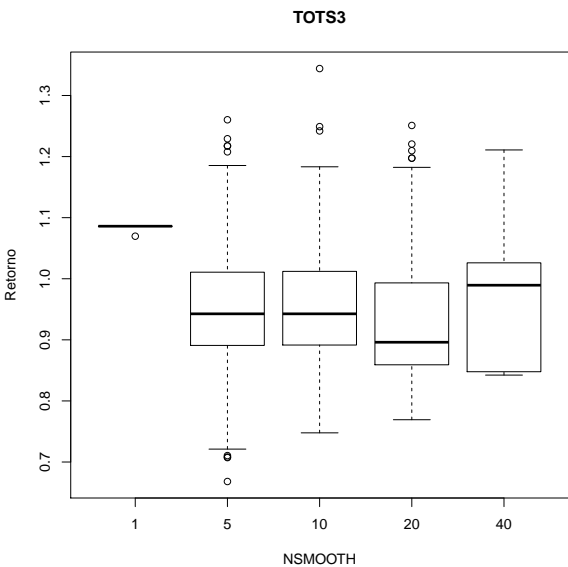
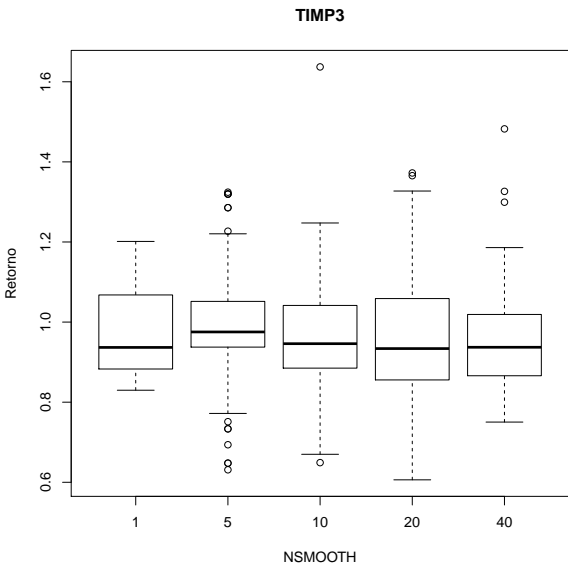
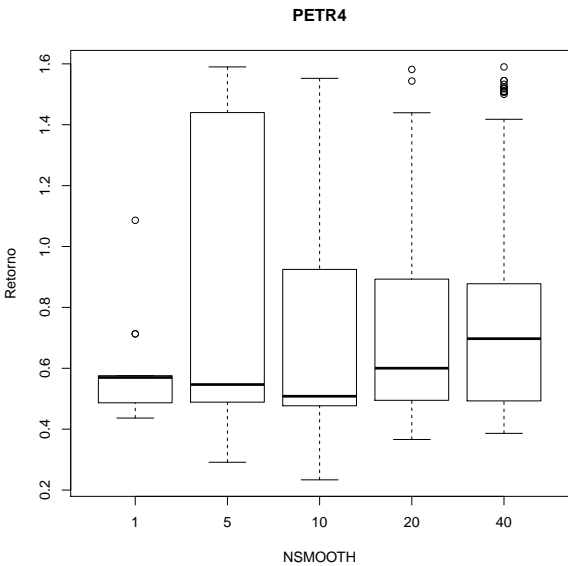
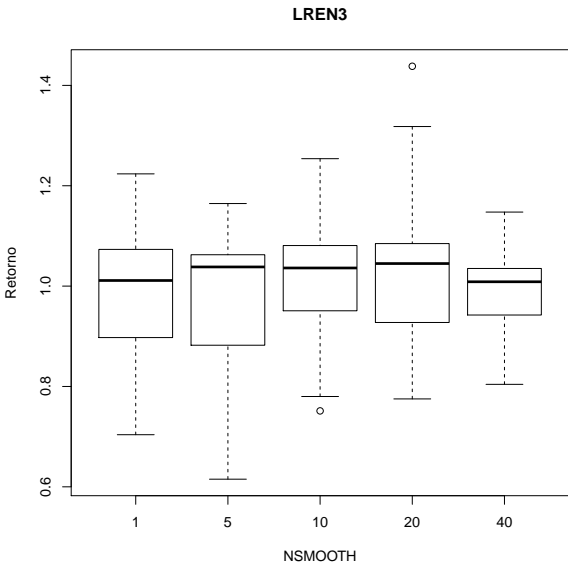
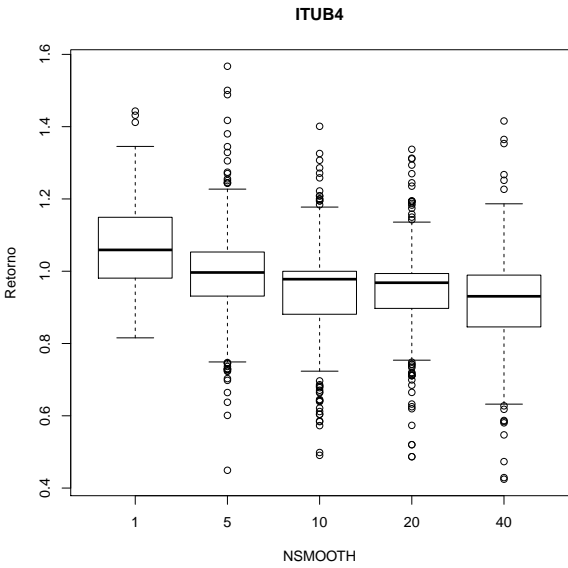




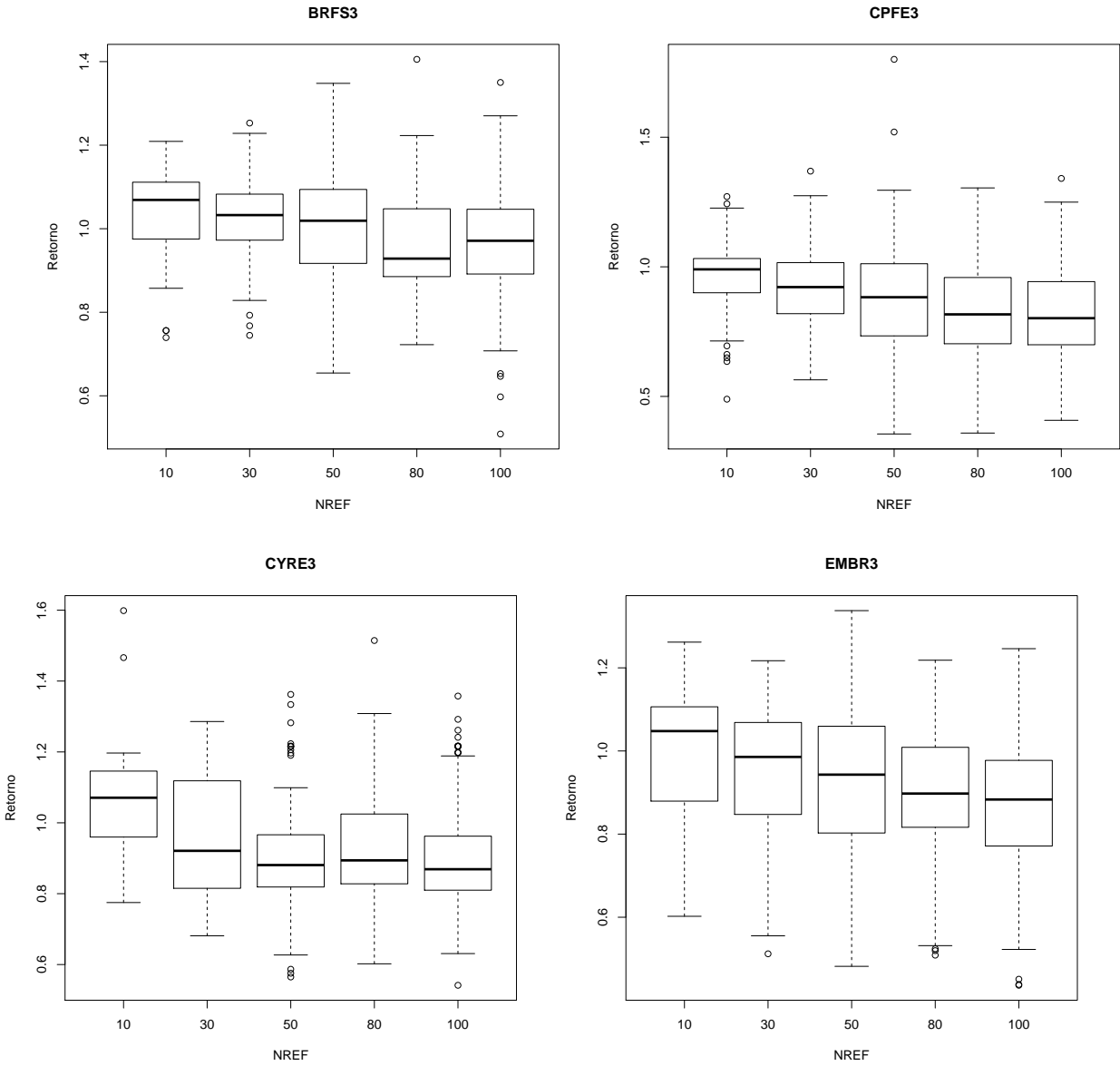


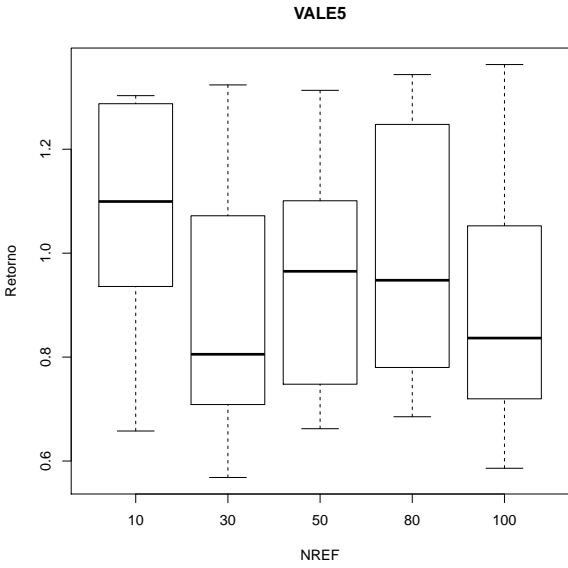
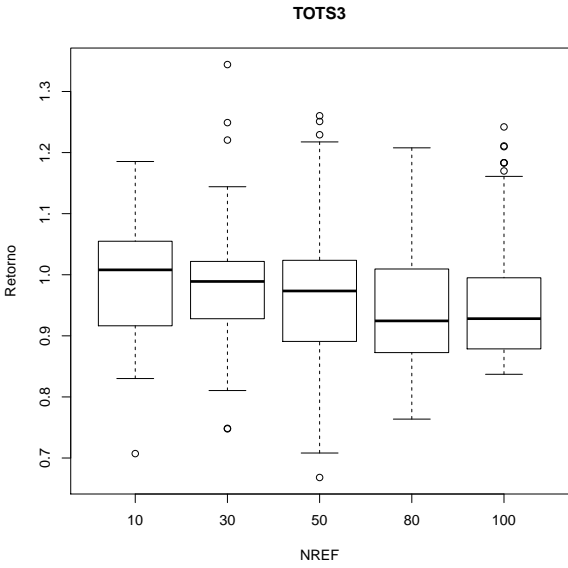
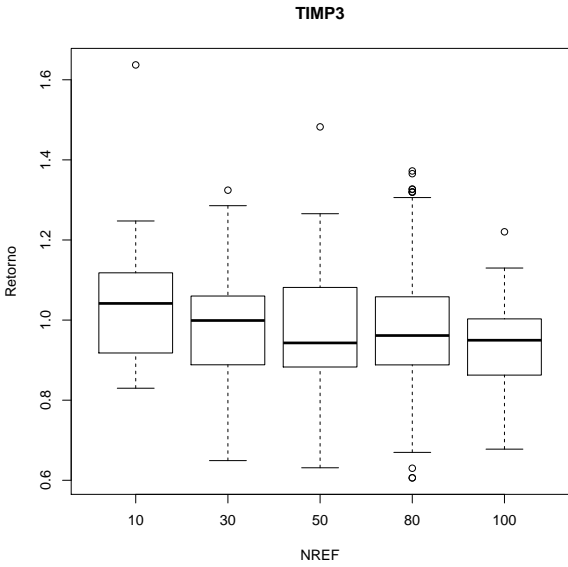
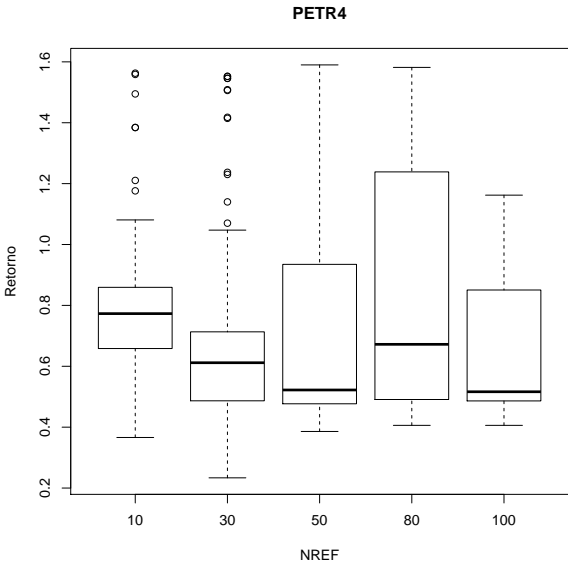
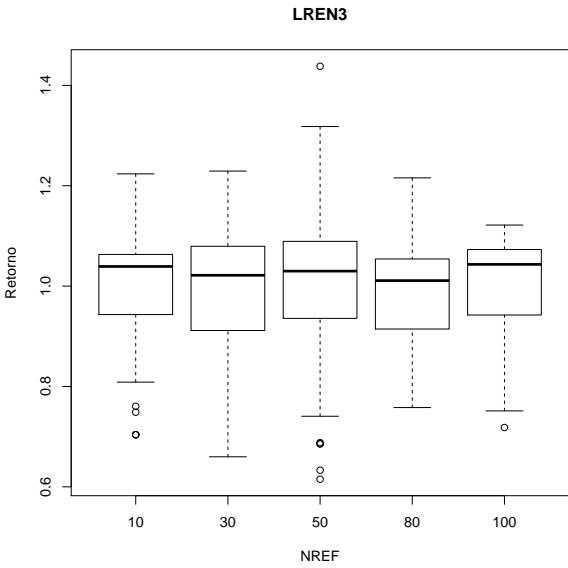
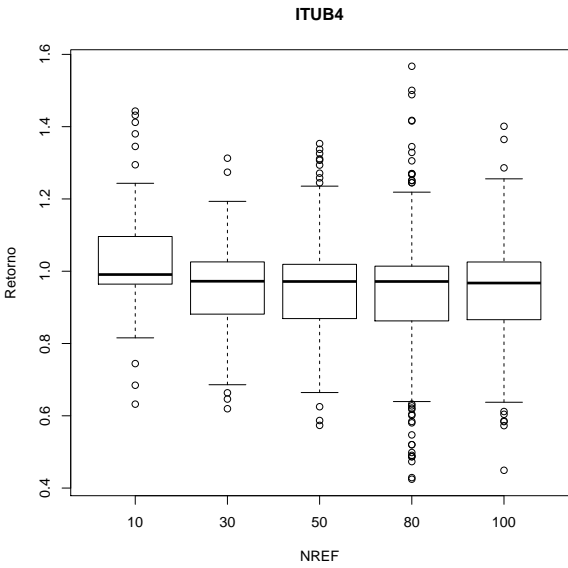
D.4 NSMOOTH



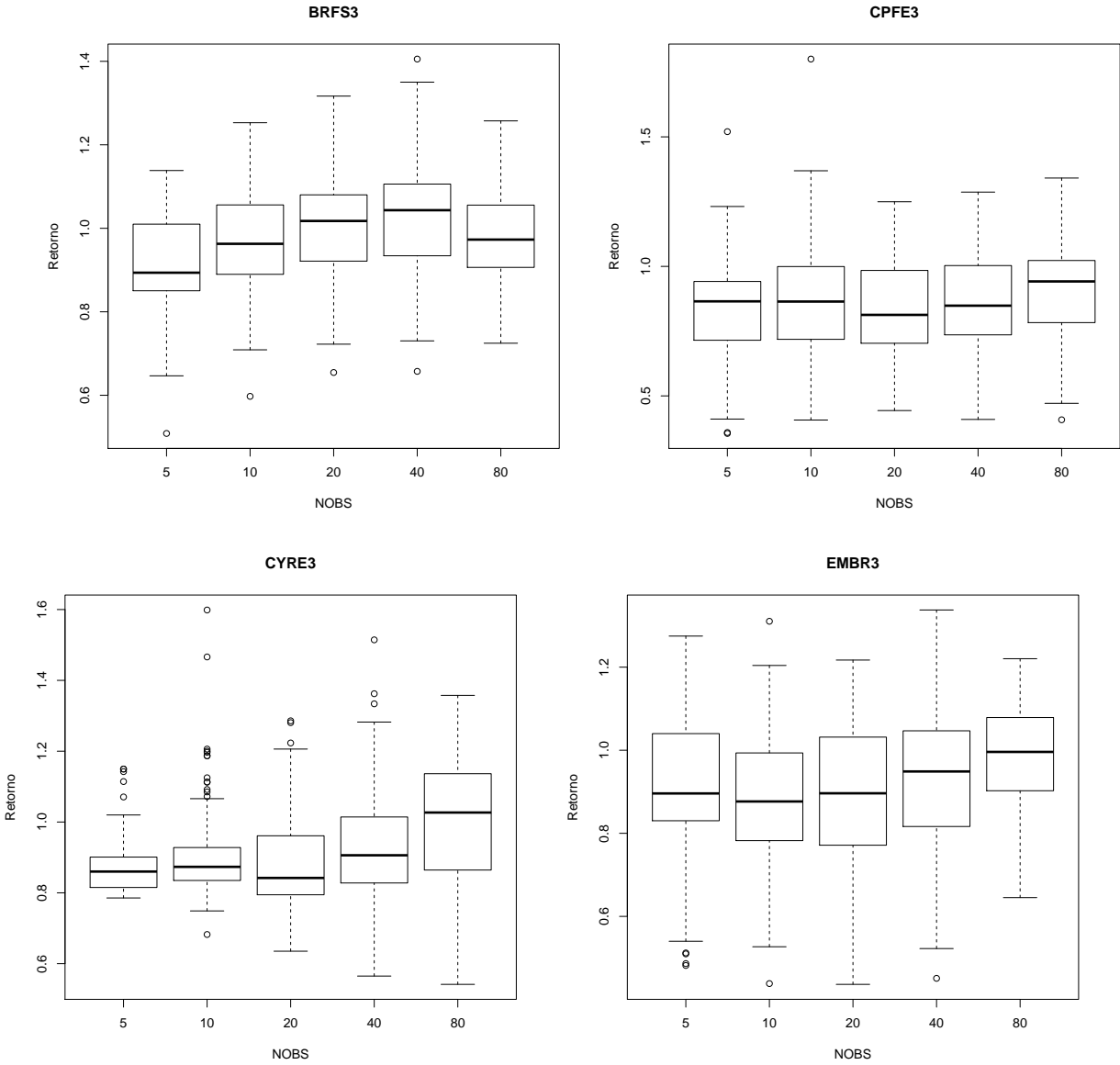


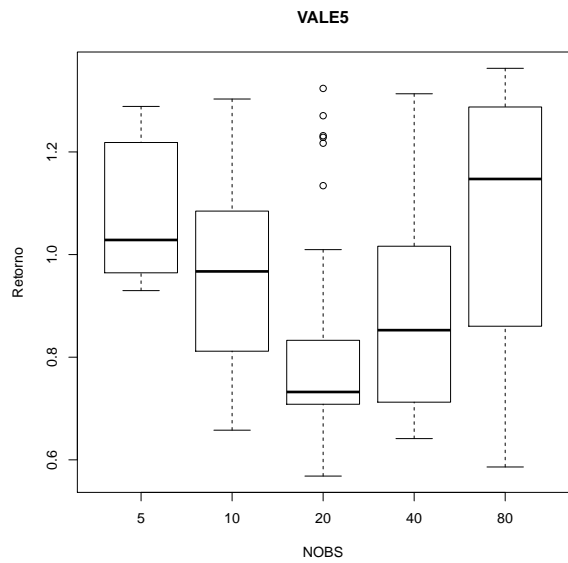
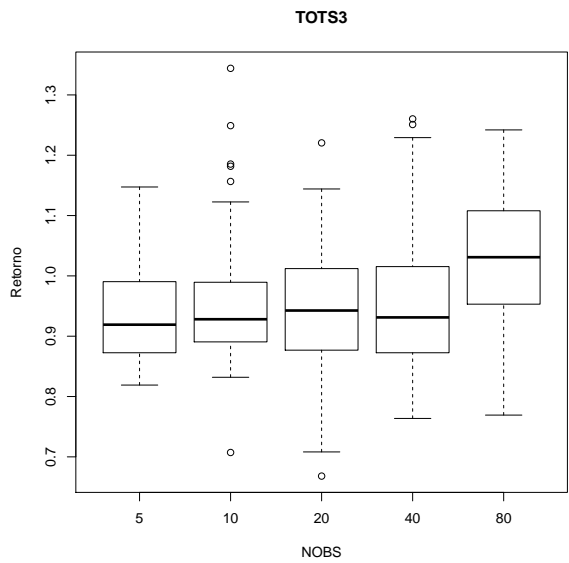
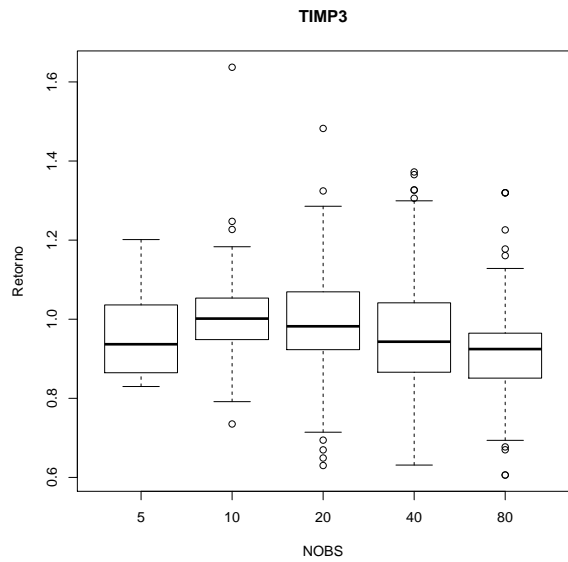
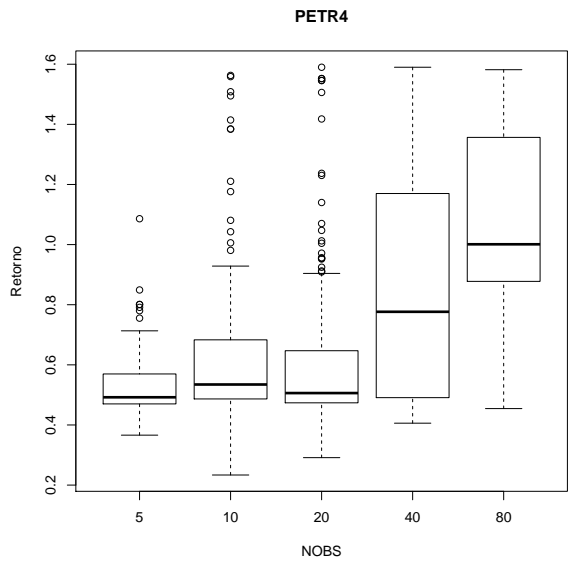
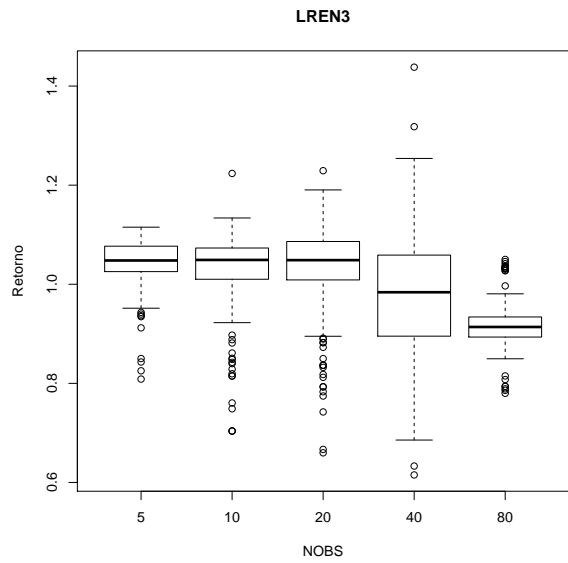
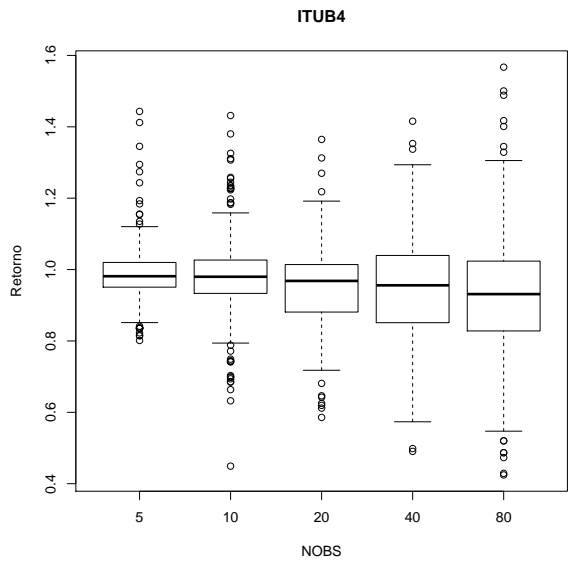
D.5 NREF





D.6 NOBS





## APÊNDICE E TABELAS DA VALIDAÇÃO DA ABORDAGEM

### 4

#### E.1 gama

BRFS3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	0.9975	0.9975	0.0171
10	1.0013	1.0013	0.0559
CYRE3			
gama	Media	Mediana	Desvio
0.1	1.0005	1.0005	0.0085
1	1.0026	1.0026	0.0377
10	0.9883	0.9883	0.0936
ITUB4			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	0.9952	0.9952	0.0323
10	0.9848	0.9848	0.0875
PETR4			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	0.9621	0.9621	0.1553
10	0.8686	0.8686	0.2699
TOTS3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	1.0019	1.0019	0.0207
10	0.9916	0.9916	0.0734

CPFE3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	1.0015	1.0015	0.0196
10	0.9922	0.9922	0.0851
EMBR3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	0.9966	0.9966	0.0279
10	1.0019	1.0019	0.0695
LREN3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	1.0004	1.0004	0.0213
10	1.0063	1.0063	0.0588
TIMP3			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	0.9995	0.9995	0.0286
10	0.9822	0.9822	0.1045
VALE5			
gama	Media	Mediana	Desvio
0.1	1	1	0
1	1.0088	1.0088	0.0517
10	0.9875	0.9875	0.1152



## E.2 theta

BRFS3			
theta	Media	Mediana	Desvio
0.65	0.9997	0.9997	0.0384
1	0.9999	0.9999	0.0327
3	0.9992	0.9992	0.0298

CPFE3			
theta	Media	Mediana	Desvio
0.65	0.9961	0.9961	0.066
1	0.9966	0.9966	0.046
3	1.001	1.001	0.0344

CYRE3			
theta	Media	Mediana	Desvio
0.65	0.9908	0.9908	0.0732
1	0.9987	0.9987	0.0562
3	1.002	1.002	0.0423

EMBR3			
theta	Media	Mediana	Desvio
0.65	1.0013	1.0013	0.0511
1	0.9994	0.9994	0.0437
3	0.9978	0.9978	0.033

ITUB4			
theta	Media	Mediana	Desvio
0.65	0.9907	0.9907	0.0679
1	0.9914	0.9914	0.0542
3	0.9978	0.9978	0.0352

LREN3			
theta	Media	Mediana	Desvio
0.65	1.0035	1.0035	0.0409
1	1.0025	1.0025	0.038
3	1.0008	1.0008	0.0286

PETR4			
theta	Media	Mediana	Desvio
0.65	0.9087	0.9087	0.2185
1	0.9438	0.9438	0.1906
3	0.9782	0.9782	0.1401

TIMP3			
theta	Media	Mediana	Desvio
0.65	0.9888	0.9888	0.0763
1	0.994	0.994	0.0632
3	0.9989	0.9989	0.0457

TOTS3			
theta	Media	Mediana	Desvio
0.65	0.9951	0.9951	0.0508
1	0.9986	0.9986	0.0469
3	0.9999	0.9999	0.0328

VALE5			
theta	Media	Mediana	Desvio
0.65	0.995	0.995	0.0879
1	0.9976	0.9976	0.0696
3	1.0037	1.0037	0.0596

### E.3 detectionsLimit

BRFS3			
dLimit	Media	Mediana	Desvio
1	0.9988	0.9988	0.0399
3	1.0004	1.0004	0.0335
5	0.9997	0.9997	0.0267

CPFE3			
dLimit	Media	Mediana	Desvio
1	0.997	0.997	0.0646
3	0.9985	0.9985	0.0484
5	0.9982	0.9982	0.0339

CYRE3			
dLimit	Media	Mediana	Desvio
1	0.9924	0.9924	0.0663
3	0.9984	0.9984	0.0604
5	1.0006	1.0006	0.0479

EMBR3			
dLimit	Media	Mediana	Desvio
1	0.9985	0.9985	0.0483
3	1.0013	1.0013	0.0438
5	0.9987	0.9987	0.0369

ITUB4			
dLimit	Media	Mediana	Desvio
1	0.9895	0.9895	0.0681
3	0.9938	0.9938	0.0534
5	0.9966	0.9966	0.036

LREN3			
dLimit	Media	Mediana	Desvio
1	1.003	1.003	0.0467
3	1.0022	1.0022	0.0328
5	1.0015	1.0015	0.026

PETR4			
dLimit	Media	Mediana	Desvio
1	0.9274	0.9274	0.1986
3	0.9452	0.9452	0.1876
5	0.9582	0.9582	0.1762

TIMP3			
dLimit	Media	Mediana	Desvio
1	0.9905	0.9905	0.0745
3	0.9958	0.9958	0.0647
5	0.9955	0.9955	0.0469

TOTS3			
dLimit	Media	Mediana	Desvio
1	0.9938	0.9938	0.0523
3	0.9994	0.9994	0.0434
5	1.0003	1.0003	0.0351

VALE5			
dLimit	Media	Mediana	Desvio
1	0.9942	0.9942	0.0777
3	0.9987	0.9987	0.0742
5	1.0034	1.0034	0.0677

## E.4 NSMOOTH

BRFS3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.9954	0.9954	0.0443
10	1.0047	1.0047	0.0503
20	0.9979	0.9979	0.032
40	1	1	0.0124

CPFE3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.998	0.998	0.0635
10	0.9973	0.9973	0.0714
20	0.9954	0.9954	0.0558
40	0.9988	0.9988	0.0233

CYRE3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	1.0001	1.0001	0.0772
10	0.9894	0.9894	0.0874
20	0.9974	0.9974	0.0495
40	0.9989	0.9989	0.0341

EMBR3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	1.0001	1.0001	0.0557
10	1.0015	1.0015	0.0577
20	0.9993	0.9993	0.0408
40	0.9967	0.9967	0.0357

ITUB4			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.9958	0.9958	0.0604
10	0.9803	0.9803	0.076
20	0.998	0.998	0.0494
40	0.9925	0.9925	0.0509

LREN3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	1.0039	1.0039	0.0463
10	1.0057	1.0057	0.0533
20	1.0019	1.0019	0.0361
40	0.9998	0.9998	0.0159

PETR4			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.9626	0.9626	0.2467
10	0.8763	0.8763	0.2253
20	0.923	0.923	0.1805
40	0.9559	0.9559	0.1554

TIMP3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	0.9996	0.9996	0.069
10	0.9824	0.9824	0.1004
20	0.9881	0.9881	0.0665
40	0.9994	0.9994	0.02

TOTS3			
smooth	Media	Mediana	Desvio
1	1	1	0
5	1.002	1.002	0.0586
10	0.9989	0.9989	0.0543
20	0.9899	0.9899	0.0523
40	0.9984	0.9984	0.0244

VALE5			
smooth	Media	Mediana	Desvio
1	1	1	0
5	1.0056	1.0056	0.1123
10	0.9874	0.9874	0.0909
20	1.0011	1.0011	0.0653
40	0.9997	0.9997	0.0407

## E.5 NREF

BRFS3			
NREF	Media	Mediana	Desvio
10	1.0002	1.0002	0.0144
30	1.0015	1.0015	0.0199
50	1.0015	1.0015	0.0395
80	0.9935	0.9935	0.0419
100	1.0028	1.0028	0.0311

CPFE3			
NREF	Media	Mediana	Desvio
10	0.9979	0.9979	0.0281
30	0.9908	0.9908	0.0508
50	1.0001	1.0001	0.0493
80	1.0025	1.0025	0.0625
100	0.9958	0.9958	0.0442

CYRE3			
NREF	Media	Mediana	Desvio
10	0.9972	0.9972	0.0292
30	1.0015	1.0015	0.0491
50	0.9966	0.9966	0.0549
80	1.0008	1.0008	0.0689
100	0.9913	0.9913	0.0642

EMBR3			
NREF	Media	Mediana	Desvio
10	0.9998	0.9998	0.019
30	1.0023	1.0023	0.0226
50	1.0004	1.0004	0.0334
80	0.9909	0.9909	0.0645
100	1.0056	1.0056	0.0388

ITUB4			
NREF	Media	Mediana	Desvio
10	0.9967	0.9967	0.0454
30	0.9918	0.9918	0.0448
50	0.9935	0.9935	0.0479
80	0.9898	0.9898	0.0672
100	0.9962	0.9962	0.0525

LREN3			
NREF	Media	Mediana	Desvio
10	0.9987	0.9987	0.0162
30	1.0008	1.0008	0.0295
50	1.0028	1.0028	0.0338
80	1.0011	1.0011	0.0404
100	1.0053	1.0053	0.0424

PETR4			
NREF	Media	Mediana	Desvio
10	0.974	0.974	0.1105
30	0.9511	0.9511	0.1637
50	0.9295	0.9295	0.2085
80	0.9496	0.9496	0.2207
100	0.9321	0.9321	0.1713

TIMP3			
NREF	Media	Mediana	Desvio
10	1.0014	1.0014	0.0551
30	0.9978	0.9978	0.0338
50	0.999	0.999	0.0808
80	0.9917	0.9917	0.0685
100	0.9867	0.9867	0.0567

TOTS3			
NREF	Media	Mediana	Desvio
10	1.0008	1.0008	0.0305
30	1.0003	1.0003	0.0269
50	0.998	0.998	0.0336
80	0.9986	0.9986	0.0572
100	0.9943	0.9943	0.0494

VALE5			
NREF	Media	Mediana	Desvio
10	1.0006	1.0006	0.0431
30	0.9989	0.9989	0.0408
50	0.9994	0.9994	0.0666
80	1.0102	1.0102	0.0966
100	0.9859	0.9859	0.0747

## E.6 NOBS

BRFS3			
NOBS	Media	Mediana	Desvio
5	1.0008	1.0008	0.0103
10	1.0014	1.0014	0.0174
20	1.0024	1.0024	0.0297
40	1.004	1.004	0.0521
80	0.9799	0.9799	0.0597

CPFE3			
NOBS	Media	Mediana	Desvio
5	0.999	0.999	0.0109
10	0.9965	0.9965	0.0248
20	0.9925	0.9925	0.0629
40	0.9972	0.9972	0.0549
80	1.0106	1.0106	0.0993

CYRE3			
NOBS	Media	Mediana	Desvio
5	0.9977	0.9977	0.0188
10	0.9931	0.9931	0.0395
20	0.99	0.99	0.0682
40	0.9986	0.9986	0.0796
80	1.018	1.018	0.0932

EMBR3			
NOBS	Media	Mediana	Desvio
5	1.0025	1.0025	0.0162
10	1.0036	1.0036	0.028
20	1.0088	1.0088	0.0415
40	0.9946	0.9946	0.0441
80	0.9704	0.9704	0.0878

ITUB4			
NOBS	Media	Mediana	Desvio
5	0.9991	0.9991	0.0113
10	0.9954	0.9954	0.0371
20	0.9956	0.9956	0.0502
40	0.989	0.989	0.0689
80	0.9758	0.9758	0.1076

LREN3			
NOBS	Media	Mediana	Desvio
5	1.0022	1.0022	0.0127
10	1.0028	1.0028	0.0249
20	1.006	1.006	0.039
40	0.9996	0.9996	0.046
80	0.9974	0.9974	0.0652

PETR4			
NOBS	Media	Mediana	Desvio
5	0.9587	0.9587	0.1347
10	0.935	0.935	0.1609
20	0.9151	0.9151	0.1991
40	0.9536	0.9536	0.2335
80	0.969	0.969	0.2494

TIMP3			
NOBS	Media	Mediana	Desvio
5	0.9987	0.9987	0.0241
10	0.9996	0.9996	0.0372
20	0.9854	0.9854	0.0722
40	0.9914	0.9914	0.1038
80	0.9885	0.9885	0.0778

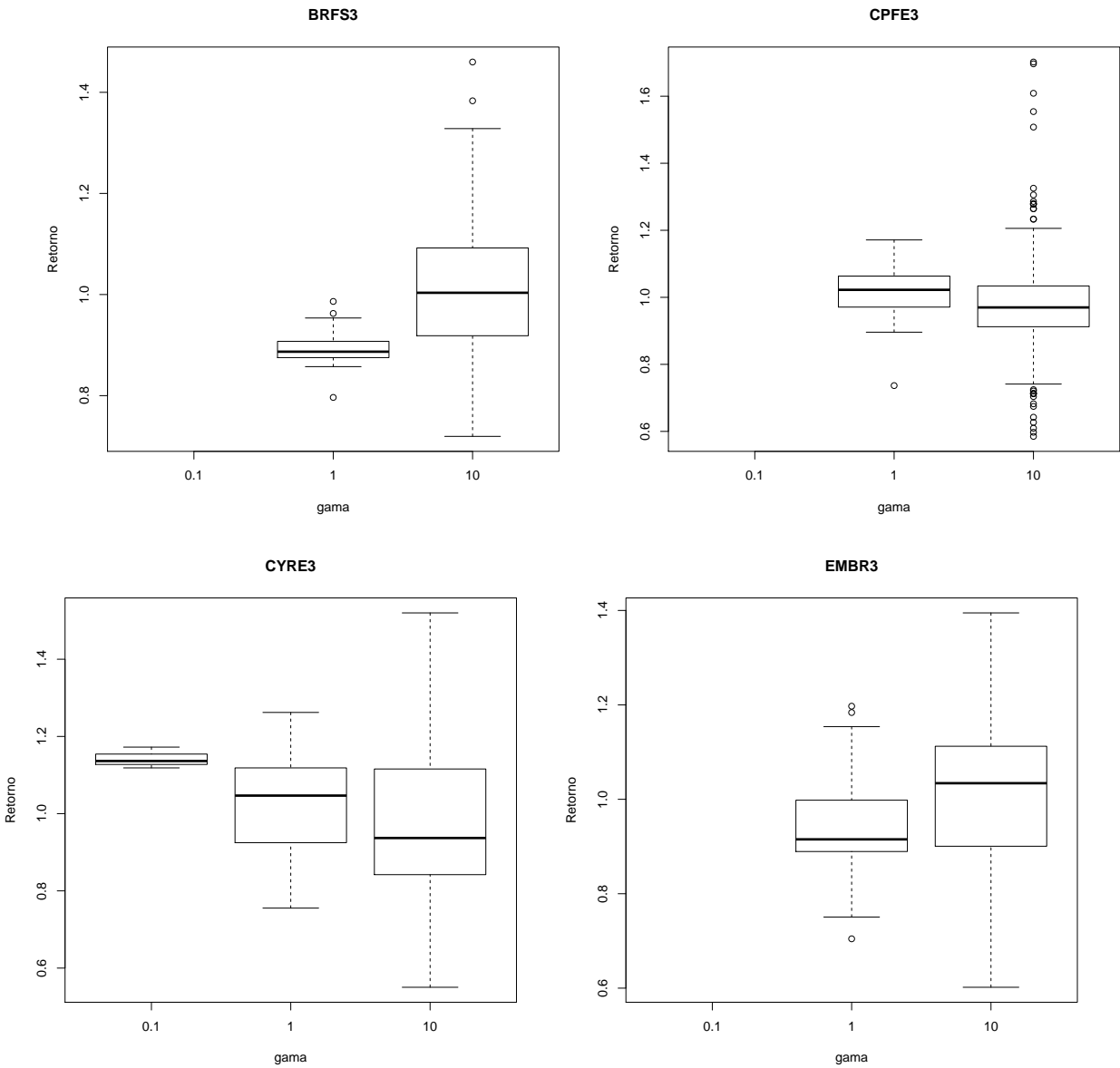
TOTS3			
NOBS	Media	Mediana	Desvio
5	0.9996	0.9996	0.0085
10	0.9964	0.9964	0.0302
20	0.9945	0.9945	0.0516
40	1.0024	1.0024	0.0504
80	0.9966	0.9966	0.0833

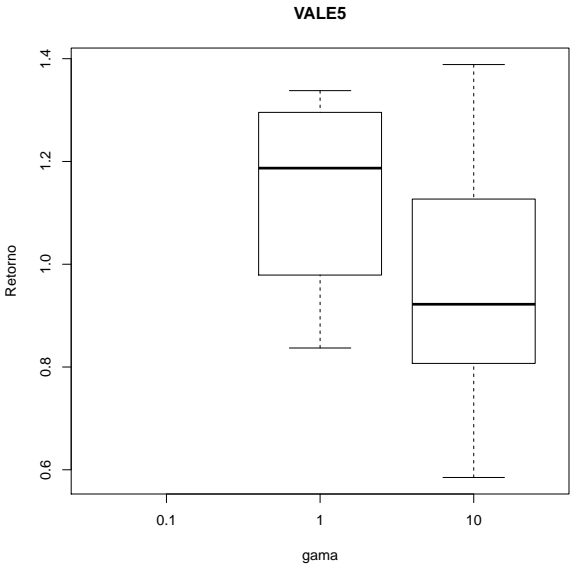
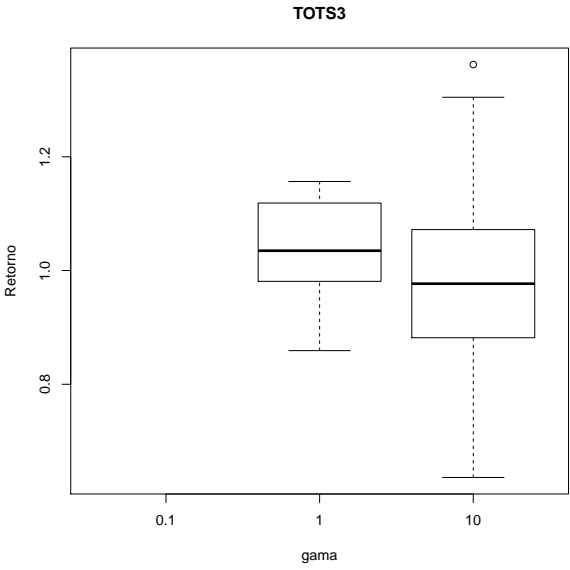
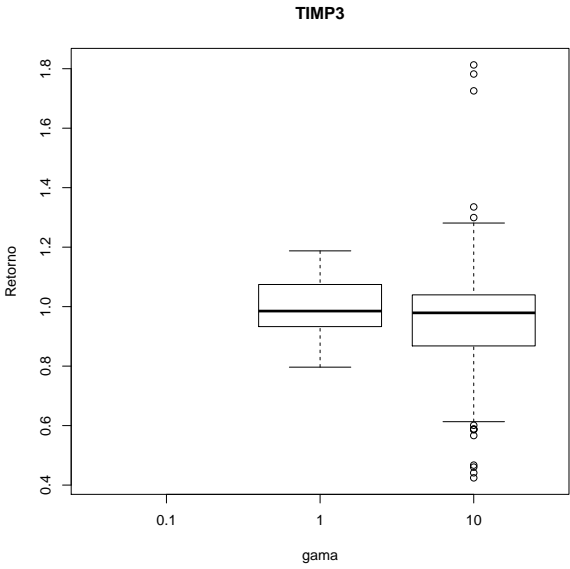
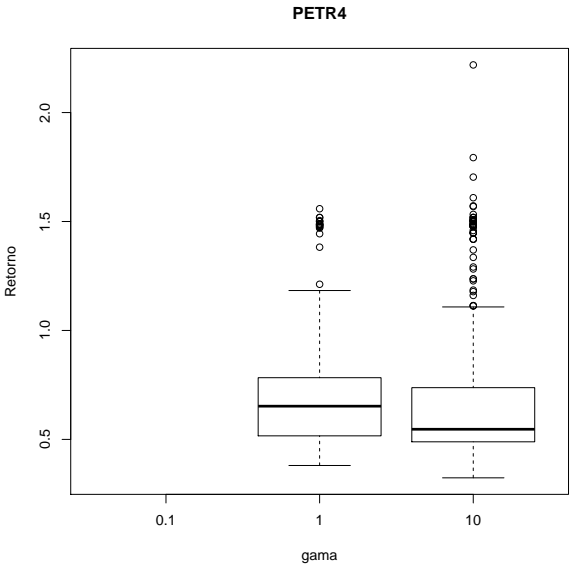
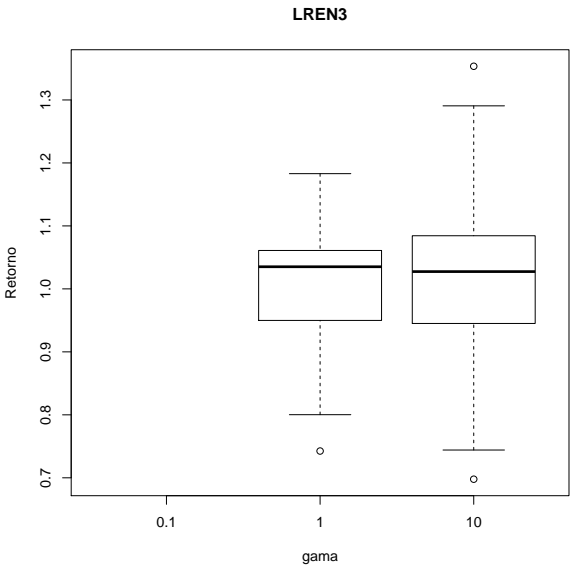
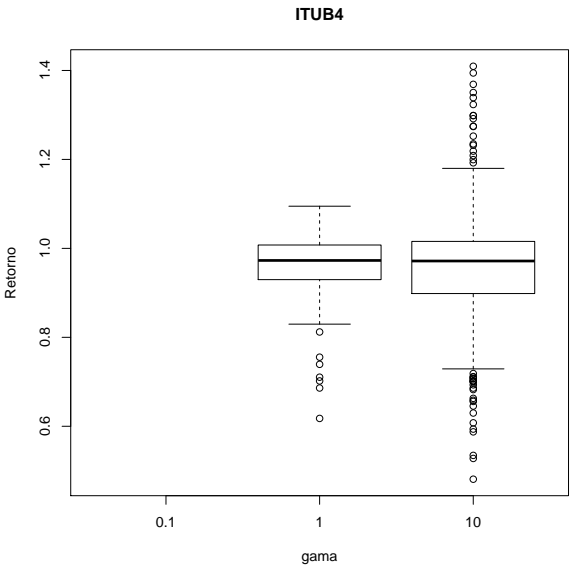
VALE5			
NOBS	Media	Mediana	Desvio
5	0.9942	0.9942	0.0341
10	0.9927	0.9927	0.0509
20	0.9907	0.9907	0.0686
40	0.9967	0.9967	0.0996
80	1.0445	1.0445	0.1234

# APÊNDICE F    BOXPLOTS DA VALIDAÇÃO DA ABORDAGEM

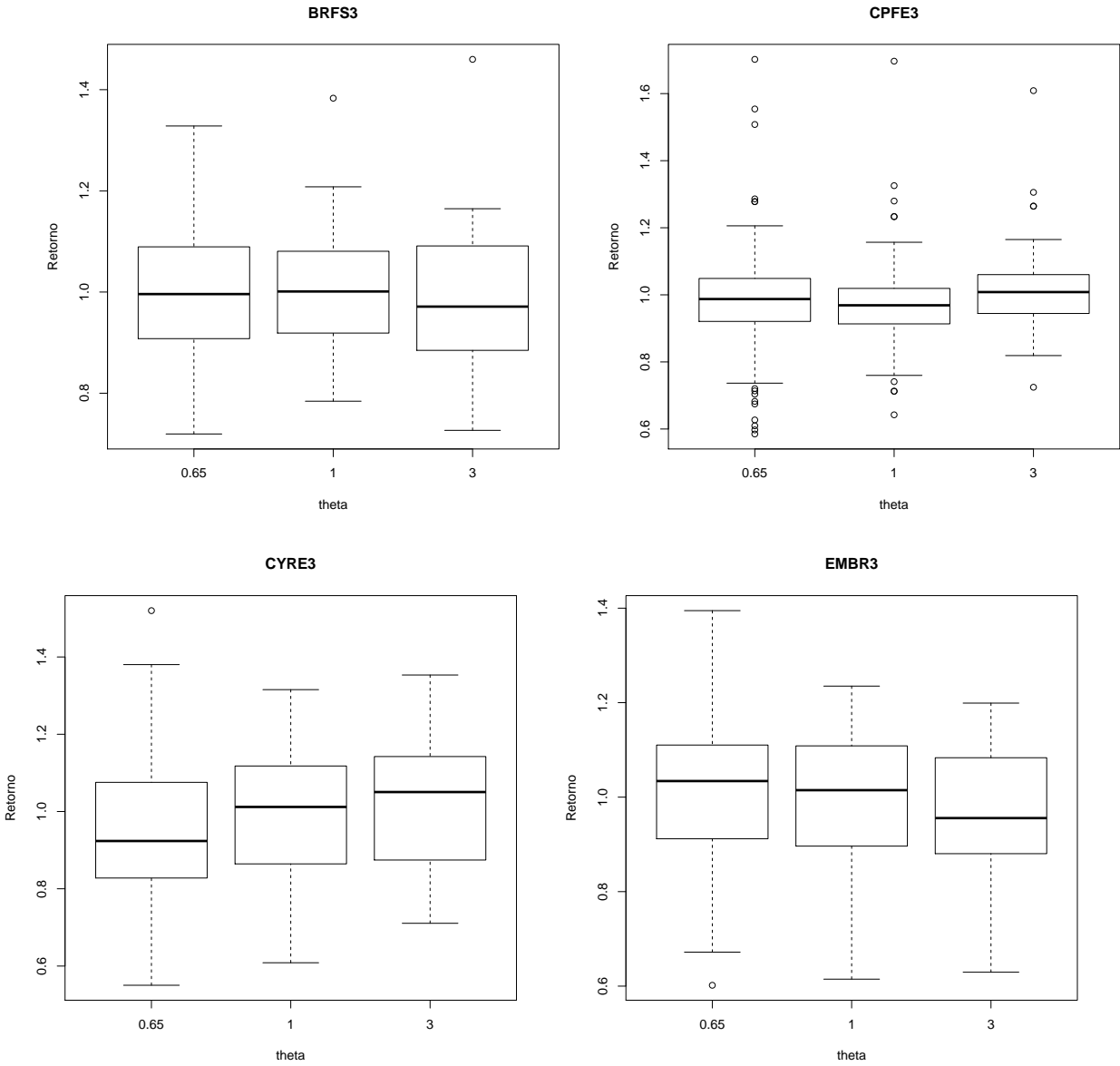
## 4

### F.1    gama

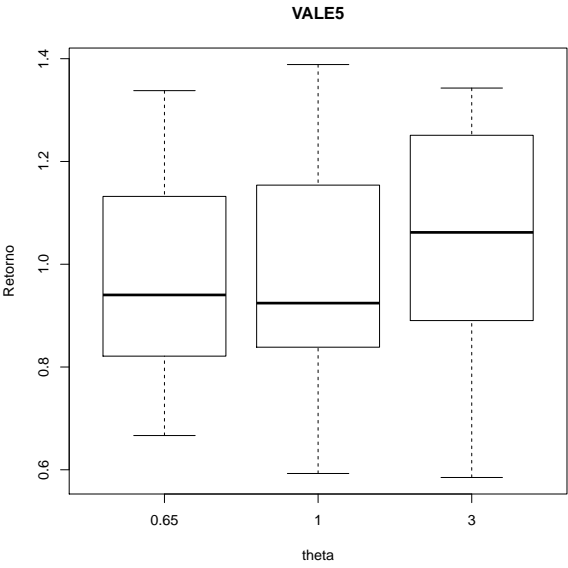
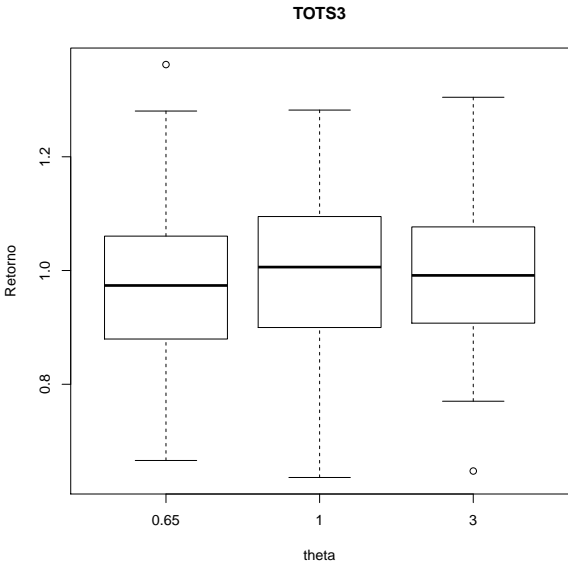
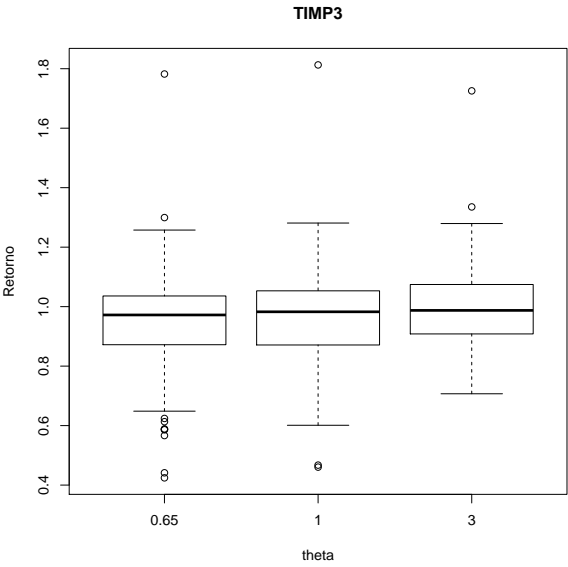
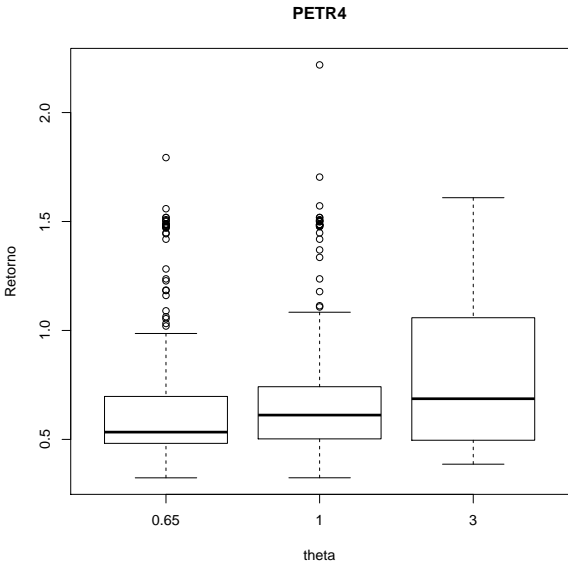
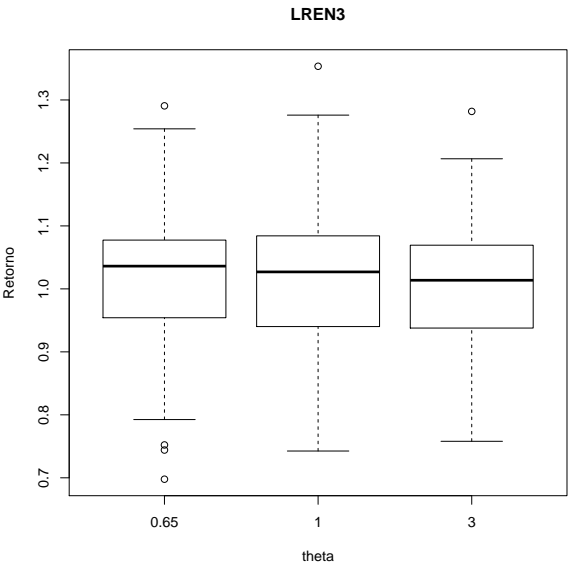
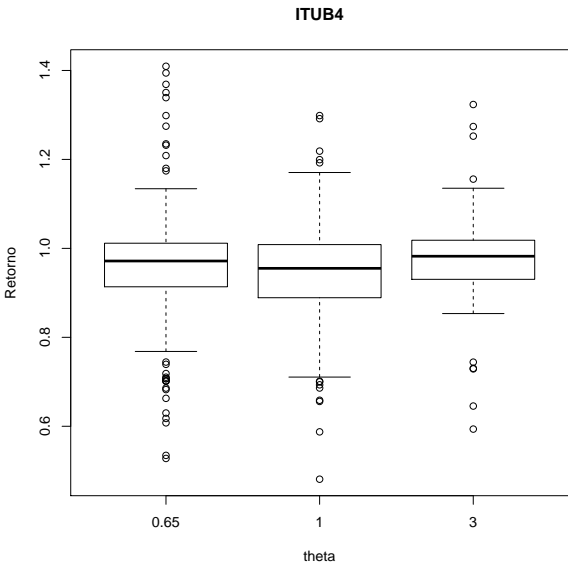




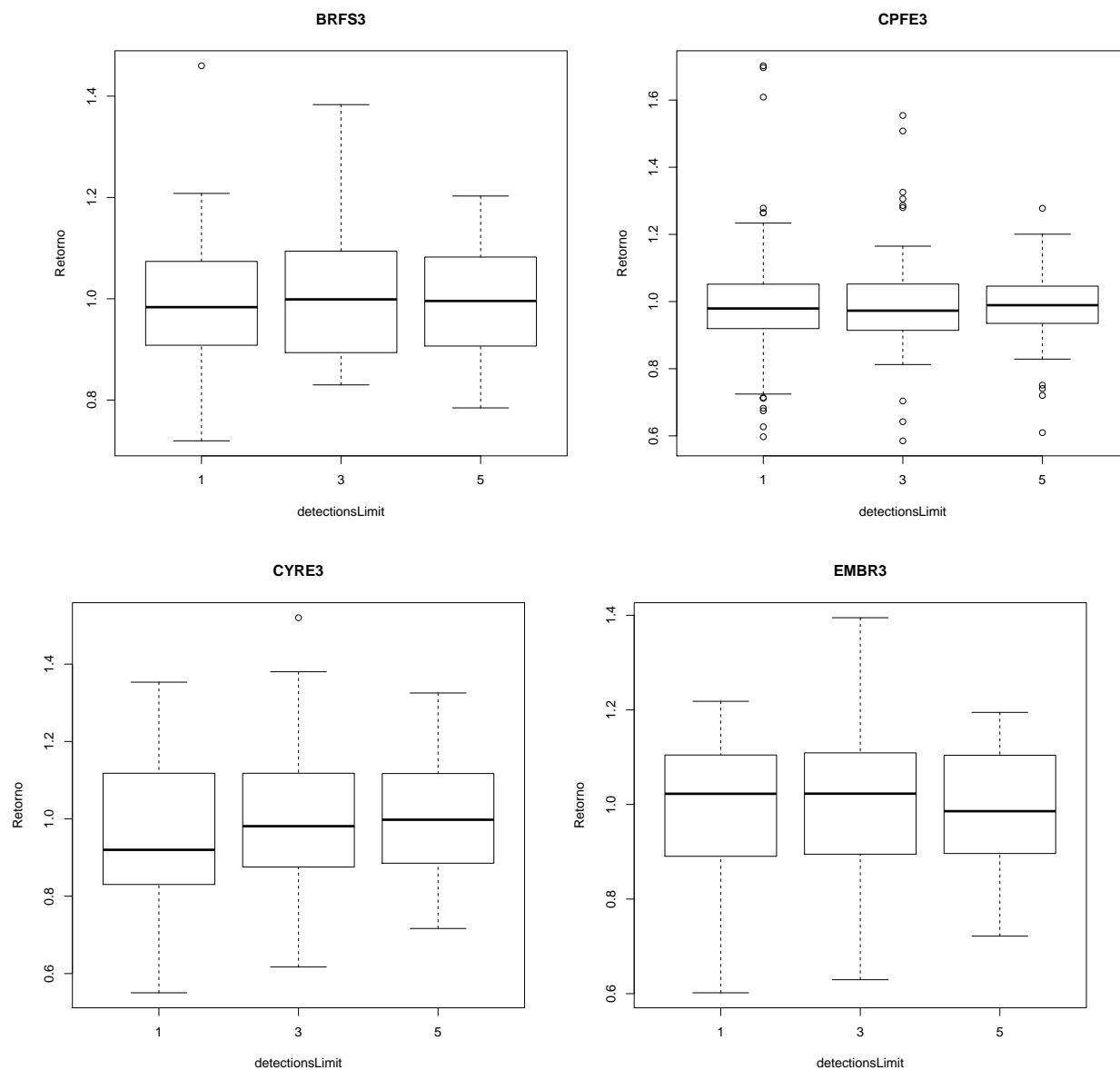
F.2 theta

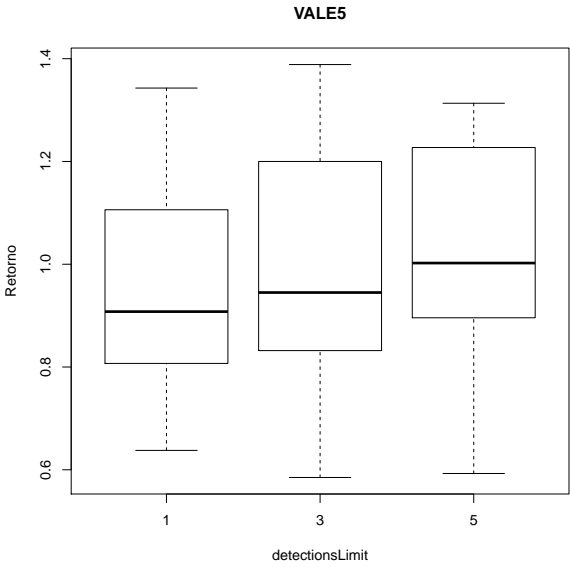
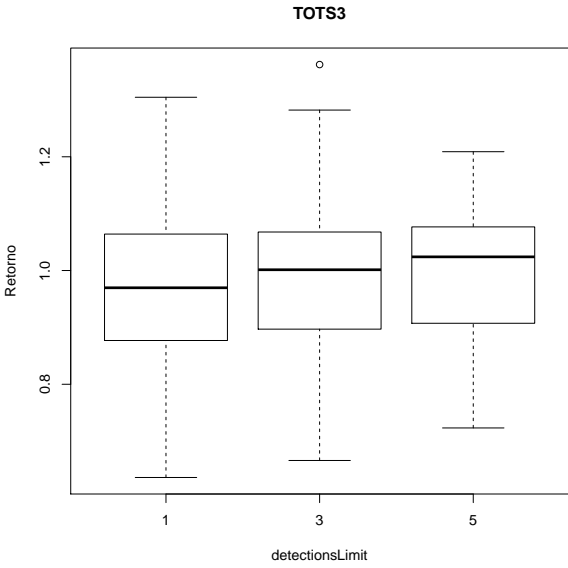
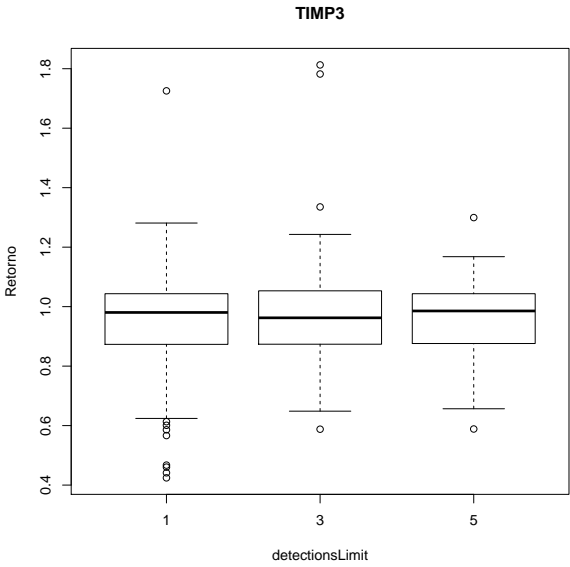
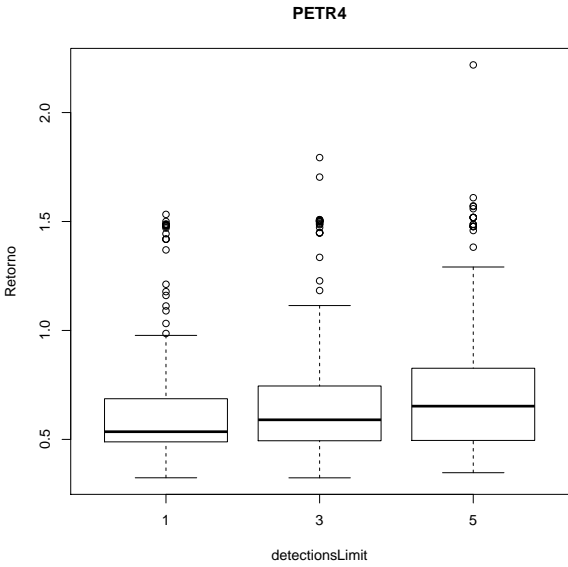
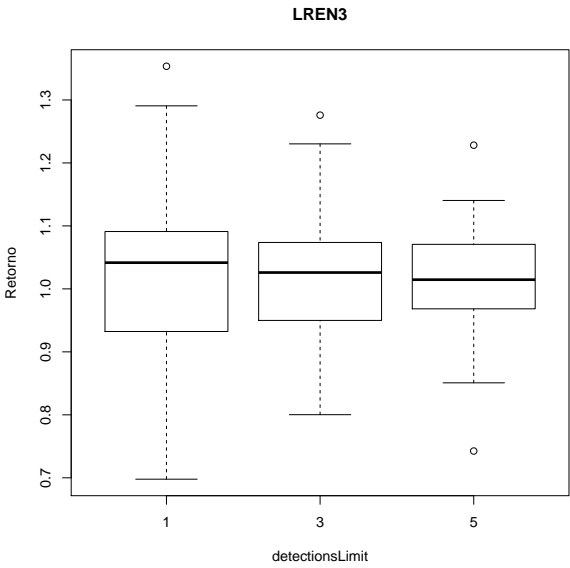
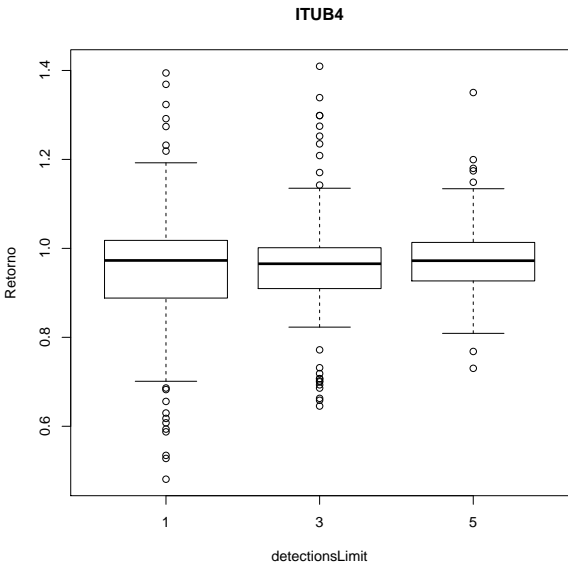




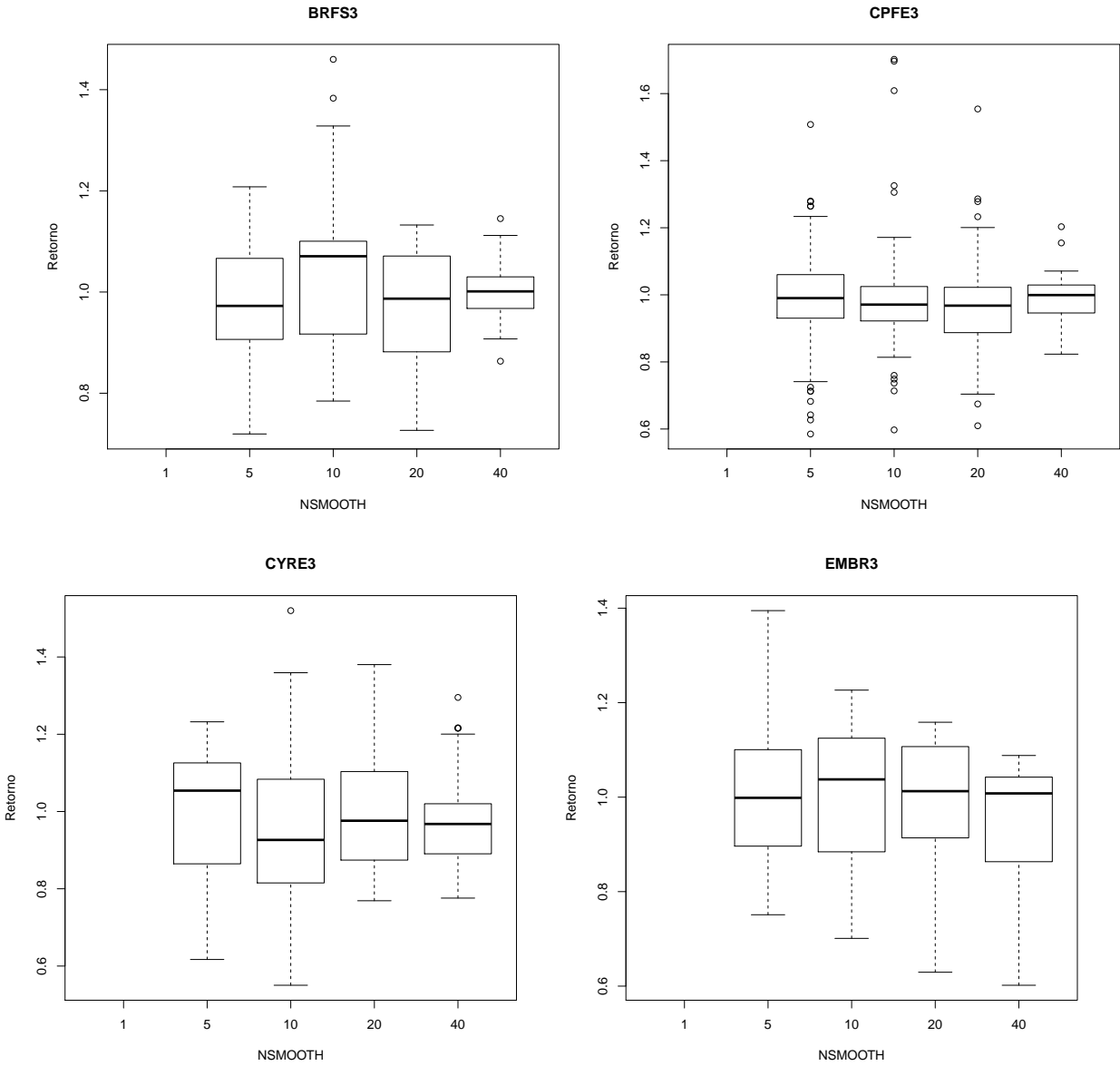


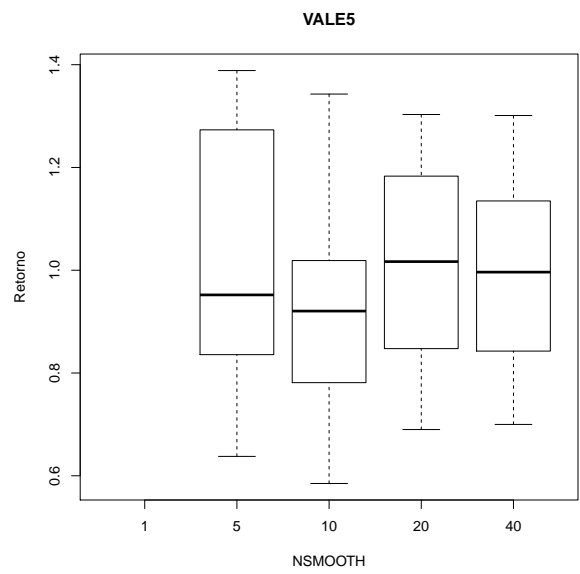
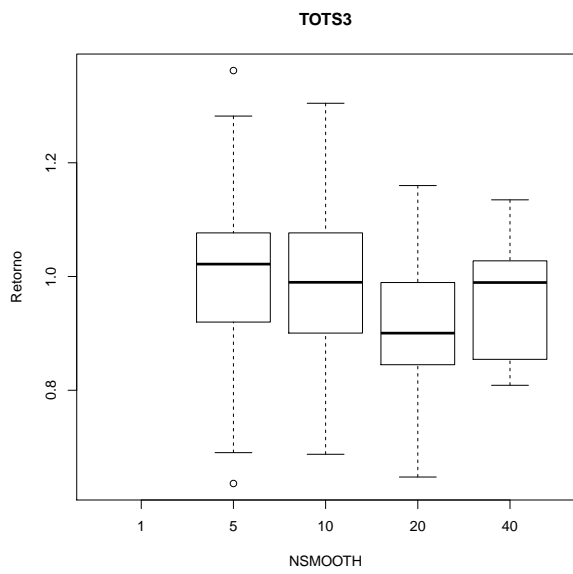
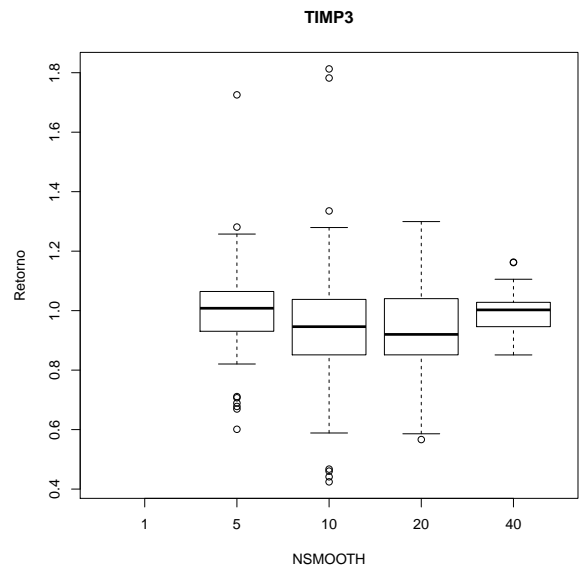
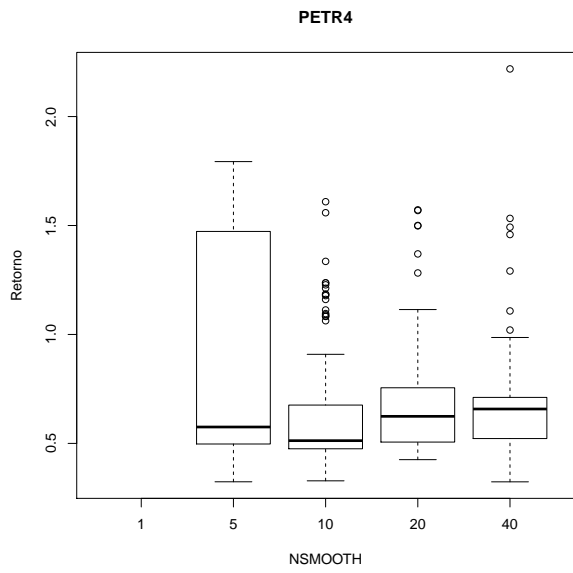
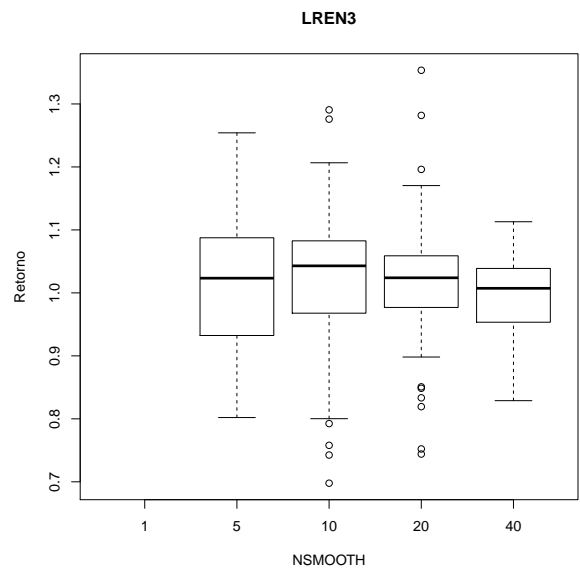
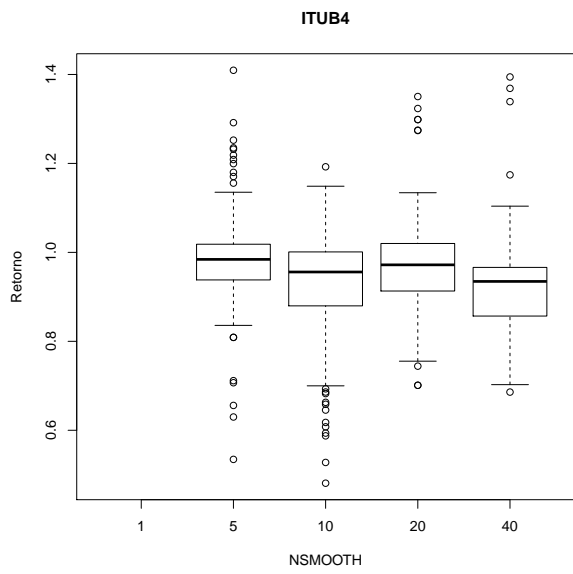
### F.3 detectionsLimit



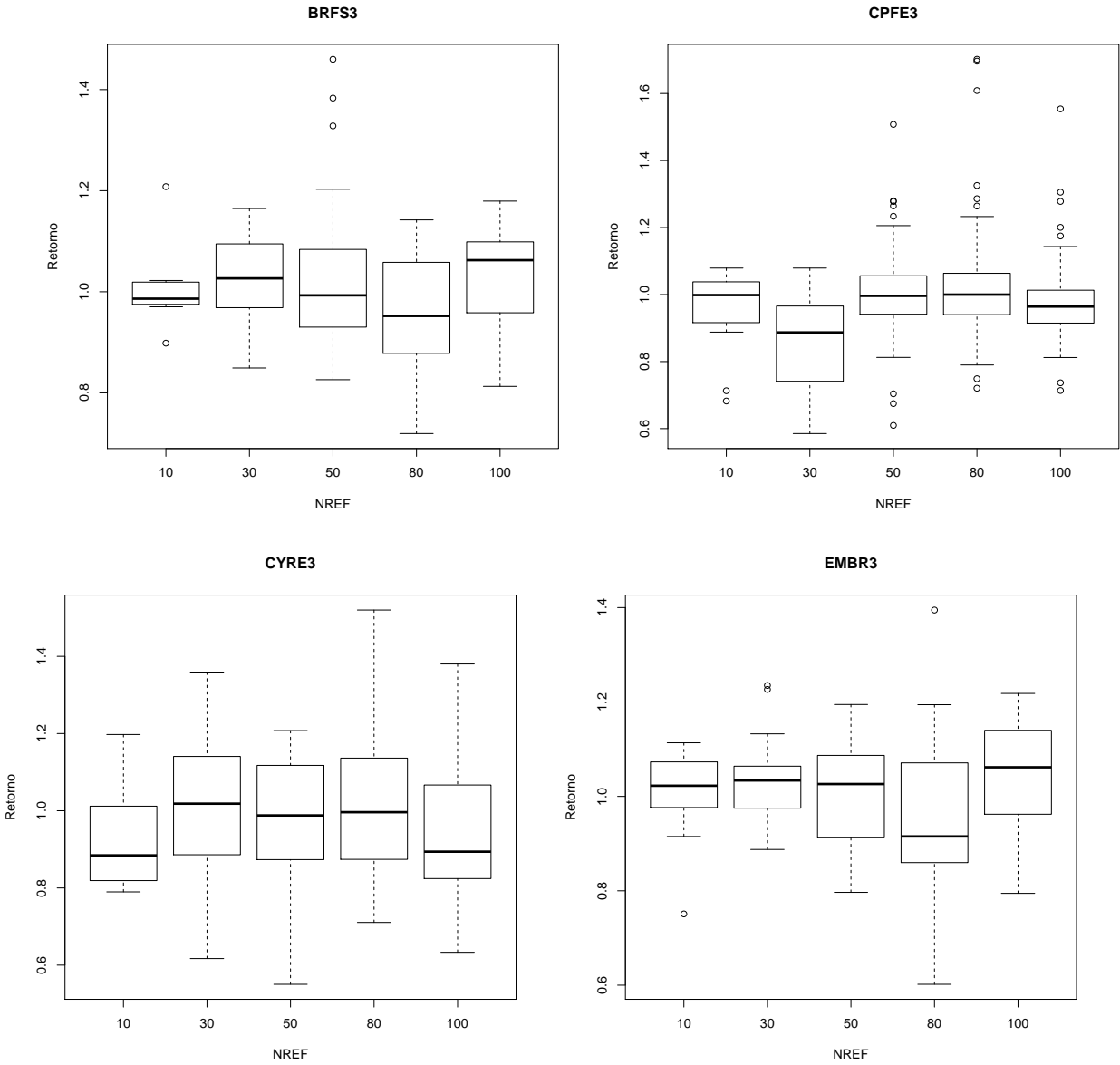


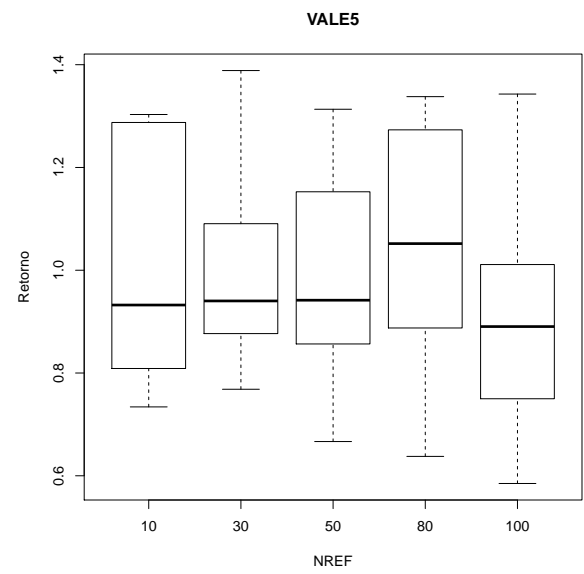
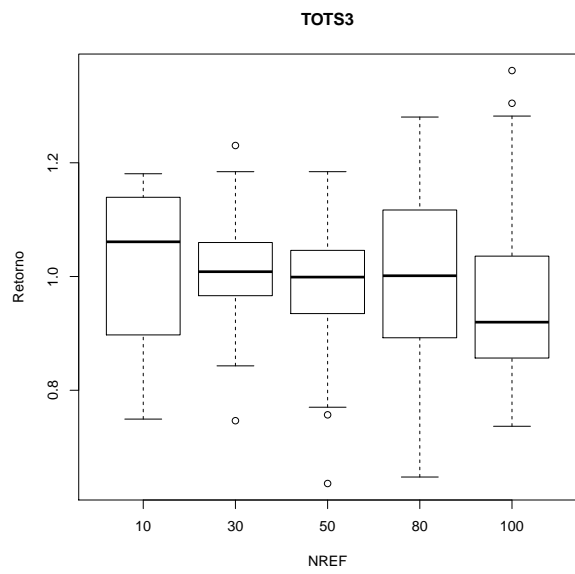
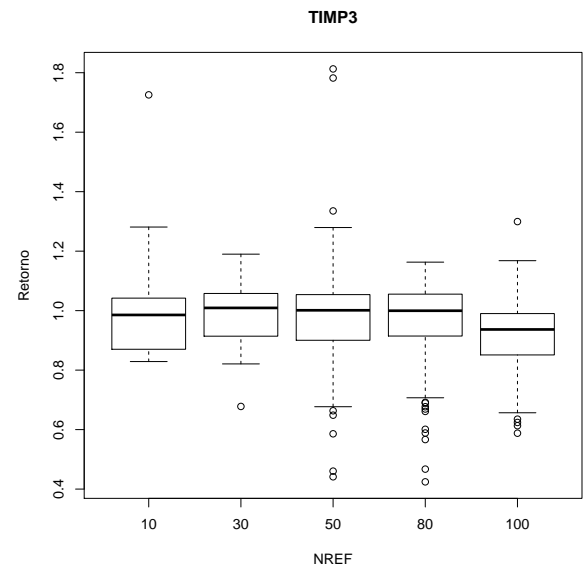
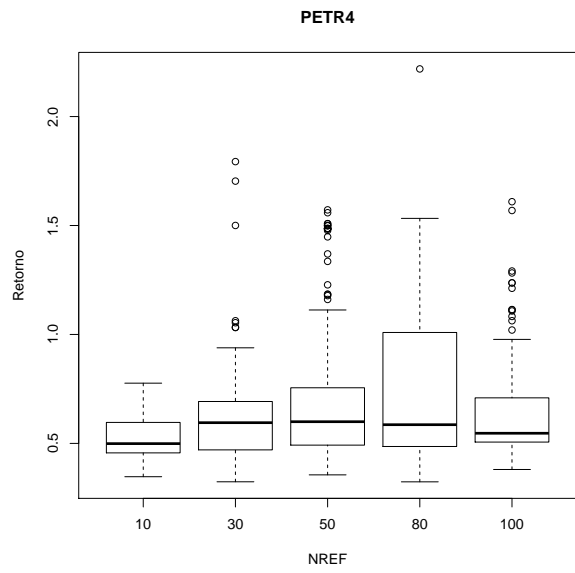
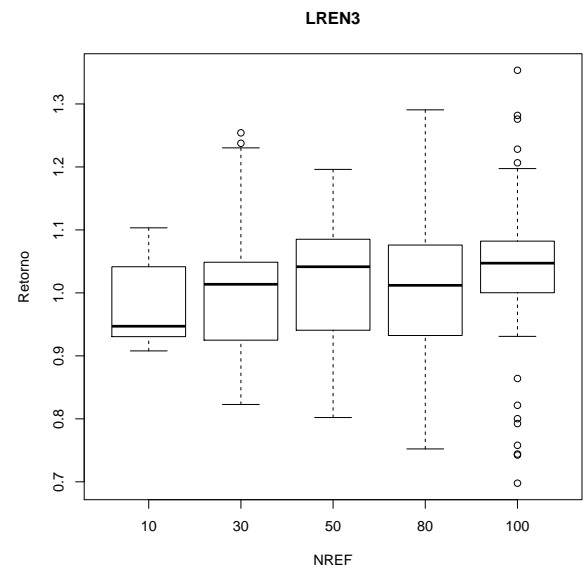
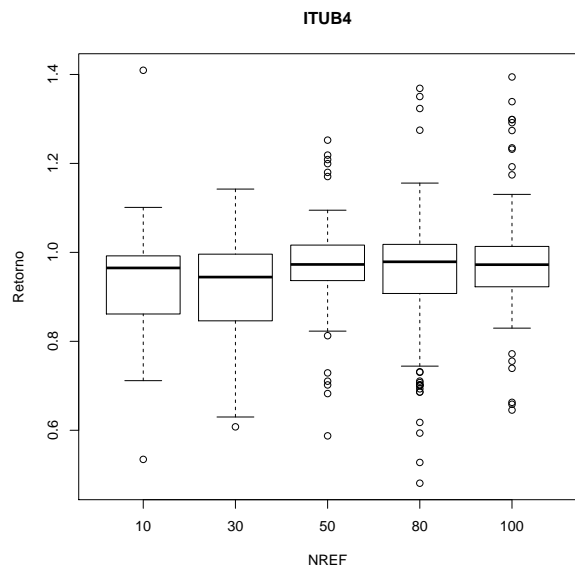
F.4 NSMOOTH



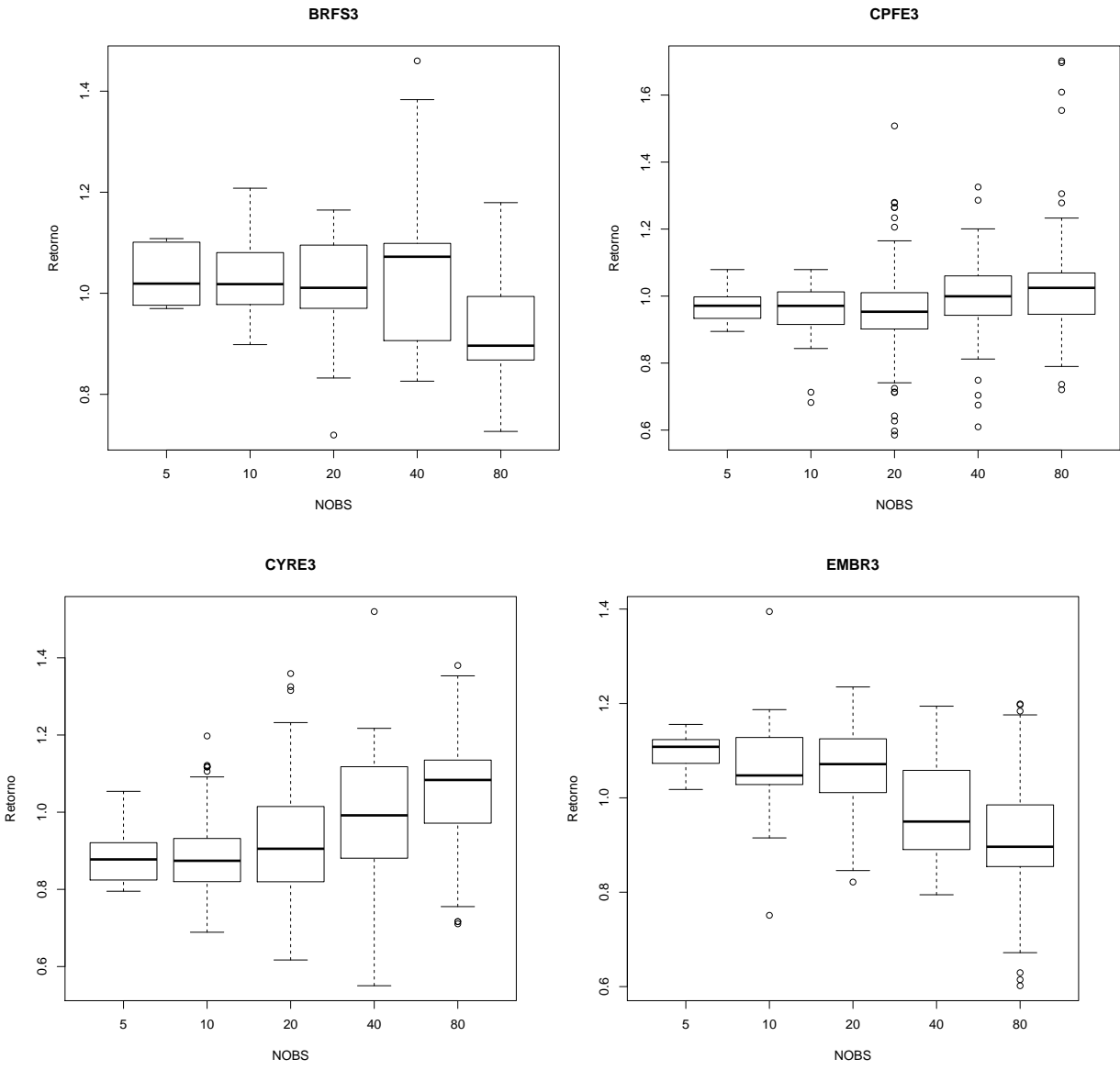


F.5 NREF

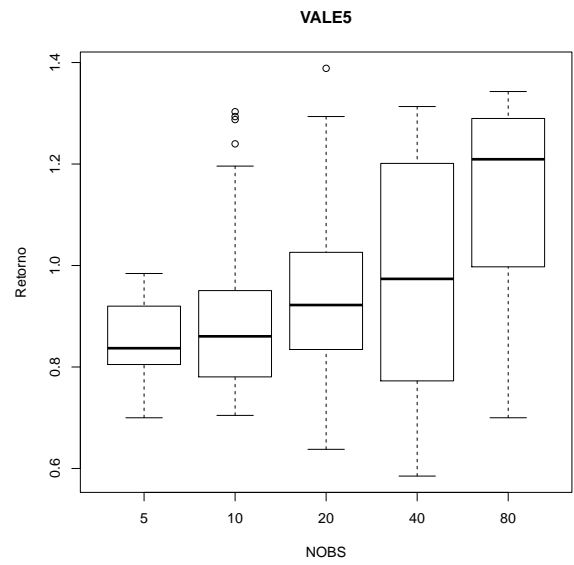
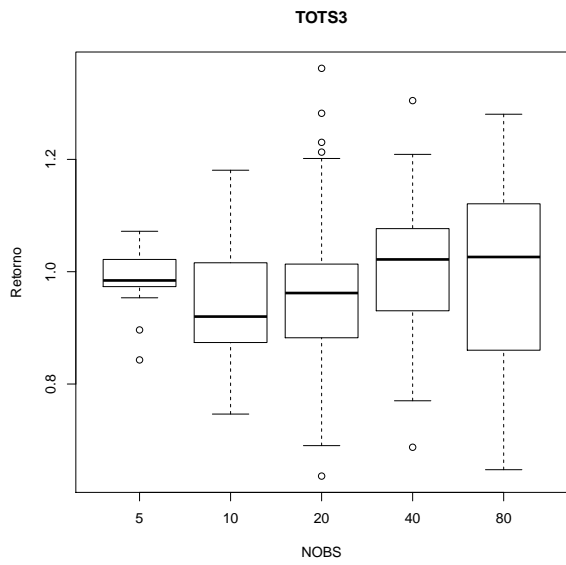
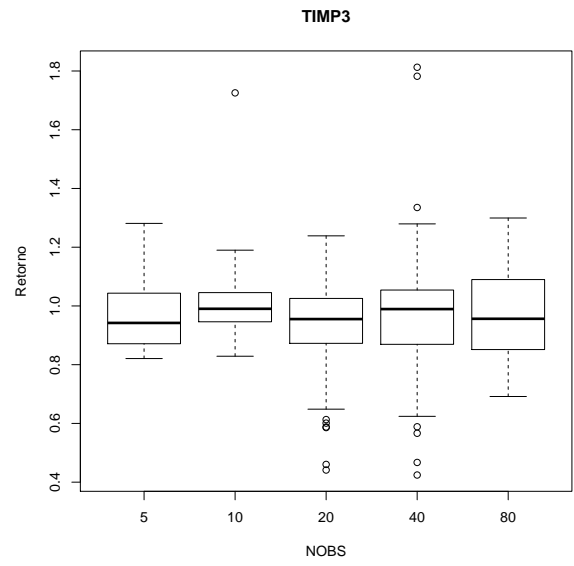
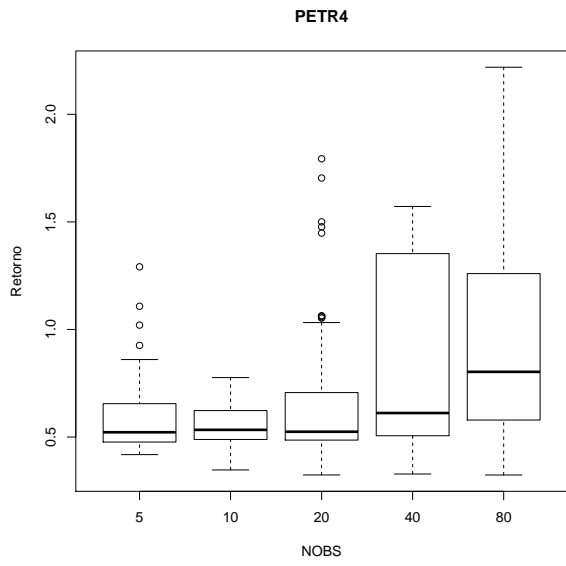
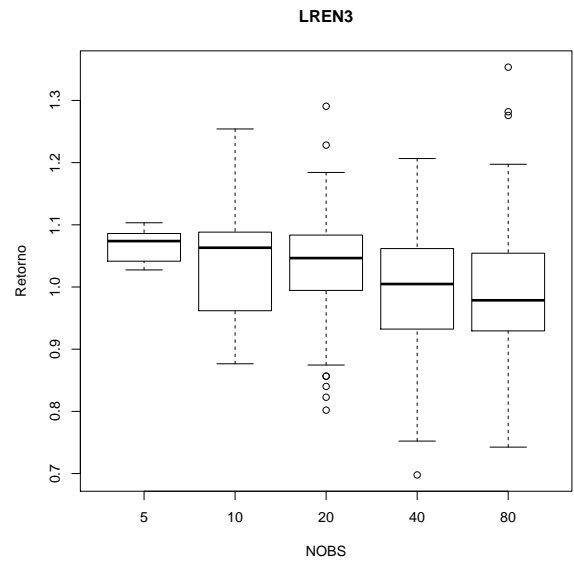
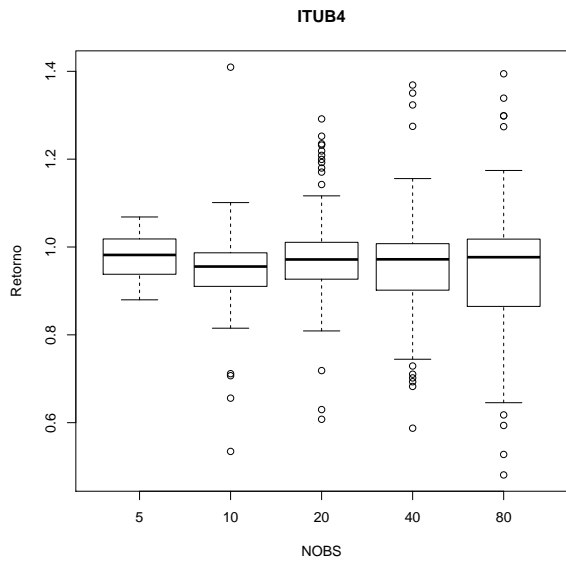




F.6 NOBS









**UNIVERSIDADE FEDERAL DE PELOTAS**  
Centro de Desenvolvimento Tecnológico  
Programa de Pós-Graduação em Computação



Dissertação

**Aprendizado de Máquina na Detecção de Tendências em  
Bolsa De Valores**

**MARCELO RUBENS GIESEL**

Pelotas, 2015