

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

# Uma Avaliação de Métodos de Previsão Aplicados à Grandes Quantidades de Séries Temporais Univariadas

**Tiago Ribeiro Pellegrini**

Orientador: Francisco Louzada Neto

Este exemplar corresponde à redação da  
Dissertação de Mestrado a ser devidamente  
corrigida e apresentada por Tiago Ribeiro  
Pellegrini e a ser avaliada pela comissão  
julgadora.

**São Carlos, 6 de dezembro de 2012**

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

P386am Pellegrini, Tiago Ribeiro.  
Uma avaliação de métodos de previsão aplicados à  
grandes quantidades de séries temporais univariadas /  
Tiago Ribeiro Pellegrini. -- São Carlos : UFSCar, 2012.  
76 f.

Dissertação (Mestrado) -- Universidade Federal de São  
Carlos, 2012.

1. Estatística. 2. Análise de séries temporais. 3. Previsão.  
I. Título.

CDD: 519.5 (20<sup>a</sup>)

**Tiago Ribeiro Pellegrini**

**Avaliação de Métodos de Previsão Aplicados a Grandes  
Quantidades de Séries Temporais Univariadas**

Dissertação apresentada à Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Aprovado em 06 de dezembro de 2012.

BANCA EXAMINADORA

Presidente

  
Prof. Dr. Francisco Louzada Neto (ICMC-USP/Orientador)

1º Examinador

  
Prof. Dr. Ricardo Sandes Ehlers (ICMC-USP)

2º Examinador

  
Profa. Dra. Maria Sílvia de Assis Moura (DEs-UFSCar)

*"A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê."*

*Arthur Schopenhauer*

## Agradecimentos

Primeiramente agradeço a Deus, por sua influência em minha vida e por ser minha direção espiritual durante este estágio de seu plano.

Minha esposa Morgana, amiga e companheira eterna, que em muitos momentos de ansiedade e estresse usou de longanimidade e paciência, me apoiando e me fortalecendo em todas as direções.

Meus pais que tanto me incentivaram e me apoiaram em minhas decisões na vida, especialmente no decorrer de minha graduação, e sem dúvida são as pessoas com quem posso sempre contar.

Não poderia deixar de agradecer a meu orientador, o professor Doutor Francisco Louzada Neto, por todo o apoio e horas dedicadas a me orientar neste trabalho.

Aos professores Maria Sílvia de Assis Moura, Márcio Alvez Diniz e Ricardo Sandes Ehlers, membros da banca do exame de qualificação e defesa, pelas sugestões feitas.

Aos funcionários do Departamento de Estatística, em especial a Isabel, secretária do departamento, por estar sempre pronta a ajudar e por propiciar boas condições de trabalho.

A todas as pessoas que, direta ou indiretamente, contribuíram para a realização dessa dissertação.

Finalmente, agradeço à CAPES pelo auxílio concedido para realização deste trabalho.

## Abstract

**Palavras-chave:** *forecasting, time series, Theta model, ARIMA models, exponential smoothing.*

Time series forecasting is probably one of the most primordial interests on economics and econometrics, and the literature on this subject is extremely vast. Due to technological growth in recent decades, large amounts of time series are daily collected; which, in a first moment, it requires forecasts according a fixed horizon; and on the second moment the forecasts must be constantly updated, making it impractical to human interaction. Towards this direction, computational procedures that are able to model and return accurate forecasts are required in several research areas. The search for models with high predictive power is an issue that has resulted in a large number of publications in the area of forecasting models. We propose to do a theoretical and applied study of forecasting methods applied to multiple univariate time series. The study was based on exponential smoothing via state space approach, automatic ARIMA methods and the generalized Theta method. Each model and method were applied in large data bases of univariate time series and the forecast errors were evaluated. We also propose an approach to estimate the Theta coefficients, as well as a redefinition of the method regarding the number of decomposition lines, extrapolation methods and a combining approach.

## Resumo

**Palavras-chave:** *previsão, séries temporais, modelo Theta, modelos ARIMA, alisamento exponencial.*

A previsão de séries temporais é provavelmente um dos interesses mais primordiais na área de economia e econometria, e a literatura referente a este assunto é extremamente vasta. Devido ao crescimento tecnológico nas últimas décadas, diariamente são geradas e disponibilizadas grandes quantidades de séries temporais; que em um primeiro momento, requerem previsões de acordo com um horizonte fixado; e no segundo momento as previsões precisam ser constantemente atualizadas, tornando pouco prática a interação humana. Desta forma, procedimentos computacionais que modelem e posteriormente retornem previsões acuradas são exigidos em diversas áreas do conhecimento. A busca por modelos com alto poder de preditivo é uma questão que tem resultado em grande quantidade de publicações na área de modelos para previsão. Neste trabalho, propõe-se um estudo teórico e aplicado de métodos de previsão aplicado à múltiplas séries temporais univariadas. O estudo foi baseado em modelos de alisamento exponencial via espaço de estados, método ARIMA automático e o método Theta generalizado. Cada modelo e método foi aplicado em grandes bases de séries temporais univariadas e avaliado o resultado em relação aos erros de previsão. Também foi proposta uma abordagem para estimação dos coeficientes Theta, assim como redefinição do método em relação a quantidade de linhas para decomposição, métodos de extrapolação e combinação das linhas para previsão.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Objetivos do Trabalho . . . . .	4
<b>2</b>	<b>Modelos de Alisamento Exponencial</b>	<b>5</b>
2.1	Modelos de Alisamento Exponencial . . . . .	6
2.1.1	Equações Pontuais . . . . .	7
2.2	Modelos de Espaço de Estados . . . . .	8
2.3	Cálculo das Equações de Espaço de Estados Para o Modelo Linear de Holt	11
2.3.1	Erros Aditivos . . . . .	12
2.3.2	Erros Multiplicativos . . . . .	12
2.4	Inicialização e Estimação . . . . .	13
2.4.1	Inicialização . . . . .	13
2.4.2	Estimação . . . . .	14
2.4.3	Intervalos de Confiança Para as Previsões . . . . .	15
2.5	Método ETS Automático . . . . .	17
<b>3</b>	<b>Métodos ARIMA Automáticos</b>	<b>19</b>
3.1	Uma Breve Revisão . . . . .	20
3.2	O Procedimento Stepwise Para Modelos ARIMA . . . . .	22
3.3	ARIMA vs Modelos de Alisamento Exponencial . . . . .	24
<b>4</b>	<b>O Método Theta Para Previsão de Séries Temporais</b>	<b>26</b>
4.1	Um Método de Decomposição de Séries Temporais . . . . .	26
4.1.1	Construção do Método Theta Generalizado . . . . .	28
4.1.2	Construção do Método a Partir de Duas Linhas-Theta . . . . .	32



4.2	Redefinindo o Método Theta . . . . .	33
4.2.1	Número de linhas-Theta para decomposição . . . . .	34
4.2.2	Métodos de Extrapolação . . . . .	34
4.2.3	Combinando as linhas-Theta . . . . .	35
4.3	Estimação dos Parâmetros do Método Theta . . . . .	36
4.4	Intervalos de Confiança para Previsão do Método Theta . . . . .	37
<b>5</b>	<b>Avaliando a Acurácia de Previsão</b>	<b>39</b>
5.1	Uma revisão sobre medidas de acurácia . . . . .	39
5.1.1	Medidas Escala-Dependentes . . . . .	39
5.1.2	Medidas Baseadas em Erros Percentuais . . . . .	40
5.1.3	Medidas Baseadas em Erros Relativos . . . . .	41
5.1.4	Medidas Relativas . . . . .	41
5.1.5	Medidas de Erro de Escala . . . . .	42
<b>6</b>	<b>Resultados</b>	<b>44</b>
6.1	As bases de dados <i>M1</i> e <i>M3-Competition</i> . . . . .	45
6.2	Resultados Para os Modelos de Alisamento Exponencial . . . . .	47
6.3	Resultados para os métodos ARIMA Automáticos . . . . .	50
6.4	Resultados Para o Método Theta . . . . .	51
6.4.1	Aplicação à base de dados M1 . . . . .	52
6.4.2	Aplicação à Base de Dados M3 Para o Método Theta Generalizado	54
<b>7</b>	<b>Considerações Finais</b>	<b>57</b>
<b>A</b>	<b>Tabelas resultados alisamento exponencial</b>	<b>60</b>
<b>B</b>	<b>Tabelas resultados métodos ARIMA Automáticos</b>	<b>62</b>
<b>C</b>	<b>Tabelas resultados método Theta generalizado</b>	<b>64</b>
<b>D</b>	<b>Provas</b>	<b>67</b>
<b>E</b>	<b>Códigos Utilizados Para Programação no Software R</b>	<b>69</b>

# Capítulo 1

## Introdução

O crescimento tecnológico das últimas décadas tornou acessível a troca e acúmulo de informações, em outras palavras, dados. Hoje, muitas organizações geram e coletam diariamente grandes volumes de dados através de operações computacionais, que como consequência necessitam de análise estatística. Surge então a necessidade de desenvolvimento de metodologias computacionais que automatizem alguns procedimentos estatísticos voltados a organização, sumarização e modelagem dos dados, o que tem refletido grande investimento em pesquisa nos últimos anos. Um bom exemplo voltado ao contexto de modelagem é a previsão de grandes quantidades de séries temporais. Por exemplo, preços de ações minuto a minuto, previsão da temperatura hora após hora, chegada de pacientes em clínicas médicas diariamente, produção semanal de um determinado produto, desemprego mês a mês em determinada região, volume de importação trimestral de um país, volume de negócios anual de determinada empresa. A quantidade de séries temporais aumenta a cada minuto, hora, dia, mês, trimestre ou ano, e em geral utilizar a informação passada a fim de prever o futuro é a primeira questão que surge durante a análise. Segundo Leonard (2005), prever grandes quantidades de séries temporais tornou-se uma problema atual. Dentro deste contexto geralmente a quantidade de séries temporais é enorme e as previsões precisam ser atualizadas constantemente.

Muitas metodologias têm sido apresentadas para modelar uma série temporal. Dentre as mais utilizadas destacam-se os modelos de regressão, os modelos de alisamento exponencial, introduzidos a partir dos trabalhos de Holt (1957), Brown (1959) e Winters (1960) que modelam a série com base nas componentes de erro, tendência e sazonalidade.

dade, e os modelos ARIMA introduzidos por Box e Jenkins (1970), que modelam tanto não estacionaridade homogênea quanto modelos estacionários e compõe uma classe mais abrangente de modelos. Posteriormente a apresentação destes modelos no contexto de previsão, Armstrong (1984) aponta para questões que começam a surgir no meio acadêmico:

- Existe um modelo que em geral resulte previsões melhores em relação a outros modelos?
- Modelos mais complexos resultam em previsões mais acuradas em relação a modelos mais simples?
- Procedimentos de combinação de modelos propiciam um aumento no poder preditivo?

Com o objetivo de responder a essas questões surgem as chamadas competições de modelos para previsão. Em resumo, as competições de modelos são comparações empíricas entre vários métodos ou modelos, com o objetivo de classificar um método baseado em sua performance de previsão. A primeira competição de modelos ocorre no início da década de 80 com o trabalho de Makridakis et al. (1982). Neste trabalho os autores consideraram 24 diferentes métodos para previsão, dentre eles; modelos de regressão e modelos de alisamento exponencial. Todos os métodos foram aplicados à 1001 séries temporais univariadas, esta base de dados foi chamada de *M1-Competition*. Naquela época, os desafios computacionais eram consideráveis; a quantidade de métodos e séries temporais envolvidos no trabalho foram uma inovação em pesquisa e um indício da necessidade em automatizar o procedimento de previsão.

Uma das competições que resultou maior impacto no meio acadêmico devido a quantidade de citações é a terceira competição de modelos, chamada *M3-Competition* (Makridakis e Hibon, 2000). Esta competição trouxe inovações em relação a quantidade de séries temporais para previsão, os métodos empregados na competição e a quantidade de medidas envolvidas para avaliar o erro de previsão. A competição reuniu 24 procedimentos de previsão aplicados à 3003 séries temporais univariadas; dentre estes, modelos de alisamento exponencial, modelos ARIMA, modelos de decomposição de séries temporais e pacotes comerciais para previsão de séries temporais. Os resultados da competição surpreenderam, por mostrarem que métodos simples desenvolvidos para retornar previsões

de forma prática, como os conhecidos modelos de alisamento exponencial, com destaque para o modelo de alisamento exponencial de Holt com *Dampen*, em muitas vezes superaram os resultados de métodos sofisticados como *Box-Jenkins Automatic*, *Forecast Pro* e algoritmos de redes neurais.

O bom desempenho dos modelos de alisamento exponencial propiciou um avanço em sua estrutura de modelagem. Através dos trabalhos de Ord et al. (1997), Hyndman et al. (2002) e Taylor (2003), uma nova estrutura estatística foi desenvolvida para os atuais 30 modelos de alisamento exponencial a partir de filtros de espaço de estados. Tal abordagem incorpora uma classe paramétrica aos modelos de alisamento, permitindo-se obter cálculo de verossimilhança, intervalos para previsão e procedimentos para seleção de modelos. "Desta forma esses modelos são trazidos para a mesma classe de modelos ARIMA, sendo amplamente aplicáveis e com base estocástica para obtenção das previsões" (Hyndman et al., 2002). Os artigos recém mencionados têm mostrado que o emprego destes modelos em áreas como economia e econometria geram ótimas previsões para todos os modelos de alisamento exponencial, incluindo a inovação de modelos não lineares de alisamento exponencial (Athanasopoulos et al., 2011).

Procedimentos de seleção de modelos ARIMA estão sempre presentes nas competições de previsão. Dentre os procedimentos mais conhecidos estão o pacote comercial *Forecast Pro* (Goodrich, 2000) e o procedimento Auto Arima proposto por Hyndman e Khandakar (2008), disponível no pacote *forecast* do software *R*.

Outro método apresentado na *M3-Competition* que despertou a atenção de pesquisadores devido a sua excelente performance foi o método de decomposição Theta desenvolvido por Assimakopoulos e Nikolopoulos (2000). O método Theta aborda de forma simples e diferenciada a decomposição de séries temporais e teve o melhor desempenho de previsão para séries que apresentam periodicidade. Muitos estudos têm sido realizados desde então, buscando generalizar o método a fim de produzir previsões melhores do que as apresentadas na competição M3 (e.g., Nikolopoulos e Assimakopoulos, 2005, Nikolopoulos et al., 2012a, 2008). No entanto, existem muitas questões abertas referentes ao método, de forma que, se respondidas, podem redefinir o método Theta. Tais questões serão abordadas neste trabalho. O método Theta também tornou-se um método de referência em várias competições de previsão (e.g., Athanasopoulos et al., 2011, Crone et al., 2011).

## 1.1 Objetivos do Trabalho

O objetivo principal deste trabalho é realizar um estudo teórico e aplicado de métodos de previsão de séries temporais com aplicação à grandes quantidades de séries temporais univariadas. Este trabalho investigará os seguintes procedimentos de previsão:

- Modelos de alisamento exponencial via espaço de estados, de acordo com os trabalhos de Hyndman et al. (2002) e Hyndman et al. (2008);
- Procedimentos ARIMA automáticos, baseados na proposta de Hyndman e Khandakar (2008);
- Método Theta generalizado proposto por Nikolopoulos et al. (2012b).

A aplicação será realizada no software R para todos os procedimentos abordados no trabalho, considerando estratégias de seleção de modelos, avaliação do erro de previsão e tempo computacional gasto por cada procedimento utilizado. As rotinas programadas pelo autor podem ser requisitadas por meio do endereço eletrônico *tiagorpe@gmail.com*.

O objetivo específico deste trabalho propõe um estudo empírico do método Theta generalizado, com o objetivo de melhorar seu desempenho de previsão baseado nas componentes que compõe o método, o qual serão estudadas no Capítulo 4. Propõe-se também um procedimento de estimação para os coeficientes do método Theta e avaliação de seu desempenho.

Este trabalho está organizado da seguinte maneira. O Capítulo 2 apresenta o objetivo do trabalho. O Capítulo 3 aborda os modelos de alisamento exponencial via espaço de estados e as recentes inovações em pesquisa nesta área. No Capítulo 4 é feita uma revisão bibliográfica de modelos ARIMA e apresentado algoritmos automáticos de previsão para esta classe de modelos. O Capítulo 5 introduz o método Theta e sua crescente importância na área de previsão, também são discutidas suas propriedades além de propostas de inovações para a construção do método Theta generalizado. O Capítulo 6 discute, a partir de um ponto de vista crítico, as métricas utilizadas para avaliação da acurácia de previsão. No Capítulo 7 são apresentados os resultados da previsão dos modelos discutidos neste trabalho aplicados à base de dados *M1* e *M3*. O Capítulo 8 trás os comentários finais e conclusões referentes a este trabalho.

## Capítulo 2

# Modelos de Alisamento Exponencial

Neste capítulo serão discutidos modelos de alisamento exponencial incorporados à estrutura de espaço de estados com base nos trabalhos de Hyndman et al. (2008) e Hyndman et al. (2002). Tais modelos têm sido extremamente utilizados no contexto de suavização e previsão de séries temporais. Gardner (2005, p. 59) contabiliza 65 trabalhos empíricos utilizando apenas estes modelos. A complementação desta classe de modelos inicia-se com os trabalho de Ord et al. (1997), trazendo uma estrutura estocástica baseada em espaço de estados, que permite estimar a verossimilhança do modelo, construção de intervalos de confiança para a previsão e procedimentos de seleção de modelos. Desta forma surge uma classe de modelos diversificada, que iniciou-se com os conhecidos quatro modelos de Holt-Winters (ver, Morettin e Tolo, 2006) até os atuais 30 modelos descritos em Hyndman et al. (2008).

A Seção 2.1 traz uma revisão dos modelos de alisamento exponencial, a inovação em suas componentes, e as equações pontuais para os trinta modelos; a Seção 2.2 apresenta a estrutura estocástica de cada modelo; a Seção 2.3 apresenta exemplos do cálculo de modelos de estado envolvendo o modelo de Holt com erros multiplicativos e aditivos. Na Seção 2.4 são apresentados os métodos de estimação para esta classe de modelos, incluindo intervalos de confiança para previsão de modelos lineares. Por fim, a Seção 2.5 discute métodos de seleção de modelos, juntamente é apresentado um algoritmo para seleção de modelos baseado em critérios de seleção, conhecido por método ETS automático.

## 2.1 Modelos de Alisamento Exponencial

Os modelos de alisamento exponencial são constituídos a partir de três componentes não observáveis, que serão modeladas através de médias móveis ponderadas, são elas:

- a componente de erro, denotada por  $\varepsilon_t$ , com  $t = 1, \dots, n$ .
- a componente de tendência, denotada por  $T_t$ , com  $t = 1, \dots, n$ .
- a componente sazonal, denotada por  $S_t$ , para  $t = 1, \dots, n$ .

desta forma, define-se a estrutura de modelagem de cada componente.

**Definição 2.1** *As estruturas de modelagem são definidas da seguinte maneira:*

1. **N:** *Nenhuma estrutura ou estrutura ausente.*
2. **A:** *Estrutura aditiva.*
3. **M:** *Estrutura multiplicativa.*
4. **A<sub>d</sub>:** *Estrutura damped aditiva.*
5. **M<sub>d</sub>:** *Estrutura damped multiplicativa.*

Desta maneira a componente de tendência assume uma dentre as cinco possíveis estruturas descritas pela Definição 2.1, (N, A, M, A<sub>d</sub>, ou M<sub>d</sub>), a componente de erro assume duas estruturas, (A ou M) e a componente sazonal assume três estruturas (N, A ou M). A estrutura de tendência *damped* é apropriada quando há uma curva suave na tendência. Tal estrutura, em geral, melhora a acurácia das previsões, como mostrado por Makridakis e Hibon (2000).

Na prática é possível perceber a presença ou ausência das componentes tendência e sazonalidade através de análise gráfica da série temporal durante o período de tempo em questão. No entanto, distinguir a estrutura de cada componente é uma tarefa mais complicada, que geralmente é realizada através de testes de hipóteses ou critérios de seleção de modelos (Fisher, 1929, Hyndman e Koehler, 2006, Morry, 1975). Gardner (1985), apresenta graficamente o comportamento intuitivo das componentes  $T_t$  e  $S_t$  de acordo com suas respectivas estruturas.

A partir destas componentes, Hyndman et al. (2002) define uma notação para o modelo de alisamento exponencial via espaço de estados através da sigla ETS referente às componentes Erro, Tendência e Sazonalidade respectivamente. A Tabela 2.1 ilustra a diversidade de modelos construídos a partir desta definição considerando fixa a componente erro.

Tabela 2.1: Classificação dos modelos de alisamento exponencial via espaço de estados

Tendência	Sazonalidade		
	N	A	M
N	N,N	N,A	N,M
A	A,N	A,A	A,M
$A_d$	$A_d,N$	$A_d,A$	$A_d,M$
M	M,N	M,A	M,M
$M_d$	$M_d,N$	$M_d,A$	$M_d,M$

Cada um dos quinze modelos tem uma componente de tendência e uma de sazonalidade. A célula (N, N) descreve o modelo de alisamento exponencial simples, (A, N) alisamento exponencial de Holt e (A, A) o modelo Holt-Winters aditivo, e assim segue a construção dos modelos.

Historicamente, a distribuição da componente de erro tem sido ignorada, pois a distinção da componente erro não interfere nas previsões pontuais. Alternando a componente de erro (aditiva ou multiplicativa), obtém-se trinta modelos de alisamento exponencial. Esta classificação de modelos é originária da taxonomia proposta por Pegels (1969), estendida por Gardner (1985) e modificada por Hyndman et al. (2002) e Taylor (2003).

### 2.1.1 Equações Pontuais

As estruturas das componentes de tendência e sazonalidade, apresentadas na Seção 2.1, são modeladas através de três equações de suavizações gerais: a equação de nível denotada por  $l_t$ , a equação de inclinação denotada por  $b_t$  e a equação de sazonalidade denotada por  $s_{t-m}$ , em que  $t = 1, \dots, n$  e  $m$  denota o período de sazonalidade; geralmente  $m = 1$  para dados anuais,  $m = 4$  para dados trimestrais e  $m = 12$  para dados mensais. A componente tendência é uma combinação das equações de nível e inclinação, da mesma forma a componente sazonal é uma combinação das equações de nível, inclinação e sazonalidade.



Tabela 2.2: Equações de suavização e respectivas previsões para cálculos recursivos

Tendência	Sazonalidade		
	<b>N</b>	<b>A</b>	<b>M</b>
<b>N</b>	$l_t = \alpha y_t + (1 - \alpha)l_{t-1}$ $\hat{y}_{t+h t} = l_t$	$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)l_{t-1}$ $s_t = \gamma(y_t - l_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t + s_{t-m+h_m^+}$	$l_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)l_{t-1}$ $s_t = \gamma(y_t/l_{t-1}) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t s_{t-m+h_m^+}$
<b>A</b>	$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$ $\hat{y}_{t+h t} = l_t + hb_t$	$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - (l_{t-1} + b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t + hb_t + s_{t-m+h_m^+}$	$l_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(l_{t-1} + b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = (l_t + hb_t)s_{t-m+h_m^+}$
<b>A<sub>d</sub></b>	$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $\hat{y}_{t+h t} = l_t + \phi_h b_t$	$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - (l_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t + \phi_h b_t + s_{t-m+h_m^+}$	$l_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(l_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = (l_t + \phi_h b_t)s_{t-m+h_m^+}$
<b>M</b>	$l_t = \alpha y_t + (1 - \alpha)l_{t-1}b_{t-1}$ $b_t = \beta^*(l_t/l_{t-1}) + (1 - \beta^*)b_{t-1}$ $\hat{y}_{t+h t} = l_t b_t^h + s_{t-m+h_m^+}$	$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)l_{t-1}b_{t-1}$ $b_t = \beta^*(l_t/l_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t - (l_{t-1}b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t b_t^h + s_{t-m+h_m^+}$	$l_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)l_{t-1}b_{t-1}$ $b_t = \beta^*(l_t/l_{t-1}) + (1 - \beta^*)b_{t-1}$ $s_t = \gamma(y_t/(l_{t-1}b_{t-1})) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t b_t^h s_{t-m+h_m^+}$
<b>M<sub>d</sub></b>	$l_t = \alpha y_t + (1 - \alpha)l_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(l_t/l_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} s_{t-m+h_m^+}$	$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)l_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(l_t/l_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t - l_{t-1}b_{t-1}^\phi) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} + s_{t-m+h_m^+}$	$l_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)l_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(l_t/l_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t/(l_{t-1}b_{t-1}^\phi)) + (1 - \gamma)s_{t-m}$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} s_{t-m+h_m^+}$

Seguindo a notação de Hyndman et al. (2008), a Tabela 2.2 apresenta as equações pontuais de nível, inclinação e sazonalidade juntamente com as respectivas previsões para cada modelo.

Na Tabela 2.2,  $\phi_h = \phi + \phi^2 + \dots + \phi^h$ ;  $h_m^+ = [(h - 1) \bmod m] + 1$ , mod: operador *resto* da divisão. As constantes  $\alpha$ ,  $\beta^*$ ,  $\gamma$  e  $\phi$  são os parâmetros de suavização de suas respectivas equações. No caso  $\alpha$ ,  $\beta^*$  e  $\phi$  são responsáveis por modelar a tendência linear dos dados, e  $\phi$  é responsável por amortizar (*damping*) a tendência. Fazendo  $\phi = 1$ , temos os modelos sem amortização da tendência;  $\gamma$  é o parâmetro que modela a sazonalidade dos dados.

Por simplificação, considerou-se  $\beta = \alpha\beta^*$  ao invés de trabalhar com os cálculos baseados em  $\beta$ . A restrição dos parâmetros é organizada cuidadosamente a fim de evitar problemas de otimização durante o processo iterativo de estimação, que será abordado na Seção 2.4, desta forma temos:

$$0 \leq \alpha \leq 1, 0 \leq \beta \leq \alpha, 0 \leq \gamma \leq 1 - \alpha \text{ e } 0 < \phi < 1$$

## 2.2 Modelos de Espaço de Estados

O estudo de modelos de espaço de estado iniciou-se com o artigo de Kalman (1960), cujo trabalho descreve um processo recursivo para solução de problemas lineares relacionados à filtragem de dados discretos. A extensão à dados contínuos deu-se com a contribuição em 1961 de Richard S. Bucy no artigo de Kalman e Bucy (1961), também

conhecido como filtro de Kalman-Bucy. Modelos de espaço de estado permitem considerável flexibilidade na especificação de uma estrutura paramétrica. Com o desenvolvimento de pesquisa, estes modelos podem ser trabalhados tanto em estruturas lineares como não lineares. Trabalhos posteriores aplicados a engenharia vieram com Jazwinski (1970) e Anderson e Moore (1979). A representação Markoviana para estatística foi desenvolvida por Akaike (1973) e Akaike (1974). Hannan e Deistler (1988) fornecem uma apresentação unificada para análise de séries temporais estacionárias. Em economia, Aoki e Havennerb (1991) trabalham com modelos de espaço de estados multivariados e sugerem procedimentos para dados estacionários e não estacionários. Uma outra vantagem do filtro de Kalman é a estimação dos parâmetros desconhecidos do modelo através da maximização da verossimilhança via decomposição do erro de previsão.

Nesta dissertação será utilizada a formulação proposta por Hyndman et al. (2008) que é baseada nos trabalhos de Anderson e Moore (1979); Aoki (1987); Hannan e Deistler (1988).

Sejam  $y_t$  as observações da série temporal no tempo  $t$ , e  $x_t$  um vetor de estados contendo componentes não observáveis que descrevem as componentes de nível, tendência e sazonalidade da série temporal, com  $t = 1, \dots, n$ . Então uma inovação linear de modelos de espaço de estados pode ser escrita como

$$y_t = w'x_{t-1} + \varepsilon_t, \quad (2.2)$$

$$x_t = Fx_{t-1} + g\varepsilon_t, \quad (2.3)$$

em que  $w$ ,  $F$  e  $g$  são coeficientes e  $\varepsilon_t$  é um processo de ruído branco. A equação (2.2) é conhecida como equação de medida ou observação, e descreve o relacionamento entre os estados não observados  $x_{t-1}$  e a observação  $y_t$ . A equação (2.3) é conhecida como equação de transição ou estado, e descreve a evolução dos estados em relação ao tempo. As equações (2.2) e (2.3) são similares ao filtro de Kalman, no entanto a diferença se dá ao considerar o uso de erros idênticos nas duas equações, fazendo disso uma inovação nos modelos de espaço de estado, o que não era considerado no filtro de Kalman.

Os modelos de espaço de estados ajustam-se bem à abordagem de alisamento exponencial por conta das equações de nível, tendência e sazonalidade terem seus estados explícitos nesses modelos. Em contraste, estes estados não podem ser escritos explicita-

mente nos modelos ARIMA.

As equações para implementação dos modelos de espaço de estados foram propostas por Ord et al. (1997) e são conhecidas pela estrutura geral OKS. Tal estrutura foi desenvolvida com o intuito de trabalhar tanto com modelos lineares quanto não lineares, envolvendo um vetor de estados  $x_t$  e equações de espaço de estado da forma

$$y_t = w(x_{t-1}) + r(x_{t-1})\varepsilon_t, \quad (2.4)$$

$$x_t = f(x_{t-1}) + g(x_{t-1})\varepsilon_t, \quad (2.5)$$

onde  $\varepsilon_t$  é ruído branco com média zero e variância  $\sigma^2$ .

Define-se  $x_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{m-1})$ , e  $\mu_t = w(x_{t-1})$ . O modelo com erros aditivos é construído através da equação  $y_t = \mu_t + \varepsilon_t$ , e neste caso  $r(x_{t-1}) = 1$ . Para o modelo com erros multiplicativos a equação é dada por  $y_t = \mu(1 + \varepsilon_t)$ , com  $r(x_{t-1}) = \mu_t$  e  $\varepsilon_t = (y_t - \mu_t)/\mu_t$ .

Todos os modelos de alisamento exponencial da Tabela 2.2 podem ser escritos na forma (2.4) e (2.5), suas respectivas equações são dadas pela Tabela 2.3. Os modelos não são únicos, pois mudanças em  $r(x_{t-1})$  levarão a outras previsões de  $y_t$ . Por exemplo, Koehler et al. (2001) e Archibald (1990) trazem diferentes previsões para o modelo de Holt-Winters apenas alterando a função  $r(x_{t-1})$ . No entanto, os modelos são idênticos em forma. A única diferença entre os modelos com erros aditivos e multiplicativos está na observação da equação (2.4). A equação de espaço (2.5) pode ser escrita exatamente da mesma forma substituindo  $\varepsilon_t = (y_t - \mu_t)/\mu_t$  dentro de cada equação. A Tabela 2.3 mostra os 15 modelos de alisamento exponencial fixado com erros aditivos.

Nem todos os 30 modelos de espaço de estado são apropriados para qualquer tipo de dados. Dados com zero ou negativos não podem ser empregados em modelos multiplicativos por não serem bem definidos neste domínio. Da mesma forma, não são considerados modelos sazonais quando não existem evidências de períodos de sazonalidade na série temporal.

## 2.3 Cálculo das Equações de Espaço de Estados Para o Modelo Linear de Holt

Nesta seção, ilustraremos a idéia de compor o modelo de espaço de estados utilizando o modelo linear de Holt com erros aditivos e multiplicativos.

As previsões para o alisamento exponencial de Holt utilizam duas constantes de alisamento  $\alpha$  e  $\beta$ , ambas variando entre 0 e 1, além das três equações:

$$\text{Nível: } l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.6)$$

$$\text{Tendência: } b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.7)$$

$$\text{Previsão: } \hat{y}_{t+h|t} = l_t + b_t h \quad (2.8)$$

Tabela 2.3: Equações de espaço de estado para cada modelo de alisamento exponencial com erros aditivos.

	Tendência		Sazonalidade	
	<b>N</b>	<b>A</b>	<b>M</b>	
<b>N</b>	$\mu_t = l_{t-1}$ $l_t = l_{t-1} + \alpha \varepsilon_t$ $\hat{y}_{t+h t} = l_t$	$\mu_t = l_{t-1} + s_{t-m}$ $l_t = l_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t + s_{t-m+h_m^+}$	$\mu_t = l_{t-1} s_{t-m}$ $l_t = l_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (l_{t-1})$ $\hat{y}_{t+h t} = l_t s_{t-m+h_m^+}$	
<b>A</b>	$\mu_t = l_{t-1} + \phi b_{t-1}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t + b_t^h$	$\mu_t = l_{t-1} + \phi b_{t-1} + s_{t-m}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t + b_t^h + s_{t-m+h_m^+}$	$\mu_t = (l_{t-1} + \phi b_{t-1}) s_{t-m}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / (s_{t-m} l_{t-1})$ $s_t = s_{t-m} + \gamma \varepsilon_t / (l_{t-1} + \phi b_{t-1})$ $\hat{y}_{t+h t} = (l_t + b_t^h) s_{t-m+h_m^+}$	
<b>A<sub>d</sub></b>	$\mu_t = l_{t-1} + \phi b_{t-1}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t + \phi_h b_t^h$	$\mu_t = l_{t-1} + \phi b_{t-1} + s_{t-m}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t + \phi_h b_t^h + s_{t-m+h_m^+}$	$\mu_t = (l_{t-1} + \phi b_{t-1}) s_{t-m}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / (s_{t-m} l_{t-1})$ $s_t = s_{t-m} + \gamma \varepsilon_t / (l_{t-1} + \phi b_{t-1})$ $\hat{y}_{t+h t} = (l_t + \phi_h b_t^h) s_{t-m+h_m^+}$	
<b>M</b>	$\mu_t = l_{t-1} b_{t-1}$ $l_t = l_{t-1} b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h}$	$\mu_t = l_{t-1} b_{t-1} + s_{t-m}$ $l_t = l_{t-1} b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} + s_{t-m+h_m^+}$	$\mu_t = l_{t-1} b_{t-1} s_{t-m}$ $l_t = l_{t-1} b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / (s_{t-m} l_{t-1})$ $s_t = s_{t-m} + \gamma \varepsilon_t / (l_{t-1} b_{t-1})$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} s_{t-m+h_m^+}$	
<b>M<sub>d</sub></b>	$\mu_t = l_{t-1} b_{t-1}^\phi$ $l_t = l_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h}$	$\mu_t = l_{t-1} b_{t-1}^\phi + s_{t-m}$ $l_t = l_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma \varepsilon_t$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} + s_{t-m+h_m^+}$	$\mu_t = l_{t-1} b_{t-1}^\phi s_{t-m}$ $l_t = l_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / (s_{t-m} l_{t-1})$ $s_t = s_{t-m} + \gamma \varepsilon_t / (l_{t-1} b_{t-1})$ $\hat{y}_{t+h t} = l_t b_t^{\phi_h} s_{t-m+h_m^+}$	

### 2.3.1 Erros Aditivos

Seja  $\mu_t = l_{t-1} + b_{t-1}$  a média do modelo de alisamento exponencial via espaço de estados dado pela Tabela 2.3,  $\varepsilon_t = y_t - \mu_t$  denota o erro a um passo de previsão no tempo  $t$ , e considere  $\alpha$  e  $\beta$  constantes conhecidas. Desta forma o modelo é escrito por:

$$y_t = l_{t-1} + b_{t-1} + \varepsilon_t \quad (2.9)$$

usando (2.2) e (2.3) pode-se escrever

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t \quad (2.10)$$

$$b_t = b_{t-1} + \beta(l_t - l_{t-1} - b_{t-1}) \quad (2.11)$$

$$b_t = b_{t-1} + \alpha\beta\varepsilon_t \quad (2.12)$$

As três equações (2.9) (2.10) e (2.12) constroem o modelo de Holt de espaço de estados. Definindo  $x_t = (l_t, b_t)'$  e utilizando (2.9) e (2.12), pode-se então escrever as equações na forma padronizada de modelos de espaço de estados como

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} x_{t-1} + \varepsilon_t,$$

$$x_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix} \varepsilon_t$$

usualmente assume-se que a distribuição do erro  $\varepsilon_t \sim N(0, \sigma^2)$  com erros não correlacionados. No entanto tal suposição de normalidade nem sempre é verificada, outra forma é supor que  $E[\varepsilon_t] = 0$  e  $\text{Var}[\varepsilon_t] = \sigma^2$ ,  $E[\varepsilon_i, \varepsilon_j] = 0$ , para  $i \neq j$ .

### 2.3.2 Erros Multiplicativos

O modelo com erros multiplicativos pode ser derivado similarmente, considerando  $\varepsilon_t = (y_t - \mu_t)/\mu_t$  um erro relativo. Seguindo uma abordagem similar aos erros aditivos

têm-se

$$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t), \quad (2.13)$$

$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t), \quad (2.14)$$

$$b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t, \quad (2.15)$$

ou

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & 1 \end{bmatrix} x_{t-1} (1 + \varepsilon_t), \\ x_t &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x_{t-1} \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix} \varepsilon_t \end{aligned}$$

assumindo  $\varepsilon_t \sim N(0, \sigma^2)$  e erros não correlacionados.

## 2.4 Inicialização e Estimação

A fim de utilizar esses modelos para previsão de séries temporais, é necessário definir dois pontos importantes. O primeiro é a inicialização do algoritmo iterativo, ou seja,  $x_0$ . O segundo ponto importante é a estimação dos parâmetros  $\alpha, \beta, \gamma$  e  $\phi$ .

### 2.4.1 Inicialização

Gardner (1985) discute soluções para o problema de valor inicial dos modelos de alisamento exponencial, de maneira a obter convergência rápida do algoritmo iterativo de estimação. Outra forma de encontrar as componentes é através de um esquema heurístico. O esquema heurístico de Hyndman et al. (2002) parece funcionar muito bem e computacionalmente é empregado no pacote *forecast* (Hyndman e Khandakar, 2008) do *software R*. Portanto será discutido este esquema para inicialização das componentes  $x_0$ .

- *Inicialização da equação sazonal ( $s_{t-m}$ ):* Para dados sazonais; calcule médias móveis de  $2 \times m$  nos primeiros anos dos dados (normalmente os 4 primeiros anos). Denote isto por  $\{f_t\}$ ,  $t = m/2 + 2, m/2 + 3, \dots$ . Para sazonalidade aditiva; retire a tendência dos dados e faça  $y_t - f_t$ . Para sazonalidade multiplicativa; retire a tendência

dos dados e faça  $y_t/f_t$ . Calcule os índices sazonais iniciais,  $s_{-m+1}, \dots, s_0$ , retirando a média e tendência dos dados para cada período. Normalize esses índices sazonais para que eles somem zero para sazonalidade aditiva, e somem  $m$  para sazonalidade multiplicativa.

- *Inicialização da equação de nível ( $l_t$ ):* Para dados sazonais; estime a tendência linear através do método de mínimos quadrados para os primeiros dez valores dos períodos ajustados (utilizando os índices sazonais obtidos acima) contra  $t = 1, \dots, 10$ . Para dados não sazonais; estime a tendência linear utilizando o método de mínimos quadrados para as primeiras dez observações contra  $t = 1, \dots, 10$ . Então inicie  $l_0$  de forma a receber o valor do intercepto do ajuste da regressão.
- *Inicialização da equação de inclinação ( $b_t$ ):* Para modelo aditivo; inicie  $b_0$  de forma a receber o valor da estimativa da inclinação da regressão ajustada. Para tendência multiplicativa; inicie  $b_0 = 1 + b/a$ , em que  $a$  denota a estimativa do intercepto e  $b$  denota a estimativa da inclinação da tendência ajustada.

### 2.4.2 Estimação

Tendo em mãos um modelo paramétrico construído a partir de espaço de estados, é possível obter estimativas para a log-verossimilhança. No apêndice D mostra-se que

$$L^*(\tau, x_0) = n \log\left(\sum_{t=1}^n \varepsilon_t^2\right) + 2 \sum_{t=1}^n \log |r(x_{t-1})| \quad (2.16)$$

é igual a duas vezes o negativo do logaritmo da função de verossimilhança, eliminando os termos constantes, condicionado nos parâmetros  $\tau = (\alpha, \beta, \gamma, \phi)'$  e os estados iniciais  $x_0 = (l_0, b_0, s_0, s_1, \dots, s_{-m+1})'$ , onde  $n$  é o número de observações. Isto é facilmente calculado utilizando as equações recursivas na Tabela 2.2. Diferente dos modelos de espaço de estados com múltiplas fontes de erro, não é necessário utilizar o filtro de Kalman para calcular a verossimilhança.

Os parâmetros  $\tau$  podem ser estimados maximizando (D.2). Alternativamente, as estimativas também podem ser obtidas minimizando-se o erro quadrático médio a um passo de previsão, minimizando-se a variância residual, ou através de outro critério que

meça o erro de previsão. Seja qual for o critério usado,  $x_0$  é obtido pelo esquema heurístico já apresentado e  $\tau_0 = (0.1, 0.001, 0.001, 0.99)'$ .

### 2.4.3 Intervalos de Confiança Para as Previsões

Além das equações recursivas de previsão dadas pelas Tabelas 2.2, os intervalos de confiança para a previsão são extremamente úteis na modelagem da incerteza dos valores futuros previstos.

Considerando que os parâmetros do modelo são conhecidos, então também conhecemos o vetor de estados  $x_n$  da última observação, pois o erro na equação de transição pode ser calculado a partir das observações até o tempo  $n$ . Desta maneira, define-se a distribuição preditiva como a distribuição dos valores futuros da série temporal dado o modelo, seus parâmetros estimados e o vetor de estados  $x_n$ .

A distribuição preditiva é denotada por  $y_{n+h|n}$  ou  $y_{n+h|x_n}$ , em que  $h$  é o horizonte de previsão  $h = 1, \dots, H$ . A média da distribuição preditiva ou média de previsão é denotada por  $\mu_{n+h|n} = E(y_{n+h|x_n})$  e a variância da previsão é denotada por  $v_{n+h|n} = V(y_{n+h|x_n})$ .

Uma maneira de quantificar esta incerteza e obter a distribuição preditiva é através de métodos de simulação da trajetória dos dados condicionado ao estado final  $x_n$ . Esta abordagem é discutida em Ord et al. (1997) e Hyndman et al. (2002), levando em consideração procedimentos de bootstrap para a simulação da distribuição preditiva. No entanto, a desvantagem destes procedimentos é o custo computacional para simular e obter a distribuição preditiva. Desta maneira, alternativas como provar analiticamente a distribuição preditiva é um caminho empregado para contornar o emprego destes procedimentos. Hyndman et al. (2008, p. 95) deriva todos os resultados da distribuição preditiva para as classes de modelos homocedásticos e heterocedásticos. Neste trabalho discutiremos apenas estes resultados para a classe de modelos homocedásticos, pois a aplicação da Seção 6 refere-se em especial a estes modelos.

A classe de modelos homocedásticos incluem os seguintes modelos de alisamento exponencial: (ANN), (AAN), ( $AA_dN$ ), (ANA), (AAA), e ( $AA_dA$ ). A média de previsão dos modelos é dada pela Tabela 2.2. Por causa da estrutura linear dos modelos, as médias de previsão são idênticas aos pontos de previsão dados na Tabela 2.3. No entanto, a



previsão para a variância é dada por:

$$v_{n+h|n} = V(y_{n+h}|x_n) = \begin{cases} \sigma^2 & \text{se } h = 1; \\ \sigma^2 \left[ 1 + \sum_{j=1}^{h-1} c_j^2 \right] & \text{se } h \geq 2 \end{cases} \quad (2.17)$$

onde  $c_j$  é dado pela Tabela 2.4.

Tabela 2.4: Média da previsão e componente  $c_j$  para o cálculo da variância das previsões de modelos homocedásticos

Modelo(TS)	Média da Previsão	$c_j$
NN	$l_n$	$\alpha$
NA	$l_n + hb_n$	$\alpha + j\beta$
$A_d$ N	$l_n + \phi b_n$	$\alpha + \beta\phi_j$
NA	$l_n + s_{n-m+h+m^+}$	$\alpha + \gamma d_{j,m}$
AA	$l_n + hb_n + s_{n-m+h+m^+}$	$\alpha + j\beta + \gamma d_{j,m}$
$A_d$ A	$l_n + \phi_h b_n + s_{n-m+h+m^+}$	$\alpha + \beta\phi_j + \gamma d_{j,m}$

Note que  $v_{n+h|n}$  depende somente de  $h$  e dos parâmetros de suavização  $\alpha, \beta, \gamma$  e  $\phi$ , pois  $d_{j,m} = 1$  se  $j \bmod m = 0$ , e 0 caso contrário;  $\phi_j = \phi + \phi^2 + \dots + \phi^j$ . Para expressões fechadas da variância (ver, p. 82 Hyndman et al., 2008).

Assumindo que  $\varepsilon_t$  é um processo gaussiano, então  $y_{n+h}|x_n$  também é gaussiano. Portanto os intervalos de previsão são facilmente obtidos a partir da equação:

$$\text{ICp}(100(1 - q)\%) = \mu_{n+h|n} \pm z_q \sqrt{v_{n+h|n}} \quad (2.18)$$

onde  $z_q$  denota o  $q$ -ésimo quantil de uma distribuição normal padrão. O estimador para a variância  $\sigma^2$  é o estimador de máxima verossimilhança  $\hat{\sigma}^2$ .

De acordo com Hyndman et al. (2008) as distribuição preditivas para classes heterocedásticas são não gaussianas, isso deve-se por causa da não-linearidade das equações de espaço de estados. No entanto, utilizar aproximações gaussianas resultam em boas aproximações para os intervalos de confiança da previsão. Na prática, quando aproximações gaussianas não são verificadas, é necessário utilizar procedimentos de simulação aos dados para obter os respectivos intervalos de confiança.

## 2.5 Método ETS Automático

Devido a disponibilidade de vários modelos para o ajuste de séries temporais, muitas vezes é necessário selecionar um único modelo que dentre as várias possibilidades produza previsões acuradas, ou selecionar mais de um método e utilizar metodologias de combinação de modelos a fim de aumentar a capacidade previsiva. Nesta Seção, são discutidas metodologias de seleção modelos, e apresentaremos uma metodologia automática para seleção de um modelo dentre as 30 possibilidades de modelos de alisamento exponencial apresentadas, conhecido como ETS automático, proposto por Hyndman et al. (2002).

O objetivo da seleção de modelos é encontrar um modelo parcimonioso dentre várias possibilidades, isto é, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Desta maneira espera-se um modelo, que reflita um bom ajuste e conseqüentemente um bom poder de previsão. A fim de selecionar um único modelo dentre diversas possibilidades, normalmente faz-se o uso de critérios de informação. Existem muitos critérios de seleção disponíveis na literatura, (ver, Bozdongan, 1987, Kadane e Lazar, 2004, Littell et al., 2002, Morettin e Toloi, 2006, Wolfinger, 1993). Na prática é comum o emprego de mais de um critério de seleção para decidir entre possíveis modelos candidatos.

Os critérios de informação geralmente penalizam a verossimilhança, que é comparada com o número de parâmetros em questão no modelo para compensar o ajuste em potencial dos dados. A seleção de modelos pode ser vista como um problema de aproximação da função densidade condicional  $f(y^*|y)$  de  $y^*$  dado  $y$  que é observado. Seja  $g(y^*|y)$  uma estimativa de  $f(y^*|y)$ . Para medir se  $g$  é uma boa aproximação para  $f$  utiliza-se o negativo da entropia, também conhecido como a distância de Kullback-Leibler ou informação discriminante condicional de Kullback-Leibler, em que seu tamanho reflete o erro de aproximação do modelo, dado por

$$I_{y^*|y}(f, g) = \int f(y^*|y) \log \frac{f(y^*|y)}{g(y^*|y)} dy^* \quad (2.19)$$

O princípio da entropia negativa é selecionar a densidade aproximada  $g$  que minimiza a esperança negativa da entropia  $E_y[I_{y^*|y}(f, g)]$ . Por exemplo, o critério de informação de

Akaike estima  $I_{y^*|y}(f, g)$  e é desenvolvido para produzir um estimador assintoticamente não viesado conforme o tamanho da amostra  $n$  aumenta. Portanto, o modelo que apresentar o menor AIC deverá ter o menor erro de previsão assintoticamente para  $y^*$  (Akaike, 1977, p. 27).

De acordo com Hyndman et al. (2008), a fórmula geral para os critérios de informação (CI) aplicada a modelos de espaço de estados é dada por

$$CI = -2 \log L(\hat{\tau}, \hat{x}_0|y_t) + q\zeta(n) \quad (2.20)$$

em que  $L(\hat{\tau}, \hat{x}_0|y_t)$  é a função de verossimilhança dos dados,  $\hat{\tau}$  é a estimativa dos parâmetros  $\alpha, \beta, \gamma$  e  $\phi$ ;  $\hat{x}_0$  é o vetor de estados iniciais de  $l_0, b_0$  e  $s_{0+m}$ ;  $q$  é a soma das componentes  $\hat{\tau}$  e  $\hat{x}_0$ , e  $\zeta(n)$  é uma função do tamanho da amostra. Na Tabela 2.5 estão os critérios de informação mais comuns (Hyndman et al., 2008).

Tabela 2.5: Critérios de seleção de modelos

Critério	$\zeta(n)$	Penalidade	Fonte
AIC	2	2q	Akaike (1974)
BIC	$\ln(n)$	$q \log(n)$	Schwarz (1978)
HQIC	$2 \log(\log(n))$	$2q \log(\log(n))$	Hannan e Quinn (1979)
AICc	$2n/(n-q-1)$	$2qn/(n-q-1)$	Sugiura (1978)
LEIC	c empírico	qc	Billah et. al (2003)

Baseado nestas considerações, Hyndman et al. (2002) propõe o método ETS automático para selecionar modelos de alisamento exponencial via espaço de estados através de critérios de informação. Em sua abordagem são considerados os critérios de informação de Akaike e Akaike corrigido (AIC e AICc) e o critérios de seleção Bayesiano de Schwarz (BIC). O procedimento é completamente automático e considera todos os possíveis modelos dentre as 30 possibilidades, selecionando dentre este espaço de modelos o que retorne o menor valor do critério escolhido (por *default*, AIC).

Se não existir presença da componente sazonal nos dados, o método considerará apenas modelos com componente Erro e Tendência, reduzindo o espaço de modelos para ajuste e evitando problemas de otimização do procedimento. Da mesma forma, se os dados apresentam valores negativos ou nulos são considerados apenas procedimentos aditivos para a componente erro. O procedimento está disponível através da função *ETS* no pacote *forecast* do *software R*, este procedimento será aplicado do Capítulo 6.

# Capítulo 3

## Métodos ARIMA Automáticos

Os modelos ARIMA em teoria são a classe mais geral de modelos com estrutura estocástica para previsão de séries temporais estacionárias ou não estacionárias homogêneas, que através de transformações simples tais como diferenças ou logaritmo estas séries tornam-se estacionárias. Tais modelos são extremamente utilizados na modelagem de séries temporais por sua competência no ajuste e interpretação do modelo. Como consequência, diversos pacotes estatísticos e softwares de economia trazem esta família de modelos a disposição do analista.

Diversos procedimentos têm sido propostos a fim de selecionar modelos dentre as diversas possibilidades da família ARIMA. Uma revisão destes procedimentos é encontrada em Ord e Lowe (1996). O trabalho de Makridakis e Hibon (2000) revela que grande parte destes procedimentos como *Autobox Automatic Forecasting Systems*, *Rule-Based Forecasting (RBF)* e *Box-Jenkins Automático*, não obtiveram bons resultados em relação a acurácia das previsões da base M3. No entanto, um programa que destacou-se devido à seu algoritmo de previsão baseado em modelos ARIMA foi o software *Forecast Pro*.

Este capítulo apresenta uma breve introdução de modelos ARIMA e aborda a metodologia proposta por Hyndman e Khandakar (2008), cujo objetivo é selecionar um modelo da família ARIMA para previsão de séries temporais, conhecida como métodos Auto ARIMA. A Seção 3.1 faz uma breve introdução dos modelos ARIMA, a Seção 3.2 apresenta os algoritmos Auto Arima e na Seção 3.3 é discutido as diferenças e semelhanças entre os modelos de alisamento exponencial e modelos ARIMA.

### 3.1 Uma Breve Revisão

Dada uma série temporal  $y_t$  com  $t = 1, \dots, n$ , a ordem do modelo ARIMA que descreva corretamente o processo estocástico é uma questão que muitos pesquisadores buscaram responder nos últimos 25 anos. Por exemplo, Hannan e Rissanen (1982) propõe um método para identificar a ordem de um modelo ARMA baseado em regressões sobre  $y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p}, \hat{\varepsilon}_t, \dots, \hat{\varepsilon}_{t-q}$ , em que  $\hat{\varepsilon}_{t-j}$ , com  $t = 1, \dots, n$ ,  $j = 0, \dots, q$  e  $p = 1, \dots, q$ , é obtido ao ajustar um modelo autorregressivo longo aos dados. São provadas propriedades assintóticas do método e um algoritmo recursivo para os cálculos. Este algoritmo foi expandido por Gomez (1998) para modelos sazonais multiplicativos e implementado por Gomez e Maravall (1998) no software TRAMO and SEATS.

Liu (1989) propõe uma metodologia para identificação de modelos sazonais ARIMA utilizando um método de filtro e regras heurísticas. Este algoritmo é utilizado no SCA-Expert Software. Outra abordagem implementada no pacote do software TSE-AX é descrita por Melard e Pasteels (2000), o qual o algoritmo para seleção de um modelo ARIMA univariado também permite análise de intervenção.

Dentre os softwares comerciais mais conhecidos, estão o programa *Forecast Pro* (Godrich, 2000), que é específico para previsões de séries temporais e o algoritmo *Autobox* (Reilly, 2000). Ambos participaram da competição M3 de Makridakis e Hibon (2000), com destaque de performance para o software *Forecast Pro*, que inclusive tornou-se um método de referência (*benchmark*) para outras competições (ver, Athanasopoulos et al., 2011). No entanto os softwares comerciais não publicam o algoritmo de seleção de modelos, portanto a metodologia só pode ser utilizada e não conhecida.

Devido a diversidade de modelos da família ARIMA, muitas vezes o esforço computacional empregado para selecionar um único modelo pode ser grande. Hyndman e Khandakar (2008) desenvolvem um procedimento baseado em testes de raízes unitárias e algoritmos chamados *stepwise*, para reduzir o esforço computacional sem prejudicar a seleção de um modelo provável para previsão. Tal procedimento é conhecido como *Auto Arima* com *stepwise* e está disponível no pacote *forecast* do software *R*. Paralelamente Halim e Bisnono (2008) também propõe um algoritmo automático baseado na transformação Box-Cox, detecção de raízes unitárias e na escolha de um modelo ARIMA baseado do critério de Akaike.

O processo ARIMA é definido por

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t \quad (3.1)$$

em que  $\varepsilon_t$  é um processo de ruído branco com média zero e variância  $\sigma^2$ ,  $B$  é o operador defasagem, e  $\phi(z)$  e  $\theta(z)$  são polinômios de ordem  $p$  e  $q$  respectivamente;  $d$  com  $d = 0, 1, 2, \dots$ , é um inteiro positivo que determina o número de diferenças na série temporal. Para garantir a estacionaridade e invertibilidade do processo, assume-se que  $\phi(z)$  e  $\theta(z)$  não tem raízes dentro do círculo unitário  $|z| < 1$  (Brockwell e Davis, 1991). Se  $c \neq 0$ , existe um polinômio de ordem  $d$  implícita na função de previsão.

O processo ARIMA sazonal  $(p, d, q)(P, D, Q)_m$  é dado por

$$\Phi(B^m)\phi(B)(1 - B^m)^D(1 - B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t \quad (3.2)$$

onde  $\Phi(z)$  e  $\Theta(z)$  são polinômios de ordem  $P$  e  $Q$  respectivamente, cada um não contendo raízes dentro do seu círculo unitário. Se  $c \neq 0$ , existe um polinômio implícito de ordem  $d + D$  na função de previsão.

O principal objetivo em prever automaticamente modelos ARIMA é selecionar um modelo de ordem apropriada  $p, q, P, Q, D, d$ . Caso  $d$  e  $D$  sejam conhecidos, seleciona-se as ordens do modelo  $p, q, P, Q$  através de critérios de seleção, por exemplo

$$AIC = -2 \log(L) + 2(p + q + P + Q + k) \quad (3.3)$$

em que  $k = 1$  se  $c \neq 0$  e  $k = 0$  caso contrário.  $L$  é a verossimilhança dos modelos ajustados aos dados diferenciados  $(1 - B)^D(1 - B)^d y_t$

Hyndman e Khandakar (2008) optam por iniciar o processo de maneira a descobrir o número de raízes unitárias que existem no processo. Na literatura existem diversos testes para raízes unitárias, normalmente assumindo sob a hipótese nula a existência de raízes unitárias (ver, Dickey e Fuller, 1981, Hylleberg et al., 1990). No entanto, por razões de simplificação Hyndman e Khandakar (2008) considera o teste que assume como hipótese nula a não existência de raízes unitárias.

Desta maneira, para dados não sazonais é aplicado o teste de raízes unitárias sucessivas KPSS, proposto por Kwiatkowski et al. (1992). Este teste simplesmente retorna a

existência de  $d$  raízes unitárias no processo, iniciando com  $d = 1$ . Compara-se a estatística teste a um nível de significância  $\alpha$  e, caso existam evidências que suportem a hipótese alternativa de que existe raiz unitária. No processo, sucessivamente testa-se  $d = 2, \dots$  até encontrar o primeiro resultado não significativo para o teste.

Para dados sazonais considera-se o modelo  $ARIMA(p, d, q)(P, D, Q)_m$ , em que  $m$  é o período de sazonalidade, e  $D = 0$  ou  $D = 1$  é obtido a partir do teste estendido de Canova-Hansen (ver, Canova e Hasen, 1995), com a modificação de que nesta versão do teste  $m > 1$ . Depois de  $D$  ser selecionado, o teste KPSS é utilizado para encontrar o valor de  $d$  para a série sazonalmente diferenciada, se  $D = 1$  ou  $D = 0$  (dados originais). Uma vez que os valores de  $d$  e  $D$  sejam admitidos, a seleção de  $p, q, P, Q$  é feita minimizando-se o critério de Akaike. Permite-se  $c \neq 0$  para modelos em que  $d + D < 2$ .

### 3.2 O Procedimento Stepwise Para Modelos ARIMA

O procedimento *stepwise* é extremamente utilizado em análise de regressão para iniciar o processo de modelagem segundo um número de parâmetros. No entanto, na prática o procedimento é utilizado em muitas ocasiões para selecionar um modelo de regressão. Hyndman e Khandakar (2008) propõe um procedimento *stepwise* para selecionar modelos ARIMA, esta seção explanará em detalhes este procedimento.

Suponha que se tem em mãos dados sazonais e que se deseja ajustar um modelo  $ARIMA(p, d, q)(P, D, Q)_m$ , em que os valores de  $p$  e  $q$  variam entre 0 e 3, e  $P$  e  $Q$  variam entre 0 e 1. Se  $c = 0$  existem 288 possíveis modelos para o ajuste dos dados e no caso em que  $c \neq 0$  existem 192 possibilidades, somando 480 modelos possíveis. Consequentemente, se o espaço paramétrico aumentar, a quantidade de possíveis modelos será enorme, fazendo com que não seja computacionalmente viável o ajuste de todos os possíveis modelos disponíveis para ajuste. Desta forma, é necessário buscar alternativas eficientes para chegar ao menor valor do critério de seleção escolhido para seleção de modelos. Esse é o objetivo do procedimento *stepwise* ARIMA, seu algoritmo é descrito da seguinte maneira:

Passo 1: Inicia-se o processo ajustando-se quatro modelos a princípio

- $ARIMA(2, d, 2)$  se  $m = 1$ , ou  $ARIMA(2, d, 2)(1, D, 1)_m$  se  $m > 1$ .

- ARIMA(0,  $d$ , 0) se  $m = 1$ , ou ARIMA(0,  $d$ , 0)(0,  $D$ , 0) $_m$  se  $m > 1$ .
- ARIMA(1,  $d$ , 0) se  $m = 1$ , ou ARIMA(1,  $d$ , 0)(1,  $D$ , 0) $_m$  se  $m > 1$ .
- ARIMA(0,  $d$ , 1) se  $m = 1$ , ou ARIMA(0,  $d$ , 1)(0,  $D$ , 1) $_m$  se  $m > 1$ .

Se  $d + D \leq 1$ , estes modelos são ajustados com  $c \neq 0$ , caso contrário  $c = 0$ . Destes quatro modelos, seleciona-se o que apresenta o menor valor AIC, denotando este modelo como modelo corrente.

Passo 2: Considera-se até treze variações do modelo corrente

- em que um dos parâmetros  $p, q, P, Q$  é permitido variar  $\pm 1$  a partir do modelo corrente;
- $p$  e  $q$  variem  $\pm 1$  a partir do modelo corrente;
- $P$  e  $Q$  variem  $\pm 1$  a partir do modelo corrente;
- a constante  $c$  é incluída no caso em que  $c = 0$  no modelo corrente, ou excluída se o modelo corrente tem  $c \neq 0$ .

Dentre os treze modelos um novo modelo com menor AIC é encontrado, e este é o novo modelo corrente. Desta forma, o procedimento é repetido até que o critério de parada seja atingido. O critério de parada é dado por,  $AIC_{novo} > AIC_{corrente}$ .

Existem várias restrições que devem ser consideradas a fim de evitar problemas no ajuste dos modelos com convergência de algoritmos iterativos ou raízes unitárias.

- Não é permitido que  $p$  e  $q$  excedam limitantes superiores especificados (por *default*  $p_{max} = 5$  e  $q_{max} = 5$ )
- Não é permitido que  $P$  e  $Q$  excedam limitantes superiores especificados (por *default*  $P_{max} = 2$  e  $Q_{max} = 2$ )
- Rejeita-se qualquer modelo próximo ao círculo unitário. Especificamente, computa-se as raízes de  $\phi(B)\Phi(B)$  e  $\theta(B)\Theta(B)$ . Se a raiz em módulo for menor que 1.001, então rejeita-se este modelo.
- Caso o modelo demore a convergir, o modelo é rejeitado. O princípio para este procedimento baseia-se na ideia de que modelos difíceis de ajustar provavelmente não são modelos adequados para os dados.



O algoritmo garante retornar um modelo válido pois o espaço de possíveis modelos é finito e ao menos um dos modelos no passo 1 será aceito (o modelo AR ou MA). Dado que um modelo foi selecionado, este é utilizado para gerar as requeridas previsões.

### 3.3 ARIMA vs Modelos de Alisamento Exponencial

"Existe um mito que diz que os modelos ARIMA são mais gerais do que os modelos de alisamento exponencial. O que não é verdade. As duas classes de modelo se sobrepõem. É verdade que os modelos lineares de alisamento exponencial são todos equivalentes a casos particulares de modelos ARIMA. No entanto, os modelos não lineares de alisamento exponencial não tem sua equivalência em modelos ARIMA. Por outro lado, existem muitos modelos ARIMA que não são casos particulares de modelos de alisamento exponencial. Portanto, conclui-se que as duas formas de modelagem em alguns casos se sobrepõem e em outros são complementares, cada qual com seus pontos fortes e fracos"(Hyndman et al., 2008).

Os modelos de alisamento exponencial são todos não estacionários. Modelos com sazonalidade ou tendência tem duas raízes unitárias; os modelos não sazonais, destendenciados ou com tendência amortecida, contém uma raiz unitária. É possível definir modelos estacionários com características similares a modelos de espaço de estados, no entanto, normalmente isso não é feito. Desta forma, a filosofia de modelos de alisamento exponencial é de que o mundo é não-estacionário. Portanto, se um modelo estacionário é requerido, então os modelos ARIMA são mais indicados.

Uma vantagem dos modelos de alisamento exponencial é que estes podem ser não lineares. Séries temporais que exibem características não lineares, incluindo heterocedasticidade, podem ser melhor modeladas pelos modelos de espaço de estados.

Para dados sazonais, existem mais possibilidades de ajuste através de modelos ARIMA do que as 30 possibilidades oferecidas pelos modelos de alisamento exponencial. Pode-se pensar que uma classe ampla de modelos ofereça mais vantagens. No entanto, os resultados de Hyndman et al. (2002) contradizem esta afirmação, de forma que os modelos de alisamento exponencial sobrepujaram os resultados de vários métodos ARIMA automáticos em relação à dados sazonais através do banco de dados *M3-Competition*. Contudo, em relação a dados anuais os modelos ARIMA apresentam uma performance melhor. Hynd-

man (2001) especula que a grande diversidade dos modelos sazonais ARIMA pode prejudicar a qualidade da previsão, por introduzir incertezas adicionais no modelo. Apesar de não tão vasta quanto os modelo ARIMA, a classe de modelos de alisamento exponencial via espaço de estados é suficientemente rica para capturar a dinâmica de séries temporais reais.

## **Capítulo 4**

# **O Método Theta Para Previsão de Séries Temporais**

Neste capítulo é apresentado o método Theta, utilizado na decomposição e previsão de séries temporais. A Seção 4.1 introduz o método Theta e sua generalização de acordo com os trabalhos de Assimakopoulos e Nikolopoulos (2000), Hyndman e Billah (2003), Nikolopoulos e Assimakopoulos (2005) e Nikolopoulos et al. (2008). A Seção 4.2 propõe redefinir o método Theta, baseando-se na quantidade de linhas-Theta para decomposição da série temporal, métodos de extrapolação e combinação das linhas-Theta a fim de aumentar a capacidade previsiva do modelo. A Seção 4.3 aborda o método Theta sob o ponto de vista de modelagem paramétrica, considerando um método de estimação para os coeficientes Theta.

### **4.1 Um Método de Decomposição de Séries Temporais**

O método Theta foi inicialmente proposto por Assimakopoulos e Nikolopoulos (2000), e chamou a atenção de pesquisadores devido sua excelente performance na previsão de séries temporais da base de dados M3, através da competição realizada por Makridakis e Hibon (2000). A ideia intuitiva do método baseia-se na decomposição de séries temporais univariadas, com o objetivo de capturar informações implícitas nos dados que não podem ser captadas durante o processo de modelagem da série original. Segundo Nikolopoulos e Assimakopoulos (2005), um único método de extrapolação é praticamente incapaz de

capturar toda a informação disponível na série temporal, que em muitas vezes depende de fatores externos aos dados; como por exemplo, crises financeiras. Por um lado, existem modelos que são muito simples para capturar toda a informação da série temporal, e outros tão complexos que no momento da otimização acabam não obtendo êxito em modelar todos os padrões da série. Por exemplo, modelos de alisamento exponencial que podem ir desde o alisamento exponencial simples (modelo menos sofisticado) até um Holt-Winters com damped (modelo mais sofisticado).

Desta maneira, o processo de decomposição resulta em séries mais simples, permitindo a aplicação de modelos mais adequados que modelarão as linhas decompostas. As linhas ajustadas pelos modelos de extrapolação são combinadas com o objetivo de reconstruir a série temporal em questão, obtendo-se por fim as previsões requeridas segundo um dado horizonte  $h$ . A ideia de combinar as linhas extrapoladas está também relacionada a diversificação de métodos de previsão. A Figura 4.1 ilustra o processo de decomposição e previsão do método Theta, a partir de uma série temporal sem presença de sazonalidade.

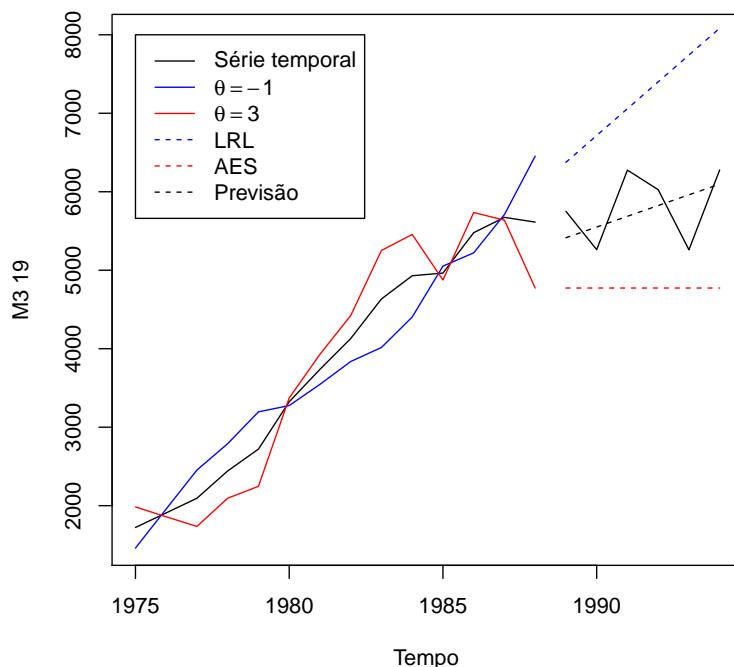


Figura 4.1: Método Theta com duas linhas de decomposição.

As componentes  $\theta = -0.5$  e  $\theta = -2.5$  representam respectivamente os coeficientes

de decomposição para construção das linhas Theta ou linhas de decomposição. A linha Theta construída a partir do coeficiente  $\theta = -0.5$  esboça o comportamento de longo prazo da série temporal, esta linha foi decomposta pelo modelo de regressão linear simples denotado por LRL, e sua previsão é dado pela linha em azul pontilhada. Da mesma forma, a linha em vermelho construída a partir do coeficiente  $\theta = 2.5$ , esboça o comportamento de curto prazo da série temporal, que é extrapolada de acordo com o modelo de alisamento exponencial simples, denotado por AES. A combinação das previsões representadas pelas linhas pontilhadas em azul e vermelho respectivamente para o modelo de regressão linear e o modelo de alisamento exponencial originam a previsão do método Theta. Concluindo a interpretação da Figura 4.1, a idéia intuitiva deste método está relacionada a diversificação e combinação de diferentes modelos de séries temporais aplicados as linhas de decomposição com o objetivo de previsão.

#### 4.1.1 Construção do Método Theta Generalizado

O método Theta é baseado no conceito de modificação das curvaturas locais da série temporal. Esta mudança é obtida a partir de um coeficiente, denominado coeficiente Theta, denotado pela letra grega ( $\theta$ ). O coeficiente é aplicado diretamente à segunda diferença da série temporal, ou seja

$$\nabla^2 Z_t(\theta) = \theta \nabla^2 y_t \quad (4.1)$$

em que  $Z_t(\theta)$  é a linha decomposta em função do coeficiente ( $\theta$ ) e do tempo  $t$ , com  $t = 1, \dots, n$ , chamada de linha-Theta; e  $\nabla$  é o operador diferença dado por  $\nabla Z_t = Z_t - Z_{t-1}$ .

Uma característica importante do método é que a média e a declividade das linhas decompostas  $Z_t(\cdot)$  são mantidas em relação aos dados originais. No entanto, as curvaturas da série temporal sofrem modificações devido a transformação dos dados.

Os cálculos propostos por Assimakopoulos e Nikolopoulos (2000) consideram que os dados da série temporal  $y_i$ , com  $i = 3, \dots, n$  podem ser escritos a partir da equação

$$y_i = y_1 + (i - 1)(y_2 - y_1) + \left( \sum_{t=2}^{i-1} (i - t) \nabla^2 y_{t+1} \right) \quad (4.2)$$

De maneira análoga, os pontos  $Z_i$  da série decomposta, chamada linha-Theta são definidos

por:

$$Z_i = Z_1 + (i - 1)(Z_2 - Z_1) + \theta \left( \sum_{t=2}^{i-1} (i - t) \nabla^2 y_{t+1} \right) \quad (4.3)$$

$Z_i, Z_1$  e  $Z_2$  são valores desconhecidos e são estimados através da minimização dos erros quadráticos, ou seja:

$$\min \left( \sum_i \varepsilon_i^2 \right) = \min \left( \sum_i (Z_i - y_i)^2 \right) \quad (4.4)$$

Após vários cálculos Assimakopoulos e Nikolopoulos (2000) chegam ao resultado para a equação (4.4).

Com base na equação (4.1), Hyndman e Billah (2003) simplificam estes cálculos baseados nos resultados de Kelley (2000), em que as linhas-Theta construídas com base na segunda diferença da série temporal, podem ser reescritas através de uma equação linear dada por,

$$Z_t(\theta) = a_\theta + b_\theta(t - 1) + \theta y_t \quad (4.5)$$

Desta forma a minimização dos erros quadráticos é equivalente a fazer uma regressão de  $(1 - \theta)y_t$  em relação a  $t - 1$ .

$$\frac{\partial \sum_i e_i^2}{\partial a_\theta} = 0 \quad (4.6)$$

$$\frac{\partial \sum_i e_i^2}{\partial b_\theta} = 0 \quad (4.7)$$

os estimadores para  $a_\theta$  e  $b_\theta$  são dados por:

$$\hat{a}_\theta = (1 - \theta)\bar{y} - b_\theta(n - 1)/2 \quad (4.8)$$

$$\hat{b}_\theta = \frac{6(1 - \theta)}{n^2 - 1} \left( \frac{2}{n} \sum_{t=1}^n t y_t - (n + 1)\bar{y} \right) \quad (4.9)$$

que é equivalente a resolver o problema levantado nas equações (4.2) e (4.3), de uma forma simplificada.

A partir da equação (4.5), (4.8) e (4.9) é possível derivar os seguintes resultados:

- $Z_t(0)$  resulta na linha de regressão linear da série  $y_t$  em relação ao tempo  $t$ ;
- $Z_t(1)$  resulta na série temporal  $y_t$ ;
- $\bar{Z}_t(\theta) = \bar{y}_t$  para todo  $\theta \in \mathfrak{R}$ ;
- A relação de simetria entre duas linhas-Theta, se dá através da equação

$$\frac{1}{2}(Z_t(\theta) + Z_t(2 - \theta)) = Z_t(1) \quad (4.10)$$

Baseado em (4.10), a previsão segundo um horizonte  $h$ ,  $h = 1, \dots, H$  para o modelo de Assimakopoulos e Nikolopoulos (2000) considerando apenas duas linhas-Theta de decomposição, e baseado nas observações da série temporal para o tempo  $t$ , é dado por:

$$\hat{y}_{t+h} = \frac{1}{2}(\hat{Z}_{t+h}(0) + \hat{Z}_{t+h}(2)) \quad (4.11)$$

em que  $\hat{Z}_{t+h}(0)$  é extrapolada segundo o modelo de regressão linear e  $\hat{Z}_{t+h}(2)$  é extrapolada segundo o modelo de alisamento exponencial simples. Este método foi utilizado para prever as séries temporais da base M3 realizada por Makridakis e Hibon (2000).

O artigo de Hyndman e Billah (2003), prova que este modelo é equivalente a um alisamento exponencial com desvio, que é dado pelas equações:

$$l_t = l_{t-1} + b + \alpha \varepsilon_t \quad (4.12)$$

$$\hat{y}_{t+h|t} = l_t + hb \quad (4.13)$$

em que o parâmetro de desvio  $b$  é dado particularmente dado por  $b = \hat{b}_0/2$ , e  $\hat{b}_0$  é a estimativa de inclinação da regressão linear dada por (4.5). O modelo também pode ser reconstruído a partir da forma reduzida de um modelo ARIMA(0,1,1) com desvio (ver Apêndice A4.2, Box et al., 1994).

Contudo, Nikolopoulos e Assimakopoulos (2005) afirmam que o alisamento exponencial com desvio é apenas um caso particular do método Theta, e que o método é mais geral do que considerar apenas duas linhas-Theta para a construção do modelo. Esforços posteriores considerando generalizações do método Theta foram feitos por (Nikolopoulos e Assimakopoulos, 2005, Nikolopoulos et al., 2012a, 2008).

Por fim, Nikolopoulos et al. (2008) provam que a maneira de calcular as linhas-Theta proposta por Hyndman e Billah (2003) é equivalente a seguinte equação

$$Z_t(\theta) = LRL + \theta e_{LRL} \quad (4.14)$$

$$= \theta y_t + (1 - \theta)LRL \quad (4.15)$$

em que LRL é a linha da regressão linear e  $e_{LRL}$  são os resíduos desta regressão no tempo  $t = 1, \dots, n$ .

**Prova 4.16** *A prova consiste em mostrar que a equação 4.5 é igual a equação 4.15. Desta forma temos:*

$$\begin{aligned} \hat{a}_\theta + \hat{b}_\theta(t-1) + \theta y_t &= \theta y_t + (1 - \theta)LRL \\ \Leftrightarrow \hat{a}_\theta + \hat{b}_\theta(t-1) &= (1 - \theta) \cdot (\hat{a}_0 + \hat{b}_0 t) \\ \Leftrightarrow (1 - \theta)\bar{y} - \hat{b}_\theta \frac{n-1}{2} + \hat{b}_\theta(t-1) &= (1 - \theta) \left( \bar{y} - \hat{b}_0 \frac{n+1}{2} + \hat{b}_0 t \right) \\ \Leftrightarrow \left[ -\frac{(n-1)}{2} + (t-1) \right] \hat{b}_\theta &= (1 - \theta) \left[ -\frac{(n+1)}{2} + t \right] \hat{b}_0 \end{aligned} \quad (4.17)$$

A relação entre  $\hat{b}_\theta$  e  $\hat{b}_0$  é dada pela equação (4.9), e é igual a

$$\hat{b}_\theta = (1 - \theta)\hat{b}_0 \quad (4.18)$$

Substituindo (4.17) em (4.18) temos:

$$\begin{aligned} -\frac{(n-1)}{2} + (t-1) &= -\frac{(n+1)}{2} + t \\ \Leftrightarrow -(n-1) + 2(t-1) &= -(n+1) + 2t \\ \Leftrightarrow -n + 1 + 2t - 2 &= -n - 1 + 2t \\ \Leftrightarrow -n - 1 + 2t &= -n - 1 + 2t \end{aligned}$$

o que é verdadeiro, como queríamos demonstrar.

Após mais de oito anos de pesquisa, com contribuições de diferentes autores, a equação (4.14) prova que a linha-Theta para decomposição de séries temporais univaria-



das é uma função do modelo de regressão linear aplicada diretamente aos dados.

De acordo com Assimakopoulos e Nikolopoulos (2000), a interpretação das linhas-Theta tem relação com as curvaturas locais da série temporal. Para o caso em que as curvaturas locais são gradualmente reduzidas, define-se a série temporal como deflacionada, neste caso  $\theta < 1$ . Como já mostrado, um caso particular de deflação se dá quando  $\theta = 0$ , dessa maneira a decomposição resulta no caso particular da linha de regressão linear. No caso em que  $\theta > 1$ , ocorre o aumento das curvaturas locais. Neste caso, a série é definida como dilatada e para  $\theta = 1$  tem-se a série original. A Figura 4.2 ilustra o comportamento de linhas-Theta deflacionadas e a Figura 4.3 o comportamento de linhas-Theta dilatadas.

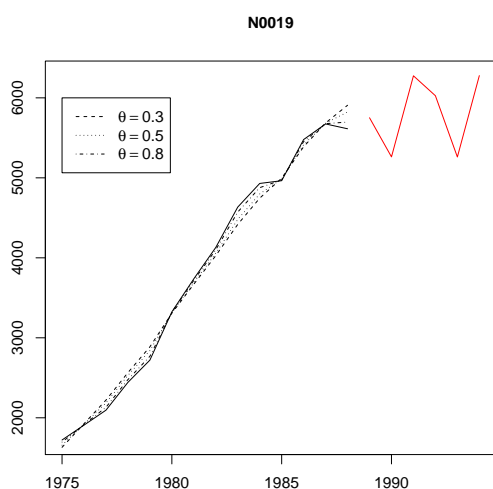


Figura 4.2: Linhas-Theta deflacionadas

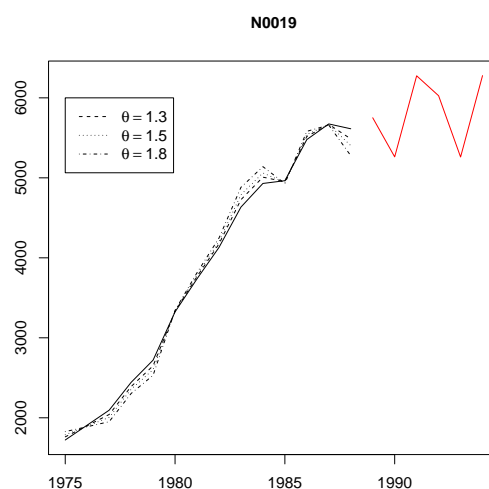


Figura 4.3: Linhas-Theta dilatadas

### 4.1.2 Construção do Método a Partir de Duas Linhas-Theta

O caso mais simples do método Theta para previsão de séries temporais, é o caso em que considera-se apenas duas linhas-Theta para decomposição (Assimakopoulos e Nikolopoulos, 2000, Hyndman e Billah, 2003, Makridakis e Hibon, 2000).

Neste caso, as linhas-Theta podem ser enxergadas como uma linha de tendência, o qual descreve o comportamento de longo ou médio prazo da série através da linha de regressão linear, e a segunda diferença que descreve os movimentos de curto-prazo da série temporal. Estes componentes são então extrapolados separadamente e os elementos

são combinados para produzir as previsões finais (Nikolopoulos et al., 2003).

Os passos do algoritmo para implementação são dados a seguir:

1. *Testando Sazonalidade*: Teste cada série temporal buscando existência de comportamento sazonal via Teste T-Student para a função de autocorrelação, com lag de um ano para séries anuais, lag de 12 meses para séries mensais e lag de 4 meses para séries trimestrais. Compara-se a estatística teste de 1.645 à probabilidade de 0.10 na cauda da função de densidade.
2. *Desazonalização*: Desazonalize as séries temporais identificadas no passo 1 através do método de decomposição clássica (multiplicativo).
3. *Decomposição*: Cada série temporal é decomposta em duas linhas-Theta,  $Z_t(0)$  e  $Z_t(2)$ .
4. *Extrapolação*:  $Z_t(0)$  e  $Z_t(2)$  são extrapoladas respectivamente através do modelo regressão linear e do modelo de alisamento exponencial simples.
5. *Combinação*: As previsões produzidas por  $\hat{Z}_{t+h}(0)$  e  $\hat{Z}_{t+h}(2)$  são combinadas a partir da média aritmética.
6. *Resazonalização*: As previsões do passo anterior são resazonalizadas resultando nas previsões finais.

## 4.2 Redefinindo o Método Theta

A partir das conclusões de Hyndman e Billah (2003), que provam que o método Theta é um caso particular de alisamento exponencial com desvio, em que o grau de desvio  $b$  é dado por  $b = \hat{b}_0/2$ , surgem questões interessantes sobre o modelo. Por exemplo, caso as linhas-Theta  $\hat{Z}_t(0)$  e  $\hat{Z}_t(2)$  não sejam combinadas por pesos iguais, então tem-se um caso geral do alisamento exponencial com desvio.

Nikolopoulos e Assimakopoulos (2005, p. 4) levantam questões importantes e até o momento não respondidas referentes ao método Theta em relação aos seguintes pontos:

- Quantas linhas-Theta devem ser utilizadas no modelo?

- Quais métodos de extrapolação devem ser utilizados?
- Como combiná-las adequadamente?
- Como propor uma metodologia de estimação para os coeficientes do modelo?

Apesar de esforços em responder tais questões (ver, Nikolopoulos e Assimakopoulos, 2005, Nikolopoulos et al., 2008, 2012b), nenhuma conclusão concisa foi fornecida. Esta seção traz uma abordagem baseada em estudos empíricos do método Theta com o objetivo de responder a tais questões. Desta maneira, são defidos três pontos importantes. Primeiro, qual o número máximo de linhas-Theta para decompor a série temporal. Segundo, quais os métodos de extrapolação aplicados a cada linha-Theta. Terceiro, de que forma combinar as linhas-Theta e produzir as respectivas previsões. Para isso, será definida uma notação para o método Theta generalizado, ainda não existente; e por fim é proposto uma abordagem computacional para estimação dos coeficientes Theta do modelo.

#### **4.2.1 Número de linhas-Theta para decomposição**

É importante ressaltar que é possível utilizar qualquer número de linhas-Theta para decompor a série temporal. Porém, quando a quantidade de linhas-Theta aumenta é necessário maior esforço computacional para construir o método. Como já apresentado, o método com duas linhas-Theta resulta em um bom desempenho de acurácia. Nikolopoulos et al. (2012a), selecionou seu ajuste dentre cinco possibilidades de linhas-Theta variando no intervalo  $\theta = [-1; 3]$ . Contudo seus melhores resultados foram obtidos dentro do intervalo  $\theta = [0; 2]$ . Tirando vantagem desta experiência, propõe-se construir diferentes vertentes do método Theta e comparar o desempenho de cada um. Serão construídas vertentes com duas linhas (configuração usual), com três e quatro linhas-Theta, em que os coeficientes Theta de cada vertente não ultrapassarão o intervalo  $\theta = [0; 2]$ .

#### **4.2.2 Métodos de Extrapolação**

Baseando-se em evidências empíricas, Fildes et al. (1998) afirmam que qualquer método de extrapolação pode ser utilizado a fim de obter previsões de acordo com a experiência do analista. Utilizando da própria experiência, propõe-se utilizar três modelos de alisamento exponencial para extrapolação via modelos de estados de espaço, que são

o alisamento exponencial simples denotado por (ANN), alisamento exponencial de Holt (AAN) e alisamento exponencial de Holt com *dampen* (AA<sub>d</sub>N). A linha de regressão linear é usada na extrapolação de  $Z_t(0)$ .

A fim de simplificar a notação do método Theta em relação a construção de linhas e métodos de extrapolação, considerou-se a seguinte notação:

$$ME_0 - ME_1, \dots, -ME_k(\theta_0, \theta_1, \theta_2, \dots, \theta_k)$$

em que  $ME_i$  refere-se ao método de extrapolação aplicado à  $i$ -ésima linha-Theta  $Z_t(\theta_i)$ . Também por fins de simplificação, abreviou-se os nomes dos métodos de extrapolação da seguinte forma: **L**: modelo de regressão linear, **H**: modelo de alisamento exponencial de Holt, **D**: modelo de alisamento exponencial de Holt com *dampen* e **S**: modelo de alisamento exponencial simples. Seguindo tal notação, o método L-H-D-S(0,0.3,1.7,2) refere-se a vertente com quatro linhas-Theta; em que a primeira linha-Theta  $Z_t(0)$ , é extrapolada pelo modelo de regressão linear. A segunda linha-Theta  $Z_t(0.3)$ , é extrapolada pelo alisamento de Holt. A terceira linha-Theta  $Z_t(1.7)$ , é extrapolada pelo modelo de Holt com *dampen*. A quarta linha-Theta  $Z_t(2)$ , é extrapolada pelo modelo de alisamento exponencial simples. Seguindo esta notação, o modelo proposto por Assimakopoulos e Nikolopoulos (2000) é denotado por L-S(0,2).

### 4.2.3 Combinando as linhas-Theta

A combinação de linhas-Theta é um passo importante e necessário para a construção do método. De acordo com Assimakopoulos e Nikolopoulos (2000), combinações diferentes podem ser empregadas a fim de obter previsões. No entanto, a combinação proposta pelos mesmos é baseada em pesos iguais para as duas linhas-Theta. Constantinidou et al. (2012), abordam a combinação de linhas-Theta estimando os respectivos pesos para o método com duas linhas através de algoritmos de redes neurais, a fim de produzir as previsões. Este procedimento requer considerável esforço computacional e utiliza somente duas linhas-Theta para otimizar seus pesos.

Por simplicidade de interpretação, propõe-se uma combinação através de média ponderada para os pesos das linhas-Theta, a fim de definir os pesos e combiná-las para produzir as previsões finais. As evidências dos resultados mostram que esta abordagem funci-

ona muito bem e não requer grande esforço computacional, além de permitir a utilização de várias linhas-Theta para a combinação.

Considere o seguinte vetor de coeficientes Theta na construção do modelo,  $\Theta = (\theta_0, \dots, \theta_k)$ , em que  $\Theta \in \mathfrak{R}$  e  $\Theta \subset 1$ , incluindo  $k$  coeficientes Theta,  $k = 2, 3, \dots$  e  $k \in \mathfrak{N}$ . O vetor de pesos para combinação das respectivas linhas de extrapolação é dado por

$$W_\theta = \frac{1}{|\theta_0 - \theta_k|} \nabla \Theta \quad (4.19)$$

em que  $\nabla \Theta = \theta_k - \theta_{k-1}$  representa o operador diferença. Os pesos são então combinados com as respectivas linhas-Theta, produzindo então as previsões finais da forma

$$\hat{y}_{t+h} = \sum_{j=1}^{k-1} W_{\theta_j} \text{ME}_i(\theta_j). \quad (4.20)$$

**Exemplo 4.21** Considere  $\Theta$  ter 5 componentes,  $\Theta = (0, 0.3, 1, 1.5, 2)$ , que construirão respectivamente cada linha-Theta. Neste caso

$$\begin{aligned} W_\theta &= \frac{1}{2} \nabla \Theta \\ W_\theta &= (0.15, 0.35, 0.25, 0.25). \end{aligned}$$

Resumidamente, os pesos para combinação das linhas-Theta são dados pela distância entre cada componente  $\Theta = \theta_i$  e seu subsequente componente Theta. Os resultados desta proposta serão apresentados no Capítulo 6.

### 4.3 Estimação dos Parâmetros do Método Theta

Até o presente momento o método Theta foi tratado como um método de decomposição da série temporal em que a extrapolação das linhas-Theta e combinação deriva em previsões para a série temporal. No entanto, o coeficiente Theta não foi tratado como um parâmetro a ser estimado, além de permanecer uma questão aberta para esta abordagem. Nesta seção propõe-se uma abordagem com o objetivo de estimar o parâmetro Theta do modelo em questão.

Uma forma simples de obter estimativas do parâmetro Theta é simplesmente mini-

mizar o erro do modelo em relação a uma particular função perda  $\sum_i g(e_i)$ . No entanto, cálculos algébricos a fim de obter o estimador para o parâmetro Theta são por demais complicados quando aumenta-se a quantidade de linhas-Theta no modelo. Portanto, para este fim propõe-se a utilização de métodos computacionais de minimização. Em particular, neste trabalho será utilizado o algoritmo L-BFGS-B proposto por Byrd et al. (1995) e redefinido por Zhu et al. (1997). O algoritmo permite minimizar uma função particular, restringindo os parâmetros em limites definidos.

Durante o processo de otimização e estimação dos parâmetros serão considerados os modelos L-S( $0, \theta_a$ ), L-D-S( $0, \theta_b, 2$ ) e L-D-D-S( $0, \theta_c, \theta_d, 2$ ) em que  $(\theta_j \in \mathfrak{R})$   $j = a, b, c, d$  referem-se respectivamente aos parâmetros a serem estimados no modelo sob a seguinte restrição,  $\theta_a = [0; 2.5]$ ,  $\theta_b = [0; 2]$ ,  $\theta_c = [0; 1]$ ,  $\theta_d = [0; 2]$ . A função perda escolhida para critério de minimização é a soma do erro absoluto (MAE) dada por

$$\text{MAE} = \sum_i |e_i| \quad (4.22)$$

em que  $e_i = y_i - \hat{y}_i$ , desta forma tem-se um método de estimação análogo ao método de mínimos quadrados.

## 4.4 Intervalos de Confiança para Previsão do Método Theta

O cálculo de previsão para o método Theta não é uma tarefa fácil. Dado que o método pode utilizar diferentes modelos para extrapolar uma linha-Theta e depois combiná-las, o modelo torna-se extremamente geral e portanto calcular o intervalo de confiança torna-se um desafio. Existem três abordagens que possibilitam o cálculo dos intervalos de confiança para o método Theta. A primeira maneira de quantificar esta incerteza e obter a distribuição preditiva do método Theta é através de métodos de simulação da trajetória dos dados condicionado aos coeficientes Theta, métodos de extrapolação e combinação das linhas-Theta. Steffen (2010) trás uma discussão dos intervalos de confiança para o método Theta e um algoritmo bootstrap para esta finalidade.

Outra forma, seria utilizar abordagem de espaço de estados e calcular a distribuição preditiva de cada modelo da família Theta. De fato, esta é uma tarefa extremamente complicada e uma questão em aberto sob este ponto de vista de pesquisa. Um caso particular

desta abordagem é facilmente calculado para o modelo L-S(0,2), que como já mostrado neste capítulo, é um caso particular de alisamento exponencial com desvio, que é um modelo linear de espaço de estados. Para este modelo em particular, os intervalos de confiança são dados pela equação

$$\text{ICp}(100(1 - q)\%) = l_n + hb \pm z_q \sqrt{\sigma^2(h - 1)\alpha^2 + 1} \quad (4.23)$$

em que  $l_n$  é a previsão do alisamento exponencial simples,  $b = b_0/2$ , que é o coeficiente de inclinação da regressão linear simples dividido por 2;  $z_q$  denota o q-ésimo quantil de uma distribuição normal padrão,  $\alpha$  é o parâmetro de suavização do alisamento exponencial simples e  $h$  é o horizonte de previsão,  $h = 1, \dots, H$ . Em outras palavras,  $l_n + hb$  denota a média da previsão do alisamento exponencial com desvio e  $\sigma^2(h - 1)\alpha^2 + 1$  a variância da distribuição preditiva.

A terceira forma de encontrar intervalos de confiança para o método Theta é apresentada em Hyndman e Athanasopoulos (2012) e extremamente utilizada na prática quando não se conhece a distribuição preditiva, ou não deseja-se trabalhar com simulação da distribuição preditiva. Assumindo-se que os erros de previsão  $e_t = y_t - \hat{y}_t$  são não correlacionados e normalmente distribuídos, então um intervalo de previsão de  $100(1 - q)\%$  é dado por

$$\text{ICp}(100(1 - q)\%) = \hat{y}_{t+h} \pm z_q \sqrt{\hat{\sigma}^2} \quad (4.24)$$

onde  $\hat{\sigma}^2$  é variância estimada a partir dos dados ajustados  $\hat{\sigma}^2 = \sum_{t=1}^n e_t^2/n$ .

Hyndman e Athanasopoulos (2012) afirmam que esta abordagem funciona bem na prática, no entanto caso as suposições para o erro do modelo não forem verificadas, a aplicação desta abordagem é errônea.

# Capítulo 5

## Avaliando a Acurácia de Previsão

Durante os últimos 20 anos muitas medidas para avaliar o ajuste de séries temporais e a acurácia da previsão foram propostas. Apesar de inúmeras medidas disponíveis, seu uso indiscriminado pode levar a conclusões errôneas. Além disso, grande parte é inadequada principalmente quando se deseja medir o poder previsivo em uma grande quantidade de séries temporais. Neste capítulo serão apresentadas as medidas de acurácia usuais considerando uma revisão crítica das mesmas.

### 5.1 Uma revisão sobre medidas de acurácia

#### 5.1.1 Medidas Escala-Dependentes

As medidas escala-dependentes são as mais conhecidas dentre as medidas de acurácia. A característica que as torna tão utilizadas é que a escala da medida depende da escala dos dados e neste caso é aconselhável utilizá-las quando se deseja comparar ajustes de diferentes métodos aplicados a uma mesma série temporal, ou diferentes séries que estejam na mesma escala. O erro de previsão para um horizonte de  $h$  passos a frente é dado por:

$$e_t = y_{t+h} - \hat{y}_{t+h} \quad (5.1)$$

em que  $y_{t+h}$  é a série observada e  $\hat{y}_{t+h}$  é a previsão obtida segundo um modelo ou método.

Desta forma define-se as medidas escala-dependentes por:

- *Mean Square Error (MSE)*: média( $e_t^2$ )



- *Root Mean Square Error (RMSE)*:  $\sqrt{\text{média}(e_t^2)}$
- *Mean Absolut Error (MAE)*:  $\text{média}(|e_t|)$
- *Median Absolut Error (Mdae)*:  $\text{mediana}(|e_t|)$

As desvantagens destas medidas estão relacionadas a sensibilidade a outliers; principalmente ao MSE, por elevar valores de grande magnitude ao quadrado. Por esta razão, é aconselhável o uso do RMSE, MAE ou Mdae como alternativas (Armstrong e Collopy, 1992, Chatfield, 1988).

### 5.1.2 Medidas Baseadas em Erros Percentuais

Erros percentuais têm a vantagem de serem independentes da escala dos dados, e portanto são úteis na comparação de várias séries temporais que não estejam na mesma escala. O erro percentual  $p_t$  é dada por

$$p_t = 100e_t/y_t \quad (5.2)$$

as medidas mais comuns são:

- *Mean Absolute Percentage Error (MAPE)*:  $\text{média}(|p_t|)$
- *Median Absolute Percentage Error (MdAPE)*:  $\text{mediana}(|p_t|)$
- *Root Mean Square Percentage Error (RMSPE)*:  $\sqrt{\text{média}(p_t^2)}$
- *Root Median Square Percentage Error (RMdSPE)*:  $\sqrt{\text{mediana}(p_t^2)}$

A desvantagem destas medidas está relacionada a séries que assumem valores próximos a zero ou iguais a zero, levando  $p_t \rightarrow \infty$  ou a valores muito grandes. O MAPE e MdAPE também penalizam mais erros positivos do que erros negativos. Neste caso Makridakis (1993) propõe as medidas percentuais simétricas, empregadas na análise do banco *M3-Competition* (ver, Makridakis e Hibon, 2000).

- *Symetric Mean Absolut Percentage Error (sMAPE)*:  $\text{média}(200|e_t|/(y_t + \hat{y}_t))$
- *Symetric Median Absolut Percentage Error (sMdAPE)*:  $\text{mediana}(200|e_t|/(y_t + \hat{y}_t))$

Os problemas decorrentes para medidas não simétricas podem ser menos severos em medidas simétricas, no caso desta medida os valores são limitados a 200% evitando a divergência da medida. A medida também permite porcentagens negativas, o que não faz sentido em muitas aplicações e poderia ser facilmente corrigido aplicando o módulo no denominador da equação (5.2). No entanto, esta precaução não é geralmente adotada. Por fim, a simetria não pode ser considerada segundo a penalidade que a medida leva em consideração. Por exemplo, para um mesmo valor de  $y_t$  e valores baixos de previsão, a medida retorna valores altos para sMAPE, já para valores discrepantes de previsão a penalidade é mais baixa.

### 5.1.3 Medidas Baseadas em Erros Relativos

Outra maneira de medir acurácia é através de erros relativos, que são definidos através da equação:

$$r_t = e_t/e_t^* \quad (5.3)$$

em que  $e_t^*$  é o erro de previsão obtido a partir de um método de referência. Usualmente o método de referência é o passeio aleatório, que é equivalente a um modelo ARIMA(0,1,0), em que a previsão é dada pela última observação da série temporal. Este método também é conhecido como método ingênuo ou *naive*. Desta forma define-se

- Mean Relative Absolute Error (MRAE): média( $|r_t|$ ).
- Median Relative Absolute Error (MdRAE): mediana( $|r_t|$ ).
- Geometric Mean Relative Absolute Error (GMRAE): média geométrica( $|r_t|$ ).

Uma deficiência de medidas de erro relativo é o caso em que  $e_t^*$  assume valores pequenos.  $r_t$  tem variância infinita pois  $e_t^*$  tem densidade positiva no 0. Um caso particular é quando  $e_t$  e  $e_t^*$  seguem distribuição normal, o que leva  $r_t$  a seguir distribuição Cauchy.

### 5.1.4 Medidas Relativas

Além de erros relativos também existem medidas relativas. Por exemplo, denote MAEr o erro absoluto médio de um dado método de referência. Neste caso o MAE

relativo é dado por  $MAE_{rel} = MAE/MAE_r$ . Podem ser usadas quaisquer medidas escala-dependentes ou erros percentuais para criar medidas relativas.

Uma vantagem destes métodos é sua interpretação. Por exemplo, o MAE relativo mede a melhora da previsão de um dado método em relação a um método de referência. Quando  $MAE_{rel}$  é menor que 1, o método proposto supera a capacidade preditiva do método de referência, e quando  $MAE_{rel}$  é maior que 1 o método de referência supera a capacidade preditiva do método proposto.

Como desvantagens, medidas relativas podem somente ser calculadas quando existem várias previsões na mesma série, portanto não podem ser utilizadas para medir a acurácia para um único horizonte de previsão.

### 5.1.5 Medidas de Erro de Escala

Medidas relativas e medidas baseadas em erros relativos buscam remover a escala dos dados e comparar as previsões obtidas com algum método de referência, no entanto também apresentam desvantagens. Hyndman e Koehler (2006) sugerem o uso de uma nova medida para avaliar a acurácia de previsão baseada nos métodos descritos anteriormente, escalonando o erro, baseado no erro absoluto médio da amostra treinamento em relação ao passeio aleatório do erro a um passo de previsão. A medida é chamada de ASE (*Absolute Scale Error*), dada pela equação

$$q_t = \frac{|e_t|}{\frac{1}{n-m} \sum_{i=m+1}^n |y_i - y_{i-m}|} \quad (5.4)$$

em que  $m$  é o período de sazonalidade da série temporal. Desta forma derivam-se as seguintes medidas:

- Mean Absolute Scaled Error (MASE):  $\text{média}(q_t)$ .
- Median Absolute Scaled Error (MdASE):  $\text{mediana}(q_t)$ .

A interpretação da medida é idêntica a medidas relativas, ou seja, quando MASE é menor que 1, em média tem-se erros menores comparados ao passeio aleatório a um passo de previsão. Além disso é possível calcular vários horizontes de previsão. Outra vantagem é que o MASE é mais robusto a outliers, tem uma escala significativa, é amplamente aplicável, e não é afetado por problemas de indeterminação. A única maneira de

indeterminações ocorrerem é no caso em que a série tem todas as observações iguais, ou seja, a variância da série temporal é igual a zero.

Considerando esta abordagem, serão utilizadas como métricas de acurácia neste trabalho o MASE e o sMAPE. Como já justificado, o MASE é uma métrica que em relação às demais medidas de acurácia não apresenta problemas de indeterminação, com exceção do caso em que a série em questão apresentar variância igual a zero, o que muito raramente ocorre. Já o sMAPE é uma medida utilizada recursivamente por outros trabalhos (ver, Crone et al., 2011, Makridakis e Hibon, 2000) e não resulta em problemas para o caso de dados positivos, o que é comprovado para a base de dados M1 e M3 considerados para aplicação.

# Capítulo 6

## Resultados

Neste capítulo serão apresentados e discutidos os resultados referentes à aplicação empírica na base de dados *M1-Competition* e *M3-Competition*, considerando os métodos discutidos até o momento,

- Modelos de alisamento exponencial via espaço de estados;
- Métodos ARIMA automáticos; e
- Método Theta generalizado.

Os métodos serão comparados de acordo com seus respectivos horizontes de previsão. Serão utilizadas duas medidas de acurácia para avaliação do erro de previsão: o *Mean Average Scalar Error* (MASE) e o *simetric Mean Absolute Error* (sMAPE). O sMAPE oferece uma interpretação percentual dos erros de previsão, o que é intuitivo e facilmente interpretável, além de ser geralmente utilizado nas competições envolvendo a base M3 (Crone et al., 2011, Makridakis e Hibon, 2000). Também, o sMAPE é utilizado como medida de classificação para definir o método vencedor nestas competições. No entanto, devido as deficiências do sMAPE apresentados no Capítulo 5, (ver também, Goodwin e Lawton, 1999, Hyndman e Koehler, 2006); consideraremos o MASE como critério de acurácia para definir o modelo vencedor da aplicação.

## 6.1 As bases de dados *M1* e *M3-Competition*

Neste trabalho utilizaremos a base de dados *M3-Competition* para aplicação de todos os modelos e métodos discutidos. Especificamente o método Theta será aplicado à base *M1-Competition* para estimação empírica de seus coeficientes. Mais detalhes são dados na Seção 6.4.

Ambas as bases apresentam séries temporais univariadas segmentadas segundo o período de sazonalidade: anual, trimestral e mensal; para a base M1. Anual, trimestral, mensal e outros; para a base M3; em que a segmentação outros tem período anual.

A previsão é feita considerando diferentes horizontes de previsão de acordo com esta segmentação:

- Para dados anuais, é requerido um horizonte de previsão igual a seis passos.
- Para dados trimestrais, e outros é requerido um horizonte de previsão igual a oito passos.
- Para dados mensais, é requerido um horizonte de previsão igual a dezoito passos.

A série temporal foi dividida em amostra treinamento, referente aos dados para ajuste; e amostra teste, utilizada para avaliação do erro de previsão. Salientamos que o horizonte requerido e a divisão da série em amostra treinamento e teste respectivamente foram definidos por Makridakis e Hibon (2000) durante a competição M3.

As séries temporais da base M1 advém de setores Micro-economicos, Macro-economico, Industrial e Demográfico. A segmentação da base M1 é apresentada na Tabela 6.1.

Tabela 6.1: Segmentação das séries temporais do base M1

Segmentação	Micro	Indústria	Macro	Demog	Total
Anual	57	35	59	30	181
Trimestral	42	18	104	39	203
Mensal	203	183	156	75	617
Total	302	236	319	144	1001

As séries temporais da base M3 advém dos setores Micro-economico, Macro-economico, Industrial, Financeiro, Demográfico e Outros. A segmentação da base M3 é apresentada na Tabela 6.2.

Tabela 6.2: Segmentação das séries temporais do base M3

<b>Segmento</b>	<b>Micro</b>	<b>Industria</b>	<b>Macro</b>	<b>Financeira</b>	<b>Demog</b>	<b>Outros</b>	<b>Total</b>
Anual	146	102	83	58	245	11	645
Trimestral	204	83	336	76	57	0	756
Mensal	474	334	312	145	111	52	1428
Outro	4	0	0	29	0	141	174
Total	828	519	731	308	413	204	3003

As Tabelas 6.3,6.4 e 6.6, apresentadas neste Capítulo resumem a performance geral dos modelos e métodos aplicados à base M3. A construção das Tabelas leva em consideração todos os horizontes de previsão requeridos para cada segmentação da base M3 em relação as medidas MASE e sMAPE. Desta forma, as Tabelas foram construídas da seguinte forma:

1. aplica-se um modelo (qualquer) para toda a base de dados M3 fixando 18 passos de previsão (máximo requerido). Guarda-se essas previsões nas linhas de uma matriz com dimensão 3003 por 18.
2. cria-se outras matrizes para as segmentações anual, trimestral, mensal e outros, com dimensão 645 por 6, 756 por 8, 1428 por 18 e 174 por 8 respectivamente.
3. calcula-se as medidas ASE e sAPE, considerando as previsões obtidas no passo 1 e segmentadas como descrito no passo 2. Desta forma são criadas novas matrizes ASE e sAPE para cada segmentação anual, trimestral, mensal e outros com dimensão 645 por 6, 756 por 8, 1428 po 18 e 174 por 8. Por exemplo, as linhas da matriz ASE.anual contém o ASE estimado de acordo as respectivas previsões e tem dimensão 645 por 6.
4. A partir da matriz ASE ou sAPE, calcula-se a média para cada horizonte de previsão, ou seja, média das colunas da matriz. Desta forma, obtém-se um vetor MASE e sMAPE para cada segmentação da base.
5. Os vetores obtidos no passos 4 são concatenados em linhas, retira-se novamente a média para cada horizonte de previsão. Desta forma, é obtida a média do MASE geral e a média do sMAPE geral, considerando todas as previsões do banco M3.

## 6.2 Resultados Para os Modelos de Alisamento Exponencial

Para aplicação dos modelos de alisamento exponencial foram selecionados os modelos que poderiam ser aplicados a toda a base de dados M3 individualmente, desconsiderando possibilidade que poderiam resultar em problemas de otimização, como modelos não-lineares, dado que o banco não possui séries temporais no domínio negativo. Cada modelo considerado extrapolou a série temporal e forneceu previsões condicionado à existência de sazonalidade na mesma. Para identificar a sazonalidade da série temporal, utilizou-se o teste T-Student para diferença de lag de ordem 4 para séries trimestrais e 12 para séries mensais. Os modelos selecionados foram:

- ANN: Alisamento exponencial simples.
- AAN: Alisamento exponencial de Holt.
- $AA_dN$ : Alisamento exponencial de Holt com *dampen*.
- *Damped* aditivo (Dadt): Ajusta-se o modelo  $AA_dA$  em séries que apresentam presença da componente sazonal; na ausência componente sazonal o modelo  $AA_dN$  é então ajustado.
- *Damped* multiplicativo (Dmult): Ajusta-se o modelo  $AA_dM$  em séries que apresentam presença da componente sazonal; na ausência da componente sazonal o modelo  $AA_dN$  é então ajustado.
- Holt-Winters aditivo (HWadt): Ajusta-se o modelo AAA em séries que apresentam presença da componente sazonal; na ausência da componente sazonal o modelo AAN é então ajustado.
- Holt-Winters multiplicativo (HWmult): Ajusta-se o modelo AAM em séries que apresentam presença da componente sazonal; na ausência da componente sazonal o modelo AAN é então ajustado.

Também o método ETS automático proposto por Hyndman et al. (2002) foi considerado para aplicação, com o objetivo de comparar sua performance com os demais



resultados. A Tabela 6.3 apresenta os resultados de acordo com as medidas de acurácia MASE e sMAPE para a aplicação geral da base de dados M3, os resultados das demais segmentações anual, trimestral, mensal e outros são discutidos e podem ser encontrados no Apêndice A.

Tabela 6.3: Resultado dos modelos de alisamento exponencial aplicados à base M3: Todos as séries.

Método	Horizontes de previsão						Média do horizonte de previsão					Ranking Médio
	1	4	6	8	12	18	1-4	1-6	1-8	1-12	1-18	
Média MASE geral												
<b>Dadt</b>	<b>0.98</b>	<b>2.02</b>	<b>2.40</b>	<b>2.07</b>	<b>1.70</b>	<b>2.30</b>	<b>1.49</b>	<b>1.77</b>	<b>1.82</b>	<b>1.82</b>	<b>1.94</b>	<b>1.37</b>
Dmult	1.00	<b>2.02</b>	2.42	<b>2.07</b>	<b>1.70</b>	2.40	1.51	1.78	<b>1.82</b>	1.83	1.96	1.91
ETS	<b>0.98</b>	2.15	2.88	2.17	1.80	2.70	1.54	1.91	1.95	1.92	2.07	3.50
HWadt	1.00	2.10	2.60	2.30	2.10	2.90	1.53	1.85	1.94	1.98	2.23	4.15
HWmult	1.00	2.10	2.60	2.30	2.10	3.10	1.53	1.84	1.92	1.99	2.25	4.30
AAdN	1.18	2.30	2.68	2.40	2.20	3.10	1.71	2.01	2.09	2.17	2.45	6.35
ANN	1.30	2.62	3.20	3.17	2.20	3.00	1.95	2.32	2.50	2.40	2.56	6.93
AAN	1.20	2.38	2.85	2.67	2.90	4.20	1.76	2.09	2.22	2.43	2.92	7.48
Média sMAPE geral												
<b>Dmult</b>	<b>6.55</b>	<b>11.70</b>	<b>13.32</b>	<b>11.40</b>	<b>13.60</b>	18.90	<b>9.16</b>	<b>10.44</b>	<b>10.57</b>	<b>11.61</b>	<b>13.43</b>	<b>1.20</b>
Dadt	6.60	11.75	13.35	11.30	13.80	<b>18.50</b>	9.26	10.60	10.76	11.78	13.65	2.20
ETS	6.67	12.00	14.05	11.50	13.80	19.30	9.34	10.75	10.85	11.84	13.71	2.70
HWadt	6.70	12.65	14.85	12.57	15.80	21.80	9.77	11.41	11.71	12.91	15.40	5.20
HWmult	6.70	12.62	14.67	12.60	16.10	22.40	9.66	11.16	11.39	12.81	15.24	5.43
ANN	7.80	13.35	14.75	13.17	14.30	20.60	10.59	11.89	12.14	12.89	15.11	5.48
AAdN	7.53	12.88	14.10	12.27	14.90	20.90	10.17	11.46	11.61	12.78	15.17	5.89
AAN	7.60	13.70	15.65	13.63	17.60	24.40	10.63	12.22	12.53	14.17	17.09	7.91

O primeiro ponto referente à análise da Tabela 6.3 e Tabelas do Apêndice A é que efetivamente existe uma discordância relativa entre as medidas de acurácia MASE e sMAPE, o que já era esperado devido às diferenças intrínsecas de cada medida como apresentado na Capítulo 5. O MASE é uma medida mais robusta, e portanto menos sensível a outliers, o que leva a uma dificuldade em diferenciar o desempenho de modelos similares, como é o caso dos modelos *Damped*. Já o sMAPE é uma medida mais sensível, o que propicia em apontar de forma mais clara para diferenças de desempenho entre os modelos.

Considerando uma análise de acordo com as segmentações do banco M3, para os dados anuais de acordo com a medida de acurácia MASE, os melhores modelos que preveem esta categoria são o alisamento exponencial de Holt com *dampen* AAdN e o alisamento exponencial de Holt. Os métodos que apresentaram uma performance insatisfatória foram ETS automático e alisamento exponencial simples.

Em relação a segmentação trimestral destacam-se os modelos *Damped* multiplicativo e *Damped* aditivo. Apesar do método ETS automático não ter sido bem classificado em relação ao MASE, para a medida sMAPE este método tem seu primeiro lugar no Ranking

médio, com diferença de 0.07% em relação ao segundo colocado *Damped* multiplicativo.

Da mesma forma para os dados mensais, cujo período de sazonalidade é igual a 12 meses, os métodos *Damped* aditivo e ETS automático destacam-se novamente. Como já esperado modelos que não modelam a componente sazonal foram classificados nos últimos lugares do ranking. O modelo de Holt-Winters aditivo, muito aplicado na prática para modelagem da componente sazonal, esboça uma diferença percentual de 2.07 % em relação ao método *Damped* multiplicativo. Isto indica que a modelagem do parâmetro de amortização da tendência contribui significativamente para a redução do erro de previsão em geral.

O segmento Outros, contém uma miscelânea de séries temporais, que por sua vez não apresentam alta volatilidade, desta forma percebe-se que o erro de previsão é bem mais baixo em relação às outras segmentações da base M3. Neste segmento o método ETS e o modelo  $AA_dN$  destacaram-se na previsão destas séries temporais tanto para o MASE quanto para o sMAPE.

Considerando o ranking médio da Tabela 6.3 dado através do MASE, o modelo *Damped* aditivo sobrepuja o modelo com *Damped* multiplicativo. Contudo, a diferença escalar de um modelo para o outro é de 0.02 pontos escalares na média da previsão de (1-18). Ao longo dos horizontes de previsão esta diferença é zero para alguns casos, o que indica que os modelos são estatisticamente iguais. Apesar desta conclusão, quando se trata de competições de modelos, esta pequena diferença é suficiente para classificar um modelo candidato ao primeiro lugar no ranking. Desta forma, deve-se considerar que ambos os modelos obtiveram boa performance em relação as previsões obtidas e comparações com as séries teste, no entanto o modelo *Damped* aditivo é o vencedor entre os modelos de alisamento exponencial.

A notável performance dos modelos *Damped* aditivo e multiplicativo chamam a atenção. Estes modelos sobrepujam a performance do método ETS automático que por sua vez é um procedimento mais sofisticado por considerar até 30 modelos diferentes de alisamento exponencial para seleção e previsão. Devido ao método ETS automático escolher o modelo que apresenta o menor valor do critério de Akaike para previsão, ou seja, supomos que o modelo escolhido reflete um modelo parcimonioso e conseqüentemente que ajusta-se bem aos dados; vê-se que nem sempre esta seleção através de "*bruta força*", como indicado pela Figura 6.1 é a melhor opção para previsão de séries temporais.

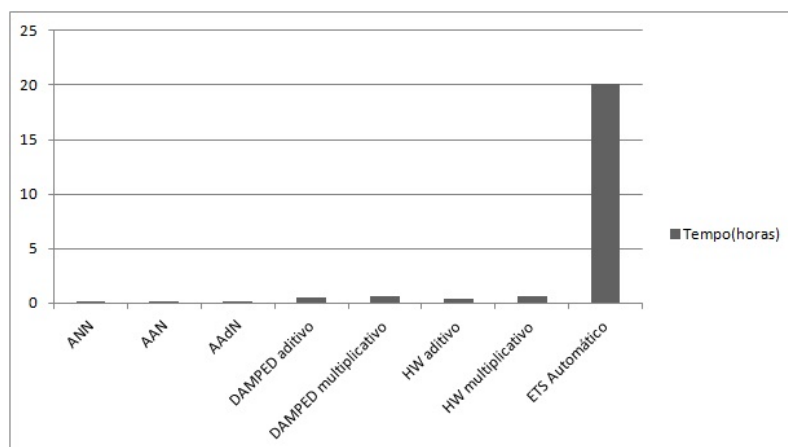


Figura 6.1: Tempo gasto no ajuste da base M3 referente aos modelos alisamento exponencial

### 6.3 Resultados para os métodos ARIMA Automáticos

Dada a importância dos modelos ARIMA no contexto de previsão de séries temporais, nesta seção serão avaliados os resultados dos procedimentos ARIMA automáticos propostos por Hyndman e Khandakar (2008) que incluem:

- Algoritmo Auto Arima com *stepwise*; e
- Algoritmo Auto Arima sem *stepwise*, em que é selecionado o modelo que apresente o menor AIC.

Os resultados gerais são apresentados na Tabela 6.4, os demais resultados seguem no Apêndice B.

Tabela 6.4: Resultado dos procedimentos Auto ARIMA aplicados à base M3: Todos as séries.

Método	Horizontes de previsão						Média do horizonte de previsão					Ranking Médio
	1	4	6	8	12	18	14	16	18	112	118	
	Média MASE geral											
Auto ARIMA ( <i>stepwise</i> )	<b>0.98</b>	<b>2.00</b>	<b>2.45</b>	<b>2.17</b>	<b>1.80</b>	2.50	<b>1.49</b>	<b>1.79</b>	<b>1.86</b>	<b>1.85</b>	<b>2.01</b>	<b>1.41</b>
Auto ARIMA	<b>0.98</b>	2.08	2.52	<b>2.17</b>	<b>1.80</b>	<b>2.40</b>	1.53	1.82	1.88	1.88	<b>2.01</b>	1.59
	Média sMAPE geral											
Auto ARIMA ( <i>stepwise</i> )	<b>6.83</b>	<b>11.82</b>	<b>13.55</b>	11.73	14.70	<b>19.60</b>	<b>9.36</b>	<b>10.64</b>	<b>10.81</b>	<b>12.01</b>	<b>14.12</b>	<b>1.28</b>
Auto ARIMA	<b>6.83</b>	12.12	13.70	<b>11.70</b>	<b>14.60</b>	19.90	9.51	10.80	10.92	12.13	14.19	1.72

O primeiro ponto que chama atenção da aplicação utilizando ambos procedimentos ARIMA automáticos é que os métodos de fato são muito próximos em relação as medidas de acurácia, o que é um indicativo de que o método ARIMA com *stepwise* de fato

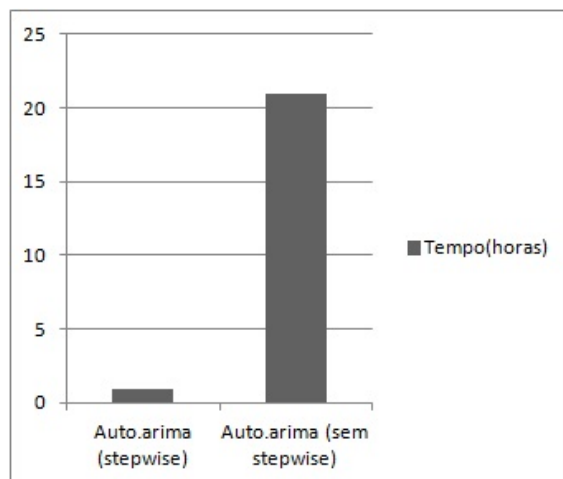


Figura 6.2: Tempo computacional gasto no ajuste banco M3 dos métodos auto.arima

cumprir o que promete, ou seja, selecionar um modelo ARIMA adequado para previsão de forma mais rápida e automática do que a proposta de ajustar todos os modelos possíveis e selecionar o que resulte no menor AIC. Além do mais, dentro do ranking médio da Tabela 6.4 o modelo ARIMA automático com *stepwise* teve a melhor performance comparando tanto o sMAPE quanto o MASE. A rotina sem *stepwise*, além de demandar um esforço computacional maior como mostrado na Figura 6.2, apenas sobrepujou a performance de seu adversário para séries com presença de sazonalidade. Isso se deve provavelmente à grande variedade de modelos sazonais ARIMA que podem ser utilizados em sua totalidade apenas através deste método, pois o algoritmo *stepwise* não ajusta cada possível modelo ARIMA aos dados.

## 6.4 Resultados Para o Método Theta

De acordo com toda metodologia proposta no Capítulo 4, com o objetivo de estabelecer uma nova abordagem na construção do método Theta generalizado, nesta seção serão apresentadas três diferentes abordagens a fim de avaliar o poder previsivo de cada um. A primeira abordagem consiste em uma avaliação empírica de diversos métodos Theta aplicados à base *M1-Competition*. A segunda abordagem consiste em selecionar os modelos que apresentaram melhor performance em relação à base M1 e aplicá-los à base M3. A terceira abordagem consiste na aplicação do procedimento de estimação dos coeficientes Theta à base M3.

### 6.4.1 Aplicação à base de dados M1

Será feita uma avaliação empírica do método Theta aplicado à base de dados M1. Serão ajustados 5061 diferentes métodos Theta à base, a construção dos modelos foi definida da seguinte forma:

1) *Método Theta com duas linhas*: o método Theta com duas linhas foi definido da seguinte forma L-ME(0,2), em que L é a linha de regressão linear e ME refere-se ao método de extrapolação aplicado a respectiva linha-Theta. Foram utilizados os três métodos propostos no Capítulo 4 para extrapolação da linha ME(2), ou seja, S: alisamento exponencial simples, H: alisamento exponencial de Holt, D: alisamento exponencial de Holt com *dampen*.

2) *Método Theta com três linhas*: o método Theta com três linhas foi definido da seguinte forma: L-ME-S(0, $\theta_1$ ,2), em que ME é extrapolado a partir dos métodos S, H e D para cada diferente combinação de linhas-Theta,  $\theta_1$  é escolhido sequencialmente de acordo com os valores  $\theta_1 = 0.1, 0.2, \dots, 1.9$ .

3) *Método Theta com quatro linhas*: da mesma forma estrutura-se o método Theta com quatro linhas, definido da seguinte forma L-ME<sub>2</sub>-ME<sub>3</sub>-S(0, $\theta_2$ , $\theta_3$ ,2), em que ME<sub>2</sub> e ME<sub>3</sub> são extrapolados a partir dos métodos S, H e D para cada diferente combinação de linhas-Theta,  $\theta_2$  e  $\theta_3$  são escolhidos sequencialmente de acordo com os valores  $\theta_2 = 0.1, 0.2, \dots, 0.9$  e  $\theta_3 = 1.1, 1.2, 1.3, \dots, 1.9$ . A melhor performance para cada modelo é selecionada e rankeada como mostrado na Tabela 6.5 .

Tabela 6.5: Resultados dos modelos Theta aplicados ao banco M1 de acordo com MASE

Modelo	Média Geral MASE				Ranking Geral
	Anual	Trimestral	Mensal	Todos	
<b>L-S (0,2)</b>	4.20	<b>1.97</b>	1.98	<b>2.39</b>	<b>1</b>
L-D (0,2)	3.47	<b>2.09</b>	2.05	2.44	2
L-H (0,2)	<b>3.36</b>	2.22	2.19	2.56	3
<b>L-D-S (0,1,7,2)</b>	3.68	1.93	<b>1.97</b>	<b>2.36</b>	<b>1</b>
L-S-D (0,1,5,2)	3.83	<b>1.92</b>	1.98	2.37	2
L-D-H (0,1,9,2)	3.45	1.97	2.04	2.38	3
L-H-D (0,1,2,2)	3.47	1.93	2.04	2.39	4
L-S-H (0,1,9,2)	4.12	2.06	<b>1.97</b>	2.42	5
L-D-D (0,0,3,2)	3.46	2.08	2.08	2.43	6
L-S-S (0,0,8,2)	4.29	2.14	<b>1.97</b>	2.45	7
L-H-H (0,1,4-2)	<b>3.37</b>	2.10	2.14	2.50	8
L-H-S (0,0,3,2)	4.63	2.27	1.98	2.51	9
<b>L-D-D-S (0,0,5,1,4,2)</b>	3.61	<b>1.91</b>	<b>1.96</b>	<b>2.34</b>	<b>1</b>
L-H-D-S (0,0,3,1,4,2)	3.59	1.96	2.00	2.37	2
L-D-D-D (0,0,2,1,4,2)	3.46	1.95	2.02	2.37	3
L-H-D-D (0,0,1,1,5,2)	3.48	1.92	2.03	2.37	4
L-S-D-S (0,0,6,1,6,2)	3.93	2.02	1.97	2.38	5
L-D-S-D (0,0,4,1,7,2)	3.89	2.01	1.97	2.38	6
L-H-S-D (0,0,2,1,7,2)	3.89	1.99	1.97	2.38	7
L-S-D-D (0,0,5,1,5,2)	3.70	1.98	2.01	2.38	8
L-D-S-H (0,0,2,1,7,2)	3.86	2.06	1.97	2.39	9
L-S-S-D (0,0,7,1,6,2)	4.03	2.01	<b>1.96</b>	2.39	10
L-D-H-D (0,0,1,1,2,2)	3.46	1.97	2.05	2.39	11
L-H-S-H (0,0,2,1,8,2)	3.91	2.06	1.98	2.40	12
L-H-H-S (0,0,2,1,2,2)	3.95	2.06	1.97	2.40	13
L-D-D-H (0,0,1,1,7,2)	3.44	2.00	2.05	2.40	14
L-H-H-D (0,0,2,1,2,2)	3.43	2.00	2.07	2.41	15
L-H-D-H (0,0,2,1,9,2)	3.42	2.04	2.06	2.41	16
L-D-S-S (0,0,3,1,8,2)	4.15	2.11	<b>1.96</b>	2.42	17
L-S-S-H (0,0,7,1,8,2)	4.13	2.12	<b>1.96</b>	2.42	18
L-S-H-S (0,0,7,1,2,2)	4.17	2.12	<b>1.96</b>	2.42	19
L-D-H-S (0,0,2,1,7,2)	3.50	2.04	2.07	2.43	20
L-H-S-S (0,0,3,1,9,2)	4.16	2.09	1.97	2.43	21
L-S-H-D (0,0,1,1,4,2)	3.92	2.06	2.02	2.44	22
L-S-D-H (0,0,1,1,7,2)	3.91	2.10	2.01	2.44	23
L-S-S-S (0,0,8,1,1,2)	4.32	2.14	1.97	2.45	24
L-D-H-H (0,0,1,1,4,2)	3.36	2.14	2.14	2.51	25
L-S-H-H (0,0,1,1,4,2)	<b>3.35</b>	2.14	2.16	2.52	26
L-H-H-H (0,0,1,1,4,2)	<b>3.35</b>	2.14	2.16	2.52	27

Os resultados da Tabela 6.5 indicam principalmente dois pontos importantes. O primeiro é que a partir dos métodos rankeados é possível perceber que aqueles que ficaram na primeira colocação do ranking advém da mesma família de métodos, em que L-S(·) é

um caso particular do método L-D-S( $\cdot$ ) que é um caso particular do método L-D-D-S( $\cdot$ ), ou seja, dos mais de 5000 modelos ajustados a base M1, uma mesma família de métodos destacou-se em relação a performance de previsão.

O segundo ponto a se destacar na análise é que conforme aumenta-se o número de linhas de extrapolação no modelo o erro de previsão diminui, isso pode ser visto com clareza nas Figuras 6.3 e 6.4, tanto para o banco M1 quanto para o banco M3.

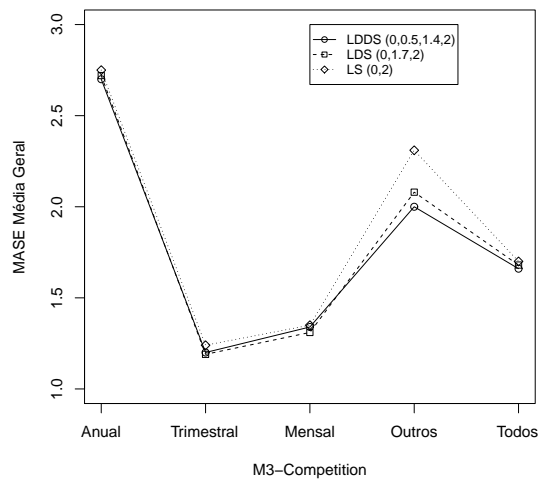
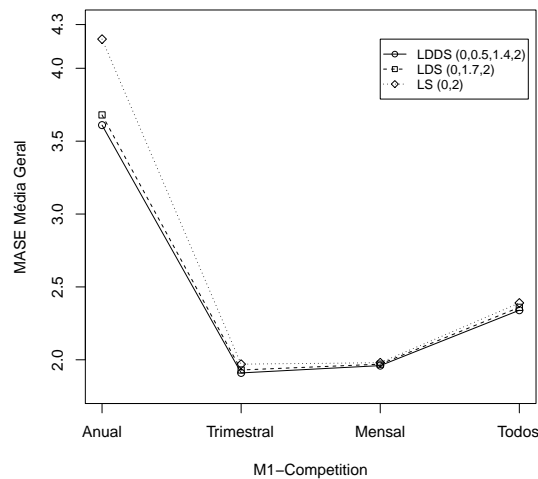


Figura 6.3: MASE média geral para a base *M1-Competition* Figura 6.4: MASE média geral para a base *M3-Competition*

### 6.4.2 Aplicação à Base de Dados M3 Para o Método Theta Generalizado

A partir dos resultados anteriores aplicou-se os seguintes métodos Theta à base de dados M3, são eles: L-D-D-S(0,0.5,1.4,2), L-D-D-S(0,0.5,1.5,2), L-D-S(0,1.7,2), L-S(0,2). Os quais foram rankeados a partir do banco M1. Também são avaliados os métodos em que é estimado o parâmetro Theta; L-S(0, $\theta_a$ ), L-D-S(0, $\theta_b$ ,2) e L-D-D-S(0, $\theta_c$ , $\theta_d$ ,2). Os resultados gerais são encontrados na Tabela 6.6, os demais resultados encontram-se no Apêndice C.

Tabela 6.6: Resultados do método Theta generalizado aplicado à base M3: Todos as séries.

Método	Horizontes de previsão						Média do horizonte de previsão					Ranking Médio
	1	4	6	8	12	18	1-4	1-6	1-8	1-12	1-18	
Média MASE geral												
<b>L-D-D-S(0,0.5,1.5,2)</b>	<b>0.80</b>	<b>1.80</b>	<b>2.27</b>	2.03	<b>1.40</b>	<b>2.10</b>	<b>1.31</b>	<b>1.60</b>	<b>1.68</b>	<b>1.58</b>	<b>1.66</b>	<b>2.04</b>
L-D-D-S(0,0.5,1.4,2)	<b>0.80</b>	1.83	2.30	2.07	<b>1.40</b>	<b>2.10</b>	<b>1.31</b>	<b>1.60</b>	1.69	<b>1.58</b>	<b>1.66</b>	2.26
L-D-S(0,1.7,2)	0.85	1.89	<b>2.27</b>	<b>2.00</b>	<b>1.40</b>	2.12	1.33	1.61	1.69	1.59	1.68	2.37
L-S(0,2)	<b>0.80</b>	1.90	2.42	2.20	<b>1.40</b>	<b>2.10</b>	1.36	1.68	1.78	1.65	1.70	3.33
Média sMAPE geral												
<b>L-D-D-S(0,0.5,1.4,2)</b>	<b>6.50</b>	<b>11.10</b>	<b>12.60</b>	<b>10.80</b>	<b>13.00</b>	<b>17.80</b>	<b>8.84</b>	<b>10.01</b>	<b>10.12</b>	<b>11.12</b>	<b>12.83</b>	<b>1.24</b>
L-D-D-S(0,0.5,1.5,2)	<b>6.50</b>	11.15	12.62	10.83	<b>13.00</b>	17.90	8.86	10.02	10.14	11.14	12.86	1.85
L-D-S(0,1.7,2)	6.58	11.20	12.78	10.87	13.20	18.00	8.94	10.12	10.23	11.25	12.99	3.09
L-S(0,2)	6.62	11.32	12.97	11.03	13.20	18.10	9.07	10.26	10.38	11.35	13.09	3.83
<b>L-D-D-S(0,<math>\theta_c, \theta_d, 2</math>)</b>	<b>6.62</b>	<b>11.18</b>	<b>12.75</b>	<b>10.77</b>	<b>13.10</b>	<b>17.80</b>	<b>8.96</b>	<b>10.12</b>	<b>10.21</b>	<b>11.20</b>	<b>12.89</b>	<b>1.07</b>
L-D-S(0, $\theta_b, 2$ )	<b>6.62</b>	11.38	12.95	11.23	13.30	18.20	9.03	10.24	10.39	11.41	13.14	1.93
L-S(0, $\theta_a$ )	7.75	11.90	13.68	11.53	14.50	19.30	9.82	10.97	11.05	12.27	14.10	3.00

A Tabela 6.6 apresenta dois tipos de resultados, os quais estão separados por uma linha: os resultados relacionados ao método empírico, em que não houve estimação da componente Theta, ou seja, esta foi selecionada de acordo com a performance dos métodos da Tabela 6.5, e os resultados do modelo Theta em que estimou-se os coeficientes Theta através do algoritmo L-BFGS-B.

Partindo dos resultados empíricos, a melhor performance de previsão é apresentada pelo método L-D-D-S(0,0.5,1.5,2) de acordo com o MASE. Esta é a melhor performance em relação a todos os métodos participantes da competição. O método também reproduz a boa performance dentro das segmentações anuais, trimestrais, mensais e outros da base M3, sendo de fato o melhor método para previsão apresentado.

Apesar da boa performance de previsão e funcionamento adequado dos modelos Theta segundo abordagem de estimação dos parâmetros para decomposição das linhas-Theta, os modelos estimados não sobrepujaram a performance dos métodos em que fixou-se os coeficientes Theta empiricamente. Isso indica que até o momento o grande esforço computacional exigido durante o processo de estimação não é suficiente para aumentar o poder de previsão. Desta forma, recomenda-se iniciar o processo de modelagem através



do método L-D-D-S(0,0.5,1.5,2), e utilizá-lo como um método de referência, para a partir de então buscar outros modelos que superem sua performance. Por exemplo, aumentando ou reduzindo componentes Theta, ou utilizando o modelo do ponto de vista de estimação. A Figura 6.5 mostra o tempo gasto durante a aplicação de cada modelo à base M3.

A Tabela 6.6 também sugere que quando a quantidade de linhas-Theta no modelo aumentam, da mesma forma o erro de previsão é reduzido. Isto ocorreu tanto no banco *MI* quanto no banco *M3* para os modelos estimados e também para os modelos empíricos.

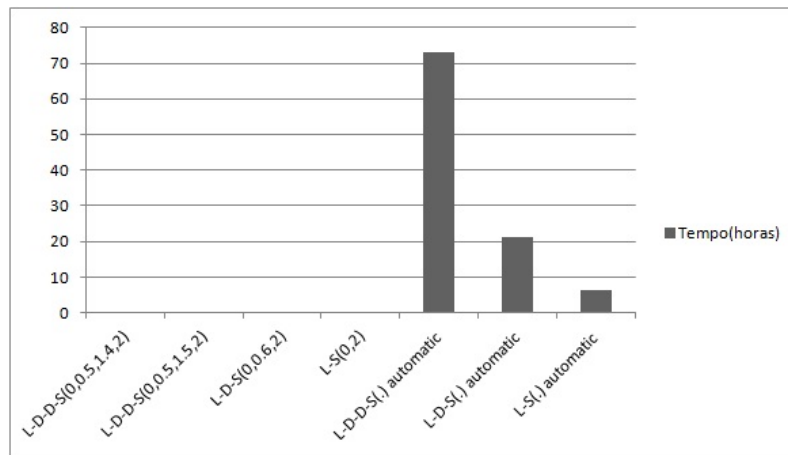


Figura 6.5: Tempo computacional gasto no ajuste banco M3 dos modelos Theta generalizados

# Capítulo 7

## Considerações Finais

A proposta deste trabalho foi desenvolver um estudo teórico e aplicado de métodos de previsão para grandes quantidades de séries temporais univariadas. O trabalho foi desenvolvido de forma empírica com as bases de dados *M1* e *M3*, que contém respectivamente 1001 e 3003 séries temporais univariadas. Buscamos atender a demanda de previsão do banco *M3* de forma a julgar a acurácia da previsão em relação as medidas MASE e sMAPE. Dentre os diversos modelos empregados com o objetivo de prever a base *M3* selecionou-se modelos da família de alisamento exponencial via espaço de estados, métodos ARIMA automáticos e o método Theta generalizado. Propomos uma redefinição do método Theta generalizado em relação a quantidade de linhas de decomposição, métodos de extrapolação e combinação das linhas-Theta com o objetivo de reduzir o erro de previsão do método.

Dentro da família de modelos de alisamento exponencial foram selecionados sete modelos e avaliou-se o poder previsivo de cada um, em que as melhores performances foram obtidas pelo modelo Damped aditivo e Damped multiplicativo. Além disso, foi empregada a rotina ETS para previsão de séries temporais baseada na seleção de até 30 modelos de alisamento exponencial. Este procedimento de modelagem não mostrou-se superior aos modelos *Damped* aditivo ou *Damped* multiplicativo em relação ao ranking geral. No entanto, quando as séries têm características sazonais o procedimento mostrou competência nas previsões sobrepujando ou igualando a performance em relação aos modelos *Damped* levando em consideração o sMAPE. Uma desvantagem no método ETS é o esforço computacional necessário para previsão, em um computador com processador Core

2 duo e 4 GBbytes de memória RAM o tempo registrado para previsão da base M3 foi de aproximadamente 27 horas, contra 30 minutos em média dos métodos Damped.

Em relação aos métodos ARIMA automáticos foram avaliados dois métodos, ambos desenvolvidos por Hyndman e Khandakar (2008). O primeiro método é baseado no algoritmo *stepwise*, o segundo baseado no ajuste de todos os possíveis modelos e posterior seleção através do critério de Akaike. A aplicação foi construída com base em toda classe de modelos  $ARIMA(p, d, q)(P, D, Q)_m$  considerando  $0 \leq d, D \leq 2$  e  $0 \leq p, q, P, Q \leq 5$ . O método ARIMA automático com *stepwise* forneceu as melhores previsões em relação ao método ARIMA automático sem *stepwise*. Além do tempo computacional do método *stepwise* ser de aproximadamente uma hora, o método sem *stepwise* leva aproximadamente 20 vezes mais tempo para completar a mesma tarefa no software R.

O método Theta proposto por Assimakopoulos e Nikolopoulos (2000), surpreendeu na competição realizada por Makridakis e Hibon (2000), ganhando o primeiro lugar entre 24 diferentes métodos de acordo com a métrica sMAPE. Como consequência, o método Theta despertou o interesse de outros pesquisadores tal como Hyndman e Billah (2003). A generalização do modelo só surgiu oito anos depois, com o trabalho de Nikolopoulos et al. (2008). No entanto segundo seus autores, questões em relação a generalização do método permanecem em aberto, tais como a quantidade de linhas-Theta no modelo, os métodos de extrapolação empregados em cada linha, a combinação das linhas e uma maneira de tratar o método Theta de um ponto de vista paramétrico, em que um método de estimação seria proposto para estimar os parâmetros do modelo.

Um dos objetivos desta pesquisa foi buscar respostas para estas questões. Primeiramente definiu-se uma notação padrão para especificar cada método Theta generalizado. Definiu-se que as vertentes a serem analisadas teriam quatro, três ou duas linhas-Theta, tendo como opções quatro métodos de extrapolação dentro da família de modelos de espaço de estados, ANN, AAN,  $AA_dN$  e o método de mínimos quadrados.

Utilizando o banco M1-Competition os resultados surpreendentemente convergiram para uma mesma família de modelos Theta. Dessa forma, os melhores resultados tanto na base M1 quanto na base de dados M3 foram dadas pelos métodos da mesma família, que foram; L-D-D-S(0,0.5,1.4,2), L-D-S(0,1.7,2) e L-S(0,2). Também foi avaliado o método L-D-D-S(0,0.5,1.5,2), que obteve resultados tão bons quanto L-D-D-S(0,0.5,1.4,2). Em particular para este modelo, as linhas-Theta são combinadas com pesos iguais.

Foi proposto uma abordagem do método Theta do ponto de vista paramétrico, para estimação das componentes Theta ou parâmetros Theta utilizando métodos numéricos segundo o algoritmo L-BFGS-B proposto por Byrd et al. (1995) e aprimorado por Zhu et al. (1997). Este algoritmo estima os parâmetros Theta minimizando a soma dos erros absolutos do modelo. Desta forma, a proposta é similar a uma regressão de mínimos quadrados. O algoritmo de estimação foi aplicado aos modelos L-S( $0, \theta_a$ ), L-D-S( $0, \theta_b, 2$ ) e L-D-D-S( $0, \theta_c, \theta_d, 2$ ) em que ( $\theta_j \in \mathfrak{R}$ )  $j = a, b, c, d$  referem-se respectivamente aos parâmetros a serem estimados no modelo sob as seguintes restrições:  $\theta_a = [0; 2.5]$ ,  $\theta_b = [0; 2]$ ,  $\theta_c = [0; 1]$ ,  $\theta_d = [0; 2]$ . A performance previsiva trouxe resultados promissores, no entanto, não sobrepujou os resultados do melhor modelo durante a competição L-D-D-S( $0, 0.5, 1.4, 2$ ). A grande desvantagem do método está relacionada ao enorme esforço computacional despendido na estimação dos coeficientes Theta. Para o caso do modelo L-D-D-S( $0, \theta_c, \theta_d, 2$ ), o tempo para estimação da base M3 foi de aproximadamente 72 horas.

Por fim, os resultados finais da competição mostraram características interessantes de cada modelo aplicado. Em relação aos modelos de alisamento exponencial foi reiterado que o parâmetro de amortização da tendência de fato reduz o erro de previsão e que estes modelos em especial podem trazer contribuições na extrapolação das linhas-Theta. De acordo com a grande diversidade do método Theta, foi proposta uma restrição do modelo até quatro linhas-Theta. Resultados empíricos mostraram que, conforme aumenta-se a quantidade de linhas para decomposição da série, o poder preditivo aumenta. Obviamente, o aumento das linhas no modelo aumenta o tempo de estimação das previsões, da mesma forma existe uma melhora em relação à acurácia que compensa essa desvantagem. Uma questão de investigação futura com certeza é a busca da otimização de ambos. Outra plataforma de programação tal como C/C++ podem reduzir o esforço computacional empregado na estimação dos modelos.

Baseados nos resultados obtidos, os próximos desafios da pesquisa são um estudo baseado na combinação de modelos a fim de avaliar a performance de previsão, aprofundar um estudo de métodos de combinação para as linhas-Theta de decomposição e trabalhar na redução do esforço computacional para estimar e prever os modelos Theta.

# Apêndice A

## Tabelas resultados alisamento exponencial

Tabela A.1: Resultado dos modelos de alisamento exponencial aplicados à base M3: 645 séries anuais.

Método	Horizontes de previsão						Média do horizonte de previsão		Ranking Médio
	1	2	3	4	5	6	1-4	1-6	
MASE									
AA <sub>d</sub> N	<b>1.10</b>	<b>1.80</b>	<b>2.80</b>	<b>3.50</b>	<b>4.00</b>	<b>4.50</b>	<b>2.30</b>	<b>2.95</b>	1.12
AAN	1.10	1.90	2.90	3.70	4.30	4.90	2.40	3.13	2.38
ANN	1.30	2.10	3.00	3.60	4.20	4.90	2.50	3.18	3.06
ETS	1.10	1.90	3.00	4.00	5.00	6.20	2.50	3.53	3.44
sMAPE									
AA <sub>d</sub> N	<b>8.20</b>	<b>12.40</b>	<b>17.20</b>	<b>19.60</b>	<b>22.20</b>	<b>23.80</b>	<b>14.35</b>	<b>17.23</b>	1.00
ANN	8.70	13.30	17.40	19.80	22.60	24.80	14.80	17.77	2.38
ETS	8.40	12.80	17.80	21.10	24.80	26.90	15.03	18.63	2.81
AAN	8.40	13.50	18.70	22.10	25.00	27.20	15.68	19.15	3.81

Tabela A.2: Resultado dos modelos de alisamento exponencial aplicados à base M3: 756 séries trimestrais.

Método	Horizontes de previsão						Média do horizonte de previsão			Ranking Médio
	1	2	3	4	6	8	1-4	1-6	1-8	
MASE										
<b>Dadt</b>	<b>0.60</b>	<b>0.70</b>	<b>0.90</b>	<b>1.10</b>	<b>1.40</b>	<b>1.80</b>	<b>0.82</b>	<b>1.00</b>	<b>1.19</b>	<b>2.18</b>
Dmult	0.60	0.80	0.90	1.10	1.40	1.80	0.85	1.02	1.19	2.59
HW mult	0.60	0.70	0.90	1.10	1.50	1.90	0.82	1.02	1.21	2.82
ETS	0.60	0.80	0.90	1.10	1.50	1.90	0.85	1.03	1.23	3.50
HW adt	0.60	0.70	0.90	1.10	1.60	2.00	0.82	1.05	1.26	3.91
AA <sub>d</sub> N	1.00	1.00	1.20	1.40	1.80	2.20	1.15	1.35	1.54	6.36
ANN	1.00	1.00	1.30	1.50	1.80	2.30	1.20	1.37	1.57	7.14
AAN	1.00	1.00	1.30	1.40	1.90	2.30	1.18	1.42	1.64	7.50
sMAPE										
<b>ETS</b>	5.00	6.90	7.80	<b>9.40</b>	12.10	14.10	<b>7.28</b>	<b>8.60</b>	<b>9.85</b>	1.77
Dmult	5.00	6.90	<b>7.70</b>	9.70	12.20	14.20	7.33	8.70	9.92	2.32
Dadt	5.10	6.80	8.10	9.70	<b>12.00</b>	<b>14.00</b>	7.42	8.73	9.95	2.82
HW mult	<b>4.80</b>	<b>6.70</b>	7.90	10.10	13.20	15.80	7.38	9.02	10.56	3.82
ANN	7.20	8.10	9.30	10.50	12.50	14.40	8.78	9.82	10.88	5.32
HW adt	5.00	7.00	8.60	10.20	13.60	16.20	7.70	9.35	10.95	5.55
AA <sub>d</sub> N	7.30	8.00	9.50	10.60	13.30	14.70	8.85	10.23	11.30	6.50
AAN	7.30	8.10	10.00	11.00	14.70	16.70	9.10	10.77	12.21	7.91

Tabela A.3: Resultado dos modelos de alisamento exponencial aplicados à base M3: 1428 séries mensais

Método	Horizontes de previsão						Média do horizonte de previsão					Ranking Médio
	1	4	6	8	12	18	1-4	1-6	1-8	1-12	1-18	
MASE												
<b>Dadt</b>	<b>1.40</b>	<b>1.60</b>	<b>1.50</b>	<b>1.70</b>	<b>1.70</b>	<b>2.30</b>	<b>1.48</b>	<b>1.50</b>	<b>1.55</b>	<b>1.64</b>	<b>1.82</b>	<b>1.52</b>
ETS	1.50	<b>1.60</b>	1.60	<b>1.70</b>	<b>1.70</b>	2.40	1.52	1.55	1.57	1.67	1.84	2.11
Dmult	1.50	1.70	1.60	1.80	1.80	2.70	1.52	1.53	1.59	1.68	1.91	2.74
HW adt	1.50	1.70	1.60	1.80	2.10	2.90	1.57	1.60	1.66	1.80	2.11	4.28
HW mult	1.50	1.70	1.70	1.90	2.10	3.10	1.57	1.60	1.65	1.81	2.13	4.50
ANN	1.90	2.30	2.10	2.20	2.20	3.00	2.02	2.07	2.11	2.14	2.39	6.00
AA <sub>d</sub> N	1.80	2.40	2.20	2.30	2.20	3.10	2.05	2.12	2.17	2.23	2.49	6.91
AAN	1.90	2.50	2.30	2.60	2.90	4.20	2.12	2.20	2.31	2.49	2.97	7.93
sMAPE												
<b>Dmult</b>	<b>11.20</b>	<b>13.10</b>	12.50	13.10	<b>13.60</b>	18.90	<b>11.80</b>	<b>12.00</b>	<b>12.29</b>	<b>12.76</b>	<b>14.20</b>	<b>1.39</b>
ETS	11.50	13.20	<b>12.40</b>	13.20	13.80	19.30	11.93	<b>12.00</b>	12.36	12.85	14.38	2.22
Dadt	11.30	13.30	12.80	<b>13.00</b>	13.80	<b>18.50</b>	12.07	12.63	12.97	13.26	14.63	2.72
HW mult	11.70	13.70	13.00	14.20	16.10	22.40	12.32	12.48	12.90	13.82	15.91	4.85
ANN	13.20	16.80	14.10	14.80	14.30	20.60	14.43	14.50	14.70	14.60	16.24	5.22
HW adt	11.50	13.70	13.30	13.70	15.80	21.80	12.45	13.15	13.65	14.20	16.27	5.48
AA <sub>d</sub> N	12.80	16.90	14.50	15.20	14.90	20.90	14.30	14.55	14.82	14.92	16.59	6.22
AAN	12.80	17.10	15.40	16.40	17.60	24.40	14.50	14.95	15.45	16.12	18.39	7.91

Tabela A.4: Resultado dos modelos de alisamento exponencial aplicados à base M3: 174 séries outras

Método	Horizontes de previsão						Média do horizonte de previsão			Ranking Médio
	1	2	3	4	6	8	1-4	1-6	1-8	
MASE										
<b>ETS</b>	<b>0.70</b>	<b>1.00</b>	<b>1.60</b>	<b>1.80</b>	<b>2.20</b>	2.80	<b>1.27</b>	<b>1.55</b>	<b>1.81</b>	<b>1.36</b>
AA <sub>d</sub> N	0.80	<b>1.00</b>	1.70	1.90	<b>2.20</b>	<b>2.70</b>	1.35	1.62	1.84	2.14
AAN	0.80	<b>1.00</b>	<b>1.60</b>	1.90	2.30	3.10	1.32	1.62	1.92	2.50
ANN	1.00	1.50	2.70	3.10	4.00	5.00	2.08	2.65	3.15	4.00
sMAPE										
<b>ETS</b>	<b>1.80</b>	<b>2.70</b>	<b>3.70</b>	<b>4.30</b>	<b>4.80</b>	7.20	<b>3.12</b>	<b>3.77</b>	4.41	<b>1.45</b>
AA <sub>d</sub> N	<b>1.80</b>	<b>2.70</b>	3.80	4.40	<b>4.80</b>	<b>6.90</b>	3.18	3.82	<b>4.39</b>	1.68
AAN	1.90	2.80	<b>3.70</b>	4.60	5.30	7.80	3.25	4.00	4.72	2.86
ANN	2.10	3.60	5.40	6.30	7.60	10.30	4.35	5.47	6.42	4.00

# Apêndice B

## Tabelas resultados métodos ARIMA Automáticos

Tabela B.1: Resultado dos procedimentos Auto ARIMA aplicados à base M3: 645 séries anuais

Método	Horizontes de previsão						Média do horizonte de previsão		Ranking Médio
	1	2	3	4	5	6	1-4	1-6	
	MASE								
Auto ARIMA ( <i>stepwise</i> )	<b>1.10</b>	<b>1.90</b>	<b>2.70</b>	<b>3.40</b>	<b>4.10</b>	<b>4.60</b>	<b>2.27</b>	<b>2.97</b>	<b>1.06</b>
Auto ARIMA	<b>1.10</b>	<b>1.90</b>	2.80	3.60	4.20	4.80	2.35	3.07	1.94
	sMAPE								
Auto ARIMA ( <i>stepwise</i> )	<b>8.40</b>	<b>12.40</b>	<b>16.70</b>	<b>19.40</b>	<b>22.40</b>	<b>24.10</b>	<b>14.22</b>	<b>17.23</b>	<b>1.12</b>
Auto ARIMA	<b>8.40</b>	13.20	17.50	20.60	23.30	24.80	14.93	17.97	1.88

Tabela B.2: Resultado dos procedimentos Auto ARIMA aplicados à base M3: 754 séries trimestrais

Método	Horizontes de previsão						Média do horizonte de previsão			Ranking Médio
	1	2	3	4	6	8	1-4	1-6	1-8	
	MASE									
Auto ARIMA	<b>0.60</b>	0.80	<b>0.90</b>	<b>1.10</b>	<b>1.40</b>	<b>1.70</b>	0.85	<b>1.00</b>	<b>1.16</b>	<b>1.45</b>
Auto ARIMA ( <i>stepwise</i> )	<b>0.60</b>	<b>0.70</b>	<b>0.90</b>	<b>1.10</b>	<b>1.40</b>	1.80	<b>0.82</b>	<b>1.00</b>	1.18	1.55
	sMAPE									
Auto ARIMA	<b>5.20</b>	<b>7.10</b>	8.30	10.00	<b>11.90</b>	<b>13.60</b>	7.65	<b>8.87</b>	<b>9.94</b>	1.32
Auto ARIMA ( <i>stepwise</i> )	5.30	<b>7.10</b>	<b>8.20</b>	<b>9.90</b>	12.10	13.90	<b>7.62</b>	8.92	10.03	1.68

Tabela B.3: Resultado dos procedimentos Auto ARIMA aplicados à base M3: 1428 séries mensais

Método	Horizontes de previsão						Média do horizonte de previsão					Ranking Médio
	1	4	6	8	12	18	1-4	1-6	1-8	1-12	1-18	
	MASE											
Auto ARIMA	<b>1.50</b>	<b>1.70</b>	<b>1.60</b>	<b>1.70</b>	<b>1.80</b>	<b>2.40</b>	<b>1.57</b>	<b>1.58</b>	<b>1.61</b>	<b>1.70</b>	<b>1.89</b>	<b>1.35</b>
Auto ARIMA ( <i>stepwise</i> )	<b>1.50</b>	<b>1.70</b>	<b>1.60</b>	<b>1.70</b>	<b>1.80</b>	2.50	<b>1.57</b>	1.60	1.64	<b>1.70</b>	1.92	1.65
	sMAPE											
<b>Auto ARIMA</b>	12.10	<b>13.50</b>	<b>12.80</b>	<b>13.60</b>	<b>14.60</b>	19.90	12.38	<b>12.50</b>	<b>12.80</b>	<b>13.38</b>	<b>15.03</b>	<b>1.35</b>
Auto ARIMA ( <i>stepwise</i> )	<b>11.90</b>	13.80	12.90	<b>13.60</b>	14.70	<b>19.60</b>	<b>12.50</b>	12.60	12.89	13.39	15.04	1.65

Tabela B.4: Resultado dos procedimentos Auto ARIMA aplicados à base M3: 174 séries outras

Método	Horizontes de previsão						Média do horizonte de previsão			Ranking Médio
	1	2	3	4	6	8	1-4	1-6	1-8	
	MASE									
Auto ARIMA ( <i>stepwise</i> )	<b>0.60</b>	<b>0.80</b>	<b>0.90</b>	<b>1.10</b>	<b>1.40</b>	<b>1.70</b>	<b>0.85</b>	<b>1.00</b>	<b>1.16</b>	<b>1.00</b>
Auto ARIMA	0.70	1.00	1.70	1.80	2.20	3.00	1.30	1.58	1.88	2.00
	sMAPE									
Auto ARIMA	1.70	2.70	<b>3.80</b>	<b>4.20</b>	<b>5.10</b>	<b>7.70</b>	<b>3.10</b>	<b>3.80</b>	<b>4.58</b>	<b>1.32</b>
Auto ARIMA ( <i>stepwise</i> )	<b>1.60</b>	<b>2.60</b>	<b>3.80</b>	4.40	5.30	7.90	3.10	3.87	4.65	1.68



# Apêndice C

## Tabelas resultados método Theta generalizado

Tabela C.1: Resultados do método Theta generalizado aplicado à base M3: 645 séries anuais

Método	Horizontes de previsão						Média do horizonte de previsão		Ranking Médio
	1	2	3	4	5	6	1-4	1-6	
MASE									
<b>L-D-D-S(0,0.5,1.4,2)</b>	<b>1.10</b>	<b>1.70</b>	<b>2.60</b>	<b>3.10</b>	<b>3.60</b>	<b>4.10</b>	<b>2.12</b>	<b>2.70</b>	<b>1.94</b>
L-D-D-S(0,0.5,1.5,2)	<b>1.10</b>	<b>1.70</b>	<b>2.60</b>	<b>3.10</b>	3.70	<b>4.10</b>	<b>2.12</b>	2.72	2.38
L-D-S(0,1.7,2)	<b>1.10</b>	<b>1.70</b>	<b>2.60</b>	<b>3.10</b>	3.70	<b>4.10</b>	<b>2.12</b>	2.72	2.38
L-S(0,2)	<b>1.10</b>	1.80	<b>2.60</b>	<b>3.10</b>	3.70	4.20	2.15	2.75	3.31
<b>L-D-D-S(0,<math>\theta_c, \theta_d, 2</math>)</b>	<b>1.10</b>	<b>1.80</b>	<b>2.60</b>	<b>3.10</b>	<b>3.70</b>	<b>4.10</b>	<b>2.15</b>	<b>2.73</b>	<b>1.12</b>
L-D-S(0, $\theta_b, 2$ )	<b>1.10</b>	<b>1.80</b>	2.70	3.30	3.90	4.40	2.23	2.87	2.19
L-S(0, $\theta_a$ )	1.60	2.00	2.80	3.30	3.80	4.30	2.42	2.97	2.69
sMAPE									
<b>L-D-D-S(0,0.5,1.4,2)</b>	<b>8.00</b>	<b>11.80</b>	<b>16.20</b>	<b>18.40</b>	21.10	<b>22.40</b>	<b>13.60</b>	<b>16.32</b>	<b>1.19</b>
L-D-D-S(0,0.5,1.5,2)	<b>8.00</b>	11.90	16.30	18.60	<b>21.00</b>	22.50	13.70	16.38	1.88
L-D-S(0,1.7,2)	8.10	11.90	16.40	18.80	21.20	22.90	13.80	16.55	3.13
L-S(0,2)	8.20	12.30	16.50	18.70	21.20	23.00	13.93	16.65	3.81
<b>L-D-D-S(0,<math>\theta_c, \theta_d, 2</math>)</b>	<b>8.20</b>	<b>12.00</b>	<b>16.50</b>	<b>18.80</b>	<b>21.30</b>	<b>22.90</b>	<b>13.88</b>	<b>16.62</b>	<b>1.19</b>
L-D-S(0, $\theta_b, 2$ )	<b>8.20</b>	12.10	16.60	<b>18.80</b>	21.40	<b>22.90</b>	13.93	16.67	1.81
L-S(0, $\theta_a$ )	10.70	13.60	17.50	19.60	21.90	24.10	15.35	17.90	3.00

Tabela C.2: Resultados do método Theta generalizado aplicado à base M3: 756 séries trimestrais

Método	Horizontes de previsão						Média do horizonte de previsão			Ranking Médio
	1	2	3	4	6	8	1-4	1-6	1-8	
MASE										
<b>L-D-S(0,1,7,2)</b>	<b>0.60</b>	<b>0.80</b>	<b>0.90</b>	<b>1.10</b>	<b>1.40</b>	<b>1.80</b>	<b>0.85</b>	<b>1.02</b>	<b>1.19</b>	<b>1.86</b>
L-D-D-S(0,0.5,1.5,2)	<b>0.60</b>	<b>0.80</b>	<b>0.90</b>	<b>1.10</b>	1.50	<b>1.80</b>	<b>0.85</b>	1.03	1.20	2.27
L-D-D-S(0,0.5,1.4,2)	<b>0.60</b>	<b>0.80</b>	<b>0.90</b>	<b>1.10</b>	1.50	<b>1.80</b>	<b>0.85</b>	1.03	1.21	2.55
L-S(0,2)	<b>0.60</b>	<b>0.80</b>	1.00	1.20	1.50	<b>1.80</b>	0.90	1.07	1.24	3.32
<b>L-D-D-S(0,θ<sub>c</sub>,θ<sub>d</sub>,2)</b>	<b>0.60</b>	<b>0.80</b>	<b>0.90</b>	<b>1.10</b>	<b>1.40</b>	<b>1.80</b>	<b>0.85</b>	<b>1.02</b>	<b>1.19</b>	<b>1.09</b>
L-D-S(0,θ <sub>b</sub> ,2)	<b>0.60</b>	<b>0.80</b>	1.00	1.20	1.60	1.90	0.90	1.10	1.27	2.00
L-S(0,θ <sub>a</sub> )	0.80	1.00	1.10	1.30	1.60	1.90	1.05	1.22	1.38	2.91
sMAPE										
<b>L-D-S(0,1,7,2)</b>	<b>5.00</b>	6.70	<b>7.30</b>	<b>8.80</b>	<b>10.60</b>	<b>12.30</b>	6.95	<b>7.95</b>	<b>8.93</b>	<b>1.59</b>
L-D-D-S(0,0.5,1.4,2)	<b>5.00</b>	6.60	<b>7.30</b>	<b>8.80</b>	10.70	12.40	<b>6.92</b>	7.97	8.96	2.00
L-D-D-S(0,0.5,1.5,2)	<b>5.00</b>	<b>6.50</b>	<b>7.30</b>	8.90	10.80	12.60	<b>6.92</b>	8.00	9.03	2.68
L-S(0,2)	5.20	6.90	7.60	9.00	11.00	12.40	7.17	8.23	9.16	3.73
<b>L-D-D-S(0,θ<sub>c</sub>,θ<sub>d</sub>,2)</b>	5.10	<b>6.70</b>	<b>7.30</b>	<b>8.70</b>	<b>10.60</b>	<b>12.10</b>	<b>6.95</b>	<b>7.93</b>	<b>8.86</b>	<b>1.09</b>
L-D-S(0,θ <sub>b</sub> ,2)	<b>5.00</b>	6.80	7.70	9.50	11.50	13.50	7.25	8.47	9.57	2.14
L-S(0,θ <sub>a</sub> )	6.20	7.70	8.40	9.60	11.60	12.70	7.97	8.97	9.81	2.77

Tabela C.3: Resultados do método Theta generalizado aplicado à base M3: 1428 séries mensais

Método	Horizontes de previsão						Média do horizonte de previsão					Ranking Médio
	1	4	6	8	12	18	1-4	1-6	1-8	1-12	1-18	
MASE												
<b>L-D-S(0,1,7,2)</b>	<b>0.70</b>	<b>1.00</b>	<b>1.10</b>	<b>1.20</b>	<b>1.40</b>	<b>2.00</b>	<b>0.80</b>	<b>0.88</b>	<b>0.96</b>	<b>1.08</b>	<b>1.31</b>	<b>1.83</b>
L-D-D-S(0,0.5,1.5,2)	<b>0.70</b>	<b>1.00</b>	<b>1.10</b>	1.30	1.40	2.10	<b>0.80</b>	<b>0.88</b>	0.98	1.11	1.34	2.48
L-D-D-S(0,0.5,1.4,2)	<b>0.70</b>	<b>1.00</b>	<b>1.10</b>	1.30	1.40	2.10	<b>0.80</b>	0.90	0.99	1.12	1.35	2.85
L-S(0,2)	<b>0.70</b>	<b>1.00</b>	<b>1.10</b>	1.30	1.40	2.10	<b>0.80</b>	0.90	0.99	1.12	1.35	2.85
<b>L-D-D-S(0,θ<sub>c</sub>,θ<sub>d</sub>,2)</b>	<b>0.70</b>	<b>1.00</b>	<b>1.10</b>	<b>1.20</b>	<b>1.40</b>	<b>2.00</b>	<b>0.80</b>	<b>0.88</b>	<b>0.96</b>	<b>1.09</b>	<b>1.32</b>	<b>1.15</b>
L-D-S(0,θ <sub>b</sub> ,2)	<b>0.70</b>	<b>1.00</b>	1.20	1.30	1.50	2.10	<b>0.80</b>	0.92	1.01	1.15	1.39	2.04
L-S(0,θ <sub>a</sub> ,2)	0.80	1.10	1.20	1.40	1.50	2.10	0.90	0.98	1.07	1.21	1.44	2.80
sMAPE												
<b>L-D-D-S(0,0.5,1.4,2)</b>	11.20	12.50	<b>12.10</b>	<b>12.70</b>	<b>13.00</b>	<b>17.80</b>	<b>11.50</b>	<b>11.68</b>	<b>12.01</b>	<b>12.38</b>	<b>13.67</b>	<b>1.59</b>
L-D-D-S(0,0.5,1.5,2)	11.20	12.50	<b>12.10</b>	<b>12.70</b>	<b>13.00</b>	17.90	<b>11.50</b>	<b>11.68</b>	<b>12.01</b>	12.39	13.69	1.89
L-D-S(0,1,7,2)	<b>11.10</b>	<b>12.40</b>	12.30	12.80	13.20	18.00	<b>11.50</b>	11.78	12.12	12.52	13.83	3.07
L-S(0,2)	11.20	12.50	12.20	12.80	13.20	18.10	11.55	11.77	12.12	12.52	13.86	3.46
<b>L-D-D-S(0,θ<sub>c</sub>,θ<sub>d</sub>,2)</b>	<b>11.10</b>	<b>12.40</b>	<b>12.10</b>	<b>12.60</b>	<b>13.10</b>	<b>17.80</b>	<b>11.47</b>	<b>11.70</b>	<b>12.03</b>	<b>12.41</b>	<b>13.69</b>	<b>1.04</b>
L-D-S(0,θ <sub>b</sub> ,2)	11.40	12.70	12.30	13.00	13.30	18.20	11.65	11.87	12.21	12.62	13.95	1.96
L-S(0,θ <sub>a</sub> )	12.30	13.50	13.50	14.20	14.50	19.30	12.55	12.83	13.28	13.75	15.09	3.00

Tabela C.4: Resultados do método Theta generalizado aplicado à base M3: 174 séries  
outras

Método	Horizontes de previsão						Média do horizonte de previsão			Ranking Médio
	1	2	3	4	6	8	1-4	1-6	1-8	
MASE										
<b>L-D-D-S(0,0.5,1.5,2)</b>	<b>0.80</b>	<b>1.10</b>	<b>1.90</b>	<b>2.00</b>	<b>2.40</b>	<b>3.00</b>	<b>1.45</b>	<b>1.75</b>	<b>2.00</b>	<b>1.36</b>
L-D-D-S(0,0.5,1.4,2)	<b>0.80</b>	<b>1.10</b>	<b>1.90</b>	2.10	2.50	3.10	1.48	1.78	2.05	2.18
L-D-S(0,1.7,2)	1.00	1.20	<b>1.90</b>	2.10	2.50	<b>3.00</b>	1.55	1.83	2.08	2.68
L-S(0,2)	<b>0.80</b>	1.20	2.10	2.30	2.90	3.50	1.60	2.00	2.31	3.77
<b>L-D-S(0,θ<sub>b</sub>,2)</b>	<b>0.80</b>	<b>1.10</b>	<b>1.80</b>	<b>1.90</b>	<b>2.30</b>	<b>2.90</b>	<b>1.40</b>	<b>1.68</b>	<b>1.92</b>	<b>1.09</b>
L-D-D-S(0,θ <sub>c</sub> ,θ <sub>d</sub> ,2)	1.00	1.20	2.00	2.20	2.60	3.20	1.60	1.90	2.16	2.41
L-S(0,θ <sub>a</sub> )	<b>0.80</b>	<b>1.10</b>	2.00	2.20	2.70	3.40	1.53	1.90	2.21	2.50
sMAPE										
<b>L-D-D-S(0,0.5,1.5,2)</b>	<b>1.80</b>	<b>2.80</b>	<b>4.10</b>	<b>4.60</b>	<b>5.10</b>	<b>7.20</b>	<b>3.32</b>	<b>4.02</b>	<b>4.61</b>	<b>1.23</b>
L-D-D-S(0,0.5,1.4,2)	<b>1.80</b>	<b>2.80</b>	<b>4.10</b>	4.70	5.20	7.30	3.35	4.07	4.66	1.91
L-D-S(0,1.7,2)	2.10	3.00	4.20	4.80	5.30	7.50	3.52	4.18	4.81	3.00
L-S(0,2)	1.90	3.00	4.40	5.10	5.70	7.90	3.60	4.38	5.05	3.86
<b>L-D-S(0,θ<sub>b</sub>,2)</b>	1.90	<b>2.80</b>	<b>3.90</b>	<b>4.50</b>	<b>5.10</b>	<b>7.20</b>	<b>3.27</b>	<b>3.97</b>	<b>4.56</b>	<b>1.14</b>
L-S(0,θ <sub>a</sub> )	<b>1.80</b>	<b>2.80</b>	4.20	4.90	5.50	7.70	3.43	4.20	4.86	2.27
L-D-D-S(0,θ <sub>c</sub> ,θ <sub>d</sub> ,2)	2.10	3.00	4.30	4.80	5.40	7.60	3.55	4.25	4.89	2.59

# Apêndice D

## Provas

**Prova D.1** *Prova Capítulo 2 Seção 2.4.2. Queremos mostrar que*

$$L^*(\theta, x_0) = n \log \left( \sum_{t=1}^n \varepsilon_t^2 \right) + 2 \sum_{t=1}^n \log |r(x_{t-1})|$$

*é igual a duas vezes o negativo do logaritmo da função de verossimilhança. A densidade conjunta da série  $y_t | \tau, x_0, \sigma^2$  é dada pelo produto ponderado das densidades individuais, isto é:*

$$P(y_t | \tau, x_0, \sigma^2) = \prod_{t=1}^n P(\varepsilon_t) / |r(x_{t-1})|. \quad (\text{D.2})$$

*como  $\varepsilon_t$  é um processo gaussiano, então a verossimilhança é dada por:*

$$L = (2\pi\sigma^2)^{-n/2} \left| \prod_{t=1}^n r(x_{t-1}) \right|^{-1} \exp \left( \frac{-1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right) \quad (\text{D.3})$$

$$\log L = \frac{-n}{2} \log(2\pi\sigma^2) - \sum_{t=1}^n \log |r(x_{t-1})| \left( \frac{-1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right) \quad (\text{D.4})$$

*substituindo  $\sigma^2$  por  $\hat{\sigma}^2$  temos a verossimilhança reduzida, dada por:*

$$L(\tau, x_0 | y) = (2\pi\hat{\sigma}^2)^{-n/2} \left| \prod_{t=1}^n r(x_{t-1}) \right|^{-1}$$

a partir de (D.5) calculamos o negativo de duas vezes o log da verossimilhança, dado por:

$$\begin{aligned} -2 \log L(\tau, x_0|y_t) &= n \log(2\pi\hat{\sigma}^2) + 2 \sum_{t=1}^n \log |r(x_{t-1})| \\ &= c_n + n \log\left(\sum_{t=1}^n \varepsilon_t^2\right) + 2 \sum_{t=1}^n \log |r(x_{t-1})| \end{aligned}$$

em que  $c_n$  é uma constante que depende apenas de  $n$ . Desta maneira, a estimativa da verossimilhança pode ser obtida minimizando-se

$$L^*(\tau, x_0) = n \log\left(\sum_{t=1}^n \varepsilon_t^2\right) + 2 \sum_{t=1}^n \log |r(x_{t-1})| \quad (\text{D.5})$$

como queremos demonstrar.

# Apêndice E

## Códigos Utilizados Para Programação no Software R

### Teste de Sazonalidade para base M3

```
TesteSazonal <- rep(FALSE,3003)
for(i in 1:3003)
{
  if(M3[[i]]$period == "MONTHLY")
  {
    x <- c(M3[[i]]$x,M3[[i]]$xx)
    xacf <- acf(x,plot=FALSE)$acf[-1,1,1]
    clim <- 1.645/sqrt(length(x))*sqrt(cumsum(c(1,2*xacf^2)))
    seas[i] <- abs(xacf[12]) > clim[12]
  }
  if(M3[[i]]$period == "QUARTERLY")
  {
    x <- c(M3[[i]]$x,M3[[i]]$xx)
    xacf <- acf(x,plot=FALSE)$acf[-1,1,1]
    clim <- 1.645/sqrt(length(x))*sqrt(cumsum(c(1,2*xacf^2)))
    seas[i] <- abs(xacf[4]) > clim[4]
  }
}
```

**Medidas de Acurácia ASE e sAPE**

```
ASE <- function (previsao, serieteste, serietreinamento, m = frequency(serie))
{
  a <- (abs(serieteste - previsao))
  b <- mean(abs(diff(serietreinamento, lag = m)))
  (abs(a/b))
}
```

```
sAPE <- function (previsao, serieteste)
{
  a = abs(serieteste - previsao)
  b = (serieteste + previsao)
  (a/b) * 200
}
```

# Referências Bibliográficas

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.
- Akaike, H. (1977). *On entropy maximization principle*. Applications of Statistics.
- Anderson, B. D. O. e Moore, J. B. (1979). *Optimal filtering*. Prentice-Hall.
- Aoki, M. (1987). *State space modeling of time series*. Berlin Heidelberg.
- Aoki, M. e Havennerb, A. (1991). State space modelling of multiple time series. *Econometric Reviews*, 10:1–59.
- Archibald, B. C. (1990). Parameter space of the holt-winters' model. *International Journal of Forecasting*, 6:199–209.
- Armstrong, J. S. (1984). Forecasting by extrapolation: Conclusions from twenty five years of research. *Interfaces*, 14(6):52–66.
- Armstrong, J. S. e Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8:69–80.
- Assimakopoulos, V. e Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16:521–530.
- Athanasopoulos, G., Hyndman, R. J., Song, H., e Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27:822–844.



- Box, G. e Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Box, G. E. P., Jenkins, G. M., e Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. Englewood Cliffs, NJ: Prentice Hall.
- Bozdongan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 3:345–370.
- Brockwell, P. J. e Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag.
- Brown, R. G. (1959). *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill.
- Byrd, R. H., Lu, P., Nocedal, J., e Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.
- Canova, F. e Hasen, B. E. (1995). Are seasonal patterns constant over time? a test for seasonal stability. *Journal of Business and Economic Statistics*, 13:237–252.
- Chatfield, C. (1988). Apples, oranges and mean square error. *International Journal of Forecasting*, 4:515–518.
- Constantinidou, C., Nikolopoulos, K., Bougioukos, N., Tsiafa, E., Petropoulos, F., e Assimakopoulos, V. (2012). A neural network approach for the theta model. *Information Engineering, Lecture Notes in Information Technology*, 25:116–120.
- Crone, S. F., Hibon, M., e Nikolopoulos, K. (2011). Advances in forecasting with neural networks: Empirical evidence from the nn3 competition on time series prediction. *International Journal of Forecasting*, 27:635–660.
- Dickey, D. A. e Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057–1071.
- Fildes, R., Hibon, M., Makridakis, S., e Meade, N. (1998). Generalising about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting*, 14:339–358.

- Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society, Series A*, 125:54–59.
- Gardner, E. S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, 4:1–28.
- Gardner, E. S. (2005). Exponential smoothing: The state of the art part ii. Technical report, Bauer College of Business University of Houston.
- Gomez, V. (1998). Automatic model identification in the presence of missing observations and outliers. Technical report, Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria.
- Gomez, V. e Maravall, A. (1998). Programs tramo and seats, instructions for the users. Technical report, Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria.
- Goodrich, R. L. (2000). The forecast pro methodology. *International Journal of Forecasting*, 16(4):533–535.
- Goodwin, P. e Lawton, R. (1999). On the asymmetry of the symmetric mape. *International Journal of Forecasting*, 15:405–408.
- Halim, S. e Bisono, I. N. (2008). Automatic seasonal auto regressive moving average models and unit root test detection. *International Journal of Management Science and Engineering Management*, 3(4):266–274.
- Hannan, E. J. e Deistler, M. (1988). *The statistical theory of linear systems*. Wiley.
- Hannan, E. J. e Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69(1):81–94.
- Holt, C. C. (2004(1957)). Forecasting trends and seasonal by exponentially weighted averages. *Office of Naval Research Memorandum 52/reprinted by International Journal of Forecasting*, 20 (1):5–10.
- Hylleberg, S., Engle, R., Granger, C., e Yoo, B. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44:215–238.

- Hyndman, R. J. e Athanasopoulos, G. (2012). *Forecasting: principles and practice*. <http://otexts.com/fpp/>.
- Hyndman, R. J. e Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, 19:287–290.
- Hyndman, R. J. e Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of Statistical Software*, 27(3):1–22.
- Hyndman, R. J. e Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688.
- Hyndman, R. J., Koehler, A. B., Ord, K., e Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin, Heidelberg.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., e Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18:439–454.
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press.
- Kadane, J. B. e Lazar, N. A. (2004). Methods and criteria for model selection. *Journal of the American Statistical Association*, 99(465):279–290.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problem. *Journal of Basic Engineering*, 82(1):35–45.
- Kalman, R. E. e Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(3):95–108.
- Kelley, W. G. (2000). *Difference Equations: An Introduction with Applications*. London: Academic Press.
- Koehler, A. B., Snyder, R. D., e Ord, K. (2001). Forecasting models and prediction intervals for the multiplicative holt-winters' method. *International Journal of Forecasting*, 17:268–286.
- Kwiatkowski, D., Schmidt, P. C. P. P., e Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

- Leonard, M. (2005). Large-scale automatic forecasting: Millions of forecasts. Technical report, SAS Institute Inc. Cary, NC, USA.
- Littell, R. C., Milliken, G. A., Stroup, W., e Wolfinger, R. D. (2002). Sas system for mixed models. Technical report, Statistical Analysis System Institute.
- Liu, L. M. (1989). Identification of seasonal arima models using a filtering method. *Communications in Statistics: Theory & Methods*, 18:2279–2288.
- Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting*, 9:527–529.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., e Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1:111–153.
- Makridakis, S. e Hibon, M. (2000). The m3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476.
- Melard, G. e Pasteels, J. M. (2000). Automatic arima modeling including intervention, using time series expert software. *International Journal of Forecasting*, 16:497–508.
- Morettin, P. A. e Toloí, C. M. C. (2006). *Análise de Séries Temporais*. ABE - Projeto Fisher.
- Morry, M. (1975). A test to determine whether the seasonality is additively or multiplicatively related to the trend-cycle component. time series and seasonal adjustment staff research papers. Technical report, Statistics Canada.
- Nikolopoulos, K. e Assimakopoulos, V. (2005). Fathoming the theta model. Technical report, Forecasting Systems Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece.
- Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., e Petropoulos, F. (2012a). The theta model: An essential forecasting tool for supply chain planning. *Advances in Automation and Robotics, Lecture Notes in Electrical Engineering*, 2,123:431–437.

- Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., e Petropoulos, F. (2008). Advances in the theta model. Technical report, University of Peloponnese, Department of Economics.
- Nikolopoulos, K., Maris, K., Pantou, G., Metaxiotis, K., e Assimakopoulos, V. (2003). Forecasting volatility with the theta model. *Empirical Economics Letters*, 6:216–227.
- Nikolopoulos, K., Thomakos, D., Litsa, A., e Assimakopoulos, V. (2012b). Forecasting s&p 500 with the theta model. *International Journal of Financial Economics and Econometrics*, Accepted.
- Ord, K., Koehler, A. B., e Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, 92:1621–1629.
- Ord, K. e Lowe, S. (1996). Automatic forecasting. *The American Statistician*, 50(1):88–94.
- Pegels, C. C. (1969). Exponential forecasting: some new variations. *Management Science*, 15(5):311–315.
- Reilly, D. (2000). The autobox system. *International Journal of Forecasting*, 16(4):531–533.
- Steffen, D. (2010). Intervalo bootstrap para previsões de séries temporais obtidas pelo métodos theta. Master's thesis, Universidade Federal do Paraná.
- Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19:715–725.
- Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.
- Wolfinger, R. D. (1993). Covariance estrutura selection in general mixed models. *Communications in Statistics*, 22:1079–1106.
- Zhu, C., Byrd, R. H., Lu, P., e Nocedal, J. (1997). L-bfgs-b: Algorithm 778: L-bfgs-b, fortran routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software*, 23 (4):550–560.