



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

# **Previsão de séries temporais: extensão sazonal e bagging para o Modelo Theta Otimizado**

por

Igor de Oliveira Barros Faluhelyi

Brasília, 21 de julho de 2025

# Previsão de séries temporais: extensão sazonal e bagging para o Modelo Theta Otimizado

por

Igor de Oliveira Barros Faluhelyi

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. José Augusto Fiorucci

Brasília, 21 de julho de 2025

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Prof. José Augusto Fiorucci  
Orientador, EST/UnB

Prof. Francisco Louzada Neto  
Banca, ICMC/USP

Prof. Helton Saulo Bezerra dos Santos  
Banca, EST/UnB

Prof. Eduardo Nakano  
Suplente, EST/UnB

“If you find that you’re spending almost all your time on theory,  
start turning some attention to practical things;  
it will improve your theories.  
If you find that you’re spending almost all your time on practice,  
start turning some attention to theoretical things;  
it will improve your practice.”.  
(Donald Knuth)

*Dedico este trabalho aos meus pais, Ligia Maria de Oliveira Barros e Peter Faluhelyi,  
e a toda a minha família — em especial aos que oraram, torceram  
e até sofreram por mim ao longo desta jornada.  
Aos meus amigos, com destaque para Guilherme (Discord), pelo apoio  
nos momentos mais críticos, e Fernando Ribeiro, cujos conselhos  
e exemplo são, para mim, fonte constante de inspiração e admiração.*

# Agradecimentos

Meus sinceros agradecimentos aos professores do PPGEST/UnB, em especial ao Prof. Eduardo Yoshio Nakano, ao Prof. Guilherme Souza Rodrigues e ao Prof. José Augusto Fiorucci, por sua orientação, apoio e ensinamentos ao longo do mestrado. Agradeço também aos prezados Professores Raul Yukihiro Matsushita e Helton Saulo Bezerra dos Santos, cujas interações foram valiosas para minha formação. Em especial, agradeço à colega Beatriz Albernaz, pelo companheirismo e incentivo ao longo dessa caminhada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

# Resumo

Este trabalho teve como objetivo principal propor uma extensão sazonal para o Modelo Theta Otimizado (OTM), com o intuito de modelar adequadamente séries temporais com sazonalidade, especialmente aquelas com sazonalidade dinâmica, além de avaliar o uso da técnica de Bagging aplicada para Modelo Theta Otimizado. Primeiramente, apresentamos uma revisão bibliográfica sobre Modelos de Espaço de Estado com inovações, métodos de alisamento exponencial e o Método Theta, destacando os modelos para sua otimização. Em seguida, desenvolvemos formalmente as versões aditiva e multiplicativa do Modelo Theta Otimizado Sazonal (SOTM), bem como suas versões dinâmicas. Avaliamos empiricamente esses modelos utilizando o conjunto de dados da competição M3, comparando-os com modelos clássicos da literatura, e testamos também o uso do Bagging aplicado para o Modelo Theta Otimizado. Os resultados empíricos mostraram que as versões sazonais dinâmicas e multiplicativas do SOTM apresentaram desempenho superior, especialmente em séries com presença de sazonalidade. Além disso, a aplicação de Bagging para o Modelo Theta Otimizado proporcionou ganhos adicionais em acurácia preditiva.

Palavras-chave: Previsão de Séries Temporais; Método Theta; Modelo Theta Otimizado; Bagging; Competição M3.

# Abstract

Time series forecasting: seasonal extension and bagging for the Optimized Theta Model

This work aimed primarily to propose a seasonal extension for the Optimized Theta Model (OTM), with the purpose of adequately modeling seasonal time series, especially those exhibiting dynamic seasonality, and also to evaluate the use of Bagging applied to the Optimized Theta Model. Initially, we presented a literature review on innovation state-space models, exponential smoothing methods, and the Theta Method, highlighting the models for their optimization. Subsequently, we formally developed the additive and multiplicative versions of the Seasonal Optimized Theta Model (SOTM), as well as their dynamic versions. We empirically evaluated these models using the dataset from the M3 competition, comparing them to classical forecasting models from the literature, and also tested the use of Bagging applied to the Optimized Theta Model. Empirical results showed that the dynamic and multiplicative seasonal versions of the SOTM exhibited superior predictive performance, particularly for series with clear seasonal patterns. Furthermore, applying Bagging to the Optimized Theta Model provided additional improvements in predictive accuracy.

Keywords: Time Series Forecasting; Theta Method; Optimized Theta Model; Bagging; M3 Competition.

# Sumário

|          |                                                    |           |
|----------|----------------------------------------------------|-----------|
| <b>1</b> | <b>Introdução</b>                                  | <b>1</b>  |
| <b>2</b> | <b>Revisão Bibliográfica</b>                       | <b>4</b>  |
| 2.1      | Modelo de Espaço de Estado com inovações . . . . . | 4         |
| 2.2      | Alisamento Exponencial . . . . .                   | 6         |
| 2.2.1    | Alisamento Exponencial Simples . . . . .           | 6         |
| 2.2.2    | Modelo Linear de Holt . . . . .                    | 8         |
| 2.2.3    | Modelo Holt-Winters . . . . .                      | 11        |
| 2.2.4    | Estimação . . . . .                                | 14        |
| 2.3      | Método Theta . . . . .                             | 17        |
| 2.4      | Modelos para Otimização do Método Theta . . . . .  | 20        |
| 2.4.1    | Modelo Theta Otimizado . . . . .                   | 21        |
| 2.4.2    | Modelo Theta Otimizado Dinâmico . . . . .          | 22        |
| 2.4.3    | Estimação . . . . .                                | 24        |
| <b>3</b> | <b>Bagging</b>                                     | <b>25</b> |
| 3.1      | Formulação do Método . . . . .                     | 26        |
| 3.2      | Uso em Séries Temporais . . . . .                  | 28        |
| 3.3      | Limitações . . . . .                               | 32        |

|          |                                                     |           |
|----------|-----------------------------------------------------|-----------|
| <b>4</b> | <b>Modelo Theta Otimizado Sazonal</b>               | <b>33</b> |
| 4.1      | Inclusão do termo sazonal de modo aditivo . . . . . | 35        |
| 4.2      | Inclusão de modo multiplicativo . . . . .           | 39        |
| 4.3      | Estimação . . . . .                                 | 42        |
| 4.4      | Versões dinâmicas . . . . .                         | 43        |
| <b>5</b> | <b>Aplicação na competição M3</b>                   | <b>47</b> |
| 5.1      | Contextualização . . . . .                          | 47        |
| 5.2      | Desenho da aplicação . . . . .                      | 49        |
| 5.3      | Resultados . . . . .                                | 50        |
| 5.4      | Resumo dos resultados . . . . .                     | 56        |
| <b>6</b> | <b>Considerações Finais e Conclusão</b>             | <b>59</b> |
| 6.1      | Possibilidade de trabalhos futuros . . . . .        | 60        |
|          | <b>Referências Bibliográficas</b>                   | <b>60</b> |
| <b>A</b> | <b>Resultados Adicionais</b>                        | <b>67</b> |



# Capítulo 1

## Introdução

A previsão de séries temporais é uma área de pesquisa estatística que tem grande relevância prática em diversas áreas como economia, finanças, meteorologia, produção industrial, saúde pública e muitas outras (Hyndman e Athanasopoulos, 2018; Shumway et al., 2017). Séries temporais são conjuntos de observações registradas sequencialmente com dependência ao longo do tempo e frequentemente apresentam padrões específicos, como tendência e sazonalidade, que devem ser considerados ao se fazer previsões. O desenvolvimento de métodos e modelos estocásticos capazes de lidar eficientemente com essas características específicas torna-se essencial para a precisão e confiabilidade das previsões feitas.

O método Theta, inicialmente proposto por Assimakopoulos e Nikolopoulos (2000), faz previsão de séries temporais sazonais usando uma estratégia que isola a componente de sazonalidade da série dos mecanismos que geram as previsões, por isso dizemos que esse método trata a sazonalidade de forma externa, diferente de quando o método trata ela internamente como por exemplo o método de Holt-Winters (Winters, 1960). O modelo Theta Otimizado (Fiorucci et al., 2016) herda essa característica do método Theta, e, portanto, também não modela a componente de sazonalidade para séries sazonais. Nesses casos, antes das previsões serem feitas, é aplicada uma decomposição da série, que se deseja fazer previsão, a fim de ajustar sua sazonalidade. Assim, em um segundo momento

é feita a ressazonalização dessas previsões. No método Theta, originalmente, esse passo é realizado multiplicando os valores previstos pelos índices sazonais correspondentes no último ciclo sazonal observado. Então, o problema é: *Como adicionar um termo sazonal no Modelo Theta Otimizado?*

Ao propormos uma versão sazonal para o Modelo Theta Otimizado, buscamos fazer com que a componente de sazonalidade da série seja propriamente modelada. Assim, para quando a série que queremos fazer previsão seja sazonal, ainda mais, quando essa sazonalidade for dinâmica, haja melhora preditiva considerando essa versão sazonal.

O *Bagging* (Breiman, 1996) constitui um importante marco no desenvolvimento dos métodos estatísticos e de aprendizado de máquina baseados em *ensembles*. Recentemente surgiram estudos que sugerem que o uso do *Bagging* para o Alisamento Exponencial melhora seus resultados preditivos, principalmente em contextos competitivos de previsão de séries temporais (Bergmeir et al., 2016; Petropoulos, Hyndman et al., 2018). Dessa forma, surgiu a motivação de testarmos sua aplicação para o Modelo Theta Otimizado, que é um Modelo de Espaço de Estado com inovações, da mesma forma que os métodos de alisamento exponencial podem ser escritos como tal.

O objetivo principal do trabalho é propor uma variação sazonal para o Modelo Theta Otimizado, a fim de modelar propriamente séries com sazonalidade, principalmente séries com sazonalidade dinâmica, bem como testar o uso de *Bagging* para o Modelo Theta Otimizado. Os objetivos específicos são: (a) testar e comparar a capacidade preditiva da versão sazonal do Modelo Theta Otimizado com sua versão original; bem como testar e comparar a capacidade preditiva do Modelo Theta Otimizado com *Bagging* com sua versão original. (b) Fazer esses testes e comparações em um contexto competitivo de previsão de séries temporais.

Esta dissertação está organizada em seis capítulos. O Capítulo 2 apresenta a Revisão Bibliográfica necessária para embasar teoricamente os modelos desenvolvidos neste trabalho. O Capítulo 3 aborda detalhadamente o método de *Bagging*, descrevendo sua

formulação e aplicação específica para séries temporais. O Capítulo 4 desenvolve formalmente a versão sazonal proposta para o Modelo Theta Otimizado, incluindo tanto as versões aditivas quanto multiplicativas e suas respectivas versões dinâmicas. O Capítulo 5 traz a aplicação empírica dos modelos propostos na competição M3, apresentando uma análise crítica dos resultados obtidos. Por fim, o Capítulo 6 apresenta as Considerações Finais e Conclusão, sintetizando os principais achados do estudo e indicando possíveis direções para trabalhos futuros.

# Capítulo 2

## Revisão Bibliográfica

Este capítulo reúne os principais fundamentos teóricos que sustentam a formulação do Modelo Theta Otimizado Sazonal e de sua versão dinâmica, que serão apresentados no Capítulo 4. Iniciamos com a apresentação dos Modelos de Espaço de Estado (MEE), com foco em sua formulação por inovações. Em seguida, abordamos alguns dos principais métodos de alisamento exponencial e apresentamos o método Theta de Assimakopoulos e Nikolopoulos (2000), juntamente com os modelos para sua otimização de Fiorucci et al. (2016).

### 2.1 Modelo de Espaço de Estado com inovações

Os modelos de espaço de estado fornecem uma estrutura unificadora para representar e analisar processos dinâmicos em que um conjunto de variáveis latentes (os estados) governam a evolução de variáveis observadas. Além disso, esses modelos oferecem grande flexibilidade na especificação de sua estrutura. Neste trabalho, focaremos na formulação por inovações (Anderson et al., 1979; Aoki, 1990; Hannan et al., 1988), conforme apresentada em Hyndman, Koehler, Ord et al. (2008).

Seja  $y_t$  a observação da série temporal no instante  $t$ . Em modelos de espaço de estado,

essa variável observada é acompanhada por variáveis auxiliares não observadas, denominadas *estados*. Esses estados são representados por um vetor indexado no tempo,  $\vec{x}_t$ , denominado *vetor de estados*, que resume, de forma parcimoniosa, o comportamento passado da série, permitindo determinar seu efeito sobre o comportamento presente e futuro da mesma. Exemplos comuns desses estados incluem o nível, a tendência e a sazonalidade da série. O modelo linear de Espaço de Estado com inovações (MEE linear) pode ser representado pelas seguintes equações:

$$y_t = \vec{w}^\top \vec{x}_{t-1} + \varepsilon_t, \quad (2.1a)$$

$$\vec{x}_t = \mathbf{F} \vec{x}_{t-1} + \vec{g} \varepsilon_t, \quad (2.1b)$$

em que  $\{\varepsilon_t\}$  é um ruído branco com variância  $\sigma^2$ . Uma vez que  $\{\varepsilon_t\}$  representa a parte nova e imprevisível da série, é referido como *inovação*. As inovações constituem a única fonte de aleatoriedade na série observada  $\{y_t\}$ . Os componentes  $\mathbf{F}$ ,  $\vec{g}$  e  $\vec{w}$  são, respectivamente, uma matriz de transição, um vetor de persistência e um vetor de coeficientes, os quais contêm parâmetros geralmente estimados a partir dos dados.

A formulação não linear é obtida a partir de:

$$y_t = w(\vec{x}_{t-1}) + r(\vec{x}_{t-1})\varepsilon_t,$$

$$\vec{x}_t = \mathbf{f}(\vec{x}_{t-1}) + \mathbf{g}(\vec{x}_{t-1})\varepsilon_t,$$

em que  $w(\cdot)$  e  $r(\cdot)$  são funções escalares,  $\mathbf{f}(\cdot)$  e  $\mathbf{g}(\cdot)$  são funções vetoriais, e  $\{\varepsilon_t\}$  representa um ruído branco com variância  $\sigma^2$ . Podemos considerar essa formulação não linear como uma generalização do MEE linear, visto que ele é um caso particular quando  $r$  e  $\vec{g}$  são constantes, e  $w$  e  $\mathbf{f}$  variam linearmente com o vetor de estados.

Se  $\{\varepsilon_t\}$  for homocedástico, a formulação do MEE linear é apropriada. Por outro lado, se esse processo for heterocedástico e, em particular, a variância do erro for proporcional

ao nível atual do processo, a formulação não linear é mais adequada (Hyndman, Koehler, Ord et al., 2008).

## 2.2 Alisamento Exponencial

Um *forecasting method* é um algoritmo que fornece, para uma série temporal, uma previsão pontual: um único valor que prediz a observação de interesse em um instante futuro. Em contrapartida, um modelo estatístico descreve um processo estocástico gerador, capaz de produzir toda uma distribuição de probabilidade para o instante futuro  $n + h$ . Dessa forma, a previsão pontual pode ser obtida facilmente ao se calcular a média (ou mediana) dessa distribuição. Além disso, o modelo permite — diferentemente do *forecasting method* — estimar a variância da previsão, viabilizando o cálculo de intervalos de confiança para a previsão pontual.

Usaremos a notação  $\hat{y}_{n+h|n}$  para denotar a previsão pontual de  $y_{n+h}$  baseada nas informações disponíveis até o tempo  $n$ .

### 2.2.1 Alisamento Exponencial Simples

O método de alisamento exponencial simples, ou SES — do inglês *simple exponential smoothing* —, consiste em fazer a previsão 1 passo à frente  $\hat{y}_{t+1|t}$  como uma soma ponderada da série observada, isto é,

$$\hat{y}_{t+1|t} = \hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots + \alpha(1 - \alpha)^{t-1} y_1.$$

Uma outra forma de escrever seria

$$\hat{y}_{t+1|t} = \sum_{i=0}^{t-1} \alpha(1 - \alpha)^i y_{t-i}. \quad (2.2)$$

Veja que para  $0 < \alpha < 1$  e  $t \rightarrow \infty$ , temos uma média (exponencialmente) ponderada, pois,

nesse caso, os pesos formam uma série geométrica. Essa previsão não considera tendência nem sazonalidade, e quanto maior o  $\alpha$ , maior é o peso para observações recentes.

A equação (2.2) pode ser reescrita como:

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}.$$

Denotaremos  $\ell_t = \hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\ell_{t-1}$ , o qual será computado recursivamente para todo  $t = 1, 2, \dots, n$ , desde que  $\ell_0$  seja conhecido. O termo  $\ell_t$  é chamado de *level*; ele descreve a previsão pontual da série sem tendência e sem sazonalidade. Desta forma, o método SES se resume:

$$\text{Previsão (ajuste)} : \quad \hat{y}_{t|t-1} = \ell_{t-1},$$

$$\text{Level} : \quad \ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}.$$

Originalmente  $\ell_0 = \hat{y}_1 \approx y_1$  foi proposto para inicializar o método, porém abordagens mais recentes têm tratado  $\ell_0 \in \mathbb{R}$  como um parâmetro a ser otimizado conjuntamente com  $\alpha \in (0, 1)$ , que é o parâmetro de alisamento.

O método SES pode ser escrito como um MEE linear. Admitindo erros aditivos  $y_t = \mu_t + \varepsilon_t$ , em que  $\mu_t = \hat{y}_{t|t-1} = E[y_t | y_{t-1}, \dots, y_1] = \ell_{t-1}$ , chegamos às equações:

$$y_t = \ell_{t-1} + \varepsilon_t,$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1},$$

o que implica que:

$$y_t = \ell_{t-1} + \varepsilon_t, \tag{2.3}$$

$$\ell_t = \ell_{t-1} + \alpha \varepsilon_t. \tag{2.4}$$

As equações (2.3) e (2.4) configuram o **modelo** SES( $\alpha, \ell_0$ ), em que  $\alpha \in (0, 1)$  e  $\ell_0 \in \mathbb{R}$  são parâmetros. Isso, pois conseguimos escrever um MEE linear equivalente:

$$\begin{aligned} y_t &= \vec{w}^\top \vec{x}_{t-1} + \varepsilon_t, \\ \vec{x}_t &= \mathbf{F} \vec{x}_{t-1} + \vec{g} \varepsilon_t, \end{aligned}$$

em que  $\vec{w} = 1$ ,  $\vec{x}_t = \ell_t \Leftrightarrow \vec{x}_{t-1} = \ell_{t-1}$ ,  $\mathbf{F} = 1$  e  $\vec{g} = \alpha$ .

A previsão pontual é  $\hat{y}_{n+h|n} = E_n[y_{n+h}]$ , em que:

$$\hat{y}_{n+h|n} = \ell_n, \quad h = 1, 2, 3, \dots \quad (2.5)$$

A variância, ou erro quadrático médio, da previsão é dada por:

$$\begin{aligned} \sigma_{n+h|n}^2 &= \text{Var}_n[y_{n+h}] \\ &= \sigma_\varepsilon^2 + \alpha^2 \sigma_\varepsilon^2 (h - 1). \end{aligned}$$

### 2.2.2 Modelo Linear de Holt

A ideia básica do método linear de Holt é fazer previsão de séries temporais considerando uma correção para a limitação do modelo de alisamento exponencial simples de fazer previsões *flat* — isto é, previsões constantes (2.5). Uma forma de resolver esse problema é incluir um termo de tendência às previsões do método SES, para que sua previsão pontual seja somada a um termo linear com o tempo. Holt implementou essa ideia da seguinte forma:

$$\hat{y}_{n+h|n} = \ell_n + h b_n, \quad h = 1, 2, 3, \dots,$$

em que  $b_n = b_t|_{t=n}$  é o crescimento, ou decrescimento, obtido a partir de um alisamento exponencial da variação dos *levels* passados.

Assim, o método linear de Holt segue:

$$\text{Previsão (ajuste)} : \hat{y}_{t|t-1} = \ell_{t-1} + b_{t-1},$$

$$\text{Level} : \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}),$$

$$\text{Crescimento} : b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1},$$

em que  $\alpha, \beta^* \in (0, 1)$  são parâmetros de alisamento.

O método linear de Holt pode ser escrito como um MEE linear. Admitindo  $\beta = \beta^* \alpha$  e erros aditivos  $y_t = \mu_t + \varepsilon_t$ , em que  $\mu_t = \hat{y}_{t|t-1} = E[y_t | y_{t-1}, \dots, y_1] = \ell_{t-1} + b_{t-1}$ , chegamos às equações:

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t, \quad (2.6)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t, \quad (2.7)$$

$$b_t = b_{t-1} + \beta \varepsilon_t, \quad (2.8)$$

em que  $\ell_0, b_0 \in \mathbb{R}$ ,  $\alpha \in (0, 1)$ ,  $\beta \in (0, \alpha)$ .

As equações (2.6)–(2.8) configuram o **modelo** linear de Holt, pois conseguimos escrever um MEE linear equivalente:

$$y_t = \vec{w}^\top \vec{x}_{t-1} + \varepsilon_t,$$

$$\vec{x}_t = \mathbf{F} \vec{x}_{t-1} + \vec{g} \varepsilon_t,$$

em que  $\vec{w} = (1, 1)^\top$ ,  $\vec{x}_t = (\ell_t, b_t)^\top$ ,  $\mathbf{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ , e  $\vec{g} = (\alpha, \beta)^\top$ .

As previsões são  $\hat{y}_{n+h|n} = E_n[y_{n+h}]$ , em que:

$$\hat{y}_{n+h|n} = \ell_n + h b_n, \quad h = 1, 2, 3, \dots$$

A variância (erro quadrático médio) da previsão será:

$$\begin{aligned}\sigma_{n+h|n}^2 &= \text{Var}_n[y_{n+h}] \\ &= \sigma_\varepsilon^2 \left( 1 + \sum_{i=1}^{h-1} (\alpha + i\beta)^2 \right).\end{aligned}$$

Para inicialização dos estados, a abordagem mais simples é tomar  $b_0 = 0$  e  $\ell_0 = y_1$ . Uma forma mais moderna seria otimizar  $b_0, \ell_0 \in \mathbb{R}$  em conjunto com os parâmetros de alisamento  $\alpha, \beta^* \in \mathbb{R}$ .

### Modelo SES com drift

Um caso particular do **modelo** linear de Holt ocorre quando fixamos  $\beta^* = 0$ , ou seja,  $\beta = 0$ . Nesse cenário, a componente de tendência torna-se constante, o que implica que  $b_t = b_0$  para todo  $t$ , representando uma taxa de crescimento (ou decrescimento) fixa — denominada *drift*. Esse caso específico resulta no chamado modelo de alisamento exponencial simples com drift (SES-d), cuja formulação é:

$$\begin{aligned}y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t, \\ \ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \\ b_t &= b_{t-1} + \cancel{\beta\varepsilon_t^0},\end{aligned}$$

ou, de forma equivalente, considerando  $b_t = b_{t-1} = b$ :

$$y_t = \ell_{t-1} + \cancel{b_{t-1}}^b + \varepsilon_t, \tag{2.9}$$

$$\ell_t = \ell_{t-1} + \cancel{b_{t-1}}^b + \alpha\varepsilon_t. \tag{2.10}$$

### 2.2.3 Modelo Holt-Winters

O método de Holt-Winters supõe que a série observada pode ser decomposta em termos de uma tendência, sazonalidade e um ruído. Por exemplo:

$$Y_t = T_t + S_t + R_t,$$

o que nos leva ao desenvolvimento de  $Y_t$  em termos de um ajuste de Holt para  $T_t$ , e, para  $S_t$ , o índice sazonal na posição equivalente a  $t$  no último ciclo sazonal. Nesse caso:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t,$$

em que  $m$  é o número de observações dentro do ciclo sazonal.

O método de Holt-Winters é dado por:

$$\text{Previsão (ajuste)} : \hat{y}_{t|t-1} = \ell_{t-1} + b_{t-1} + s_{t-m},$$

$$\text{Level} : \ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}),$$

$$\text{Crescimento} : b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1},$$

$$\text{Sazonalidade} : s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m},$$

em que  $\alpha, \beta^*, \gamma \in (0, 1)$  são parâmetros de alisamento e  $\ell_0, b_0 \in \mathbb{R}$ ,  $(s_{1-m}, s_{2-m}, \dots, s_0)^\top \in \mathbb{R}^m$  são parâmetros de inicialização. Veja que  $\{y_t - s_{t-m}\}$  consiste na série com sazonalidade ajustada e  $\{y_t - \ell_{t-1} - b_{t-1}\}$  na série com tendência ajustada.

O método de Holt-Winters pode ser escrito como um MEE linear. Admitindo erros aditivos  $y_t = \mu_t + \varepsilon_t$ , em que

$$\mu_t = \hat{y}_{t|t-1} = E[y_t \mid y_{t-1}, \dots, y_1] = \ell_{t-1} + b_{t-1} + s_{t-m},$$

chegamos às equações:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \quad (2.11)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t, \quad (2.12)$$

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

$$s_t = s_{t-m} + \gamma \varepsilon_t, \quad (2.14)$$

em que:

$$\beta = \beta^* \alpha, \quad \ell_0, b_0 \in \mathbb{R}, \quad \alpha, \gamma \in (0, 1), \quad \beta \in (0, \alpha).$$

As equações (2.11)–(2.14) configuram o **modelo** estocástico de Holt-Winters aditivo ou  $\text{HW}_a(\ell_0, b_0, s_{1-m}, \dots, s_0, \alpha, \beta, \gamma)$ , pois conseguimos escrever um MEE linear equivalente:

$$y_t = \vec{w}^\top \vec{x}_{t-1} + \varepsilon_t,$$

$$\vec{x}_t = \mathbf{F} \vec{x}_{t-1} + \vec{g} \varepsilon_t,$$

em que:

$$\vec{w} = (1, 1, 0, 0, \dots, 0, 1)^\top,$$

$$\vec{g} = (\alpha, \beta, \gamma, 0, \dots, 0)^\top,$$

$$\vec{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})^\top,$$

$$\vec{x}_{t-1} = (\ell_{t-1}, b_{t-1}, s_{t-1}, s_{t-2}, \dots, s_{t-m})^\top,$$

e:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

A previsão pontual é dada por:

$$\hat{y}_{n+h|n} = E_n[y_{n+h}] = \ell_n + hb_n + s_{n+h_m^+-m},$$

em que  $h_m^+ = [(h-1) \bmod m] + 1$ . A variância condicional é dada por:

$$\sigma_{n+h|n}^2 = \left[ 1 + \sum_{i=1}^{h-1} (\alpha + i\beta)^2 + (k-1)\gamma^2 + 2 \sum_{j=1}^{k-1} \gamma(\alpha + j\beta) \right] \sigma_\varepsilon^2,$$

em que  $k$  é o menor número inteiro positivo tal que  $n - km + h \leq n$ .

Como adotamos a suposição de que  $Y_t = T_t + S_t + R_t$ , o modelo descrito é conhecido como o modelo de Holt-Winters aditivo. É possível formular uma versão multiplicativa desse modelo, assumindo  $Y_t = T_t \times S_t + R_t$ . O Holt-Winters multiplicativo configura um MEE não linear [2.1](#).

As duas variações do modelo diferem na forma de modelagem do componente de sazonalidade. O Holt-Winters aditivo é mais adequado quando as variações sazonais permanecem aproximadamente constantes ao longo da série, enquanto o multiplicativo é mais apropriado quando essas variações mudam proporcionalmente ao nível da série (Hyndman e Athanasopoulos, [2021](#)).

### 2.2.4 Estimação

Usando a densidade de  $\mathbf{y} = (y_1, \dots, y_n)^\top$  sob a suposição de normalidade dos erros, essa expressão se reduz a:

$$p(\mathbf{y}) = (2\pi\sigma^2)^{-n/2} \left| \prod_{t=1}^n r(\vec{x}_{t-1}) \right|^{-1} \exp \left( -\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right),$$

pois  $p(\mathbf{y}) = p(y_1) \prod_{t=2}^n p(y_t | y_{t-1}, \dots, y_1)$ , em que  $p(y_t | y_{t-1}, \dots, y_1)$  é a densidade da distribuição  $N(w(\vec{x}_{t-1}), [r(\vec{x}_{t-1})]^2 \sigma^2)$ .

A função de verossimilhança baseia-se nessa densidade. Ela depende do vetor de parâmetros de alisamento e *damping*  $\vec{\theta} = (\alpha, \beta, \gamma, \phi)^\top$ , da variância das inovações  $\sigma^2$ , e também do vetor de estados iniciais  $\vec{x}_0$  (parâmetros de inicialização). Assim:

$$L(\vec{\theta}, \vec{x}_0, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \left| \prod_{t=1}^n r(\vec{x}_{t-1}) \right|^{-1} \exp \left( -\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right). \quad (2.15)$$

O logaritmo da verossimilhança é:

$$\log L = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{t=1}^n \log |r(\vec{x}_{t-1})| - \frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2. \quad (2.16)$$

Ao derivar a equação (2.16) em relação a  $\sigma^2$  e igualar a zero, obtemos o estimador de máxima verossimilhança:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2. \quad (2.17)$$

Eliminando  $\sigma^2$  da função (2.15), temos a verossimilhança concentrada:

$$L(\vec{\theta}, \vec{x}_0 | \mathbf{y}) = (2\pi e \hat{\sigma}^2)^{-n/2} \left| \prod_{t=1}^n r(\vec{x}_{t-1}) \right|^{-1}.$$

Logo, o logaritmo negativo da verossimilhança multiplicado por dois é:

$$\begin{aligned} -2 \log L(\vec{\theta}, \vec{x}_0 \mid \mathbf{y}) &= n \log(2\pi e \hat{\sigma}^2) + 2 \sum_{t=1}^n \log |r(\vec{x}_{t-1})| \\ &= c_n + n \log \left( \sum_{t=1}^n \varepsilon_t^2 \right) + 2 \sum_{t=1}^n \log |r(\vec{x}_{t-1})|, \end{aligned}$$

em que  $c_n$  depende de  $n$ , mas não de  $\vec{\theta}$  ou  $\vec{x}_0$  (Hyndman, Koehler, Ord et al., 2008).

Assim, as estimativas de máxima verossimilhança para os parâmetros de alisamento em  $\vec{\theta}$  e o vetor de estados iniciais  $\vec{x}_0$  podem ser obtidas pela minimização de  $L^*(\vec{\theta}, \vec{x}_0)$ , sob a suposição de normalidade dos erros, ou seja:

$$\hat{\vec{\theta}}, \hat{\vec{x}_0} = \arg \min_{\vec{\theta}, \vec{x}_0} L^*(\vec{\theta}, \vec{x}_0),$$

em que

$$L^*(\vec{\theta}, \vec{x}_0) = n \log \left( \sum_{t=1}^n \varepsilon_t^2 \right) + 2 \sum_{t=1}^n \log |r(\vec{x}_{t-1})|,$$

o que só pode ser obtido de forma numérica. De forma equivalente, pode-se minimizar a soma de quadrados aumentada:

$$S(\vec{\theta}, \vec{x}_0) = \left[ \exp \left( L^*(\vec{\theta}, \vec{x}_0) \right) \right]^{1/n} = \left| \prod_{t=1}^n r(\vec{x}_{t-1}) \right|^{2/n} \cdot \sum_{t=1}^n \varepsilon_t^2. \quad (2.18)$$

No caso homocedástico (modelos com erros aditivos, como todos os vistos ao longo do Capítulo 2),  $r(\vec{x}_{t-1}) = 1$ , então a equação (2.18) se reduz à soma tradicional de quadrados.

Embora a equação (2.18) seja derivada de (2.15), também se pode fazer a estimação assumindo diretamente a minimização de  $S(\vec{\theta}, \vec{x}_0)$  (Hyndman, Koehler, Ord et al., 2008). Essa abordagem, chamada de *Mínimos Quadrados Aumentados* (*Augmented Least Squares* — ALS), não exige suposições sobre a distribuição dos erros. Quando usada dessa forma ela produz os chamados *quasi-maximum likelihood estimators*.

### Valores iniciais para otimização

Tradicionalmente, e também como indica Hyndman, Koehler, Ord et al. (2008), os valores iniciais em  $\vec{x}_0$  são tratados como parâmetros a serem estimados juntamente com os parâmetros de alisamento. Nesse caso, os valores que inicializam o algoritmo para minimizar  $L^*(\vec{\theta}, \vec{x}_0)$  podem seguir o esquema heurístico de Hyndman, Koehler, Snyder et al. (2002):

- **Componente de sazonalidade inicial.** Para séries sazonais, calcula-se uma média móvel de  $2 \times m$  — isto é, uma média móvel de ordem  $m$ , seguida por uma média móvel de ordem 2 — nos primeiros anos da série. Denote isso por  $\{f_t\}$ , com  $t = \frac{m}{2} + 1, \frac{m}{2} + 2, \dots$ . Para sazonalidade aditiva, a série é ajustada subtraindo  $y_t - f_t$  para retirar sua tendência. Para sazonalidade multiplicativa, esse ajuste é feito com a divisão:  $y_t / f_t$ . Os índices sazonais iniciais,  $s_{-m+1}, \dots, s_0$ , são calculados como a média da série observada, com a tendência ajustada, para todos os tempos  $t$  no respectivo ciclo sazonal, considerando os primeiros anos da série. Esses índices sazonais são normalizados para somarem zero na sazonalidade aditiva e  $m$  na sazonalidade multiplicativa.
- **Componente de *level* inicial.** Para dados sazonais, calcula-se uma tendência (linear) por regressão linear sobre as dez primeiras observações da série, com a sazonalidade ajustada, contra o tempo  $t = 1, \dots, 10$ . Para dados não sazonais, a tendência linear é ajustada nos primeiros dez valores observados. O valor  $\ell_0$  é definido como o intercepto dessa tendência.
- **Componente de crescimento inicial.** Define-se  $b_0$  como a inclinação da tendência calculada anteriormente.

Já os valores, para inicialização do algoritmo de otimização, em  $\vec{\theta} = (\alpha, \beta, \gamma, \phi)^\top$  podem ser  $(0.1, 0.01, 0.01, 0.99)^\top$ , como indica Hyndman, Koehler, Ord et al. (2008).

## 2.3 Método Theta

O método Theta de Assimakopoulos e Nikolopoulos (2000) é aplicado a séries temporais não sazonais ou com sazonalidade ajustada (*série temporal sem sazonalidade*). Sua principal característica é a decomposição da série sem sazonalidade em duas componentes denominadas *linhas theta*, construídas a partir da aplicação de coeficientes  $\theta_1$  e  $\theta_2$  à segunda diferença da série sem sazonalidade (Fiorucci et al., 2016). Quando  $\theta < 1$ , as diferenças são suavizadas, o que favorece a representação da tendência de longo prazo. Em contrapartida, valores de  $\theta > 1$  amplificam as curvaturas locais, destacando os movimentos de curto prazo (Assimakopoulos, 1995).

As linhas resultantes, denotadas aqui por  $Z(\theta_1)$  e  $Z(\theta_2)$ , mantêm o mesmo valor médio e inclinação da série sem sazonalidade, mas apresentam curvaturas alteradas de acordo com o valor de  $\theta$ , permitindo filtrar ou realçar estruturas específicas dos dados. Essa decomposição tem como vantagem a capacidade de capturar padrões que não seriam plenamente explorados apenas por meio da extrapolação direta da série sem sazonalidade.

Cada linha theta é projetada no futuro separadamente, utilizando métodos adequados de previsão. As previsões pontuais da série sem sazonalidade são então obtidas por meio da recomposição dessas projeções via um esquema de combinação. A prática de combinar previsões é amplamente reconhecida na literatura por promover maior precisão e robustez (Clemen, 1989; Makridakis e Winkler, 1983; Petropoulos, Makridakis et al., 2014), o que justifica sua adoção no método Theta.

Além de sua simplicidade estrutural, o método destacou-se pelo bom desempenho em competições de previsão como a M3 (Koning et al., 2005; Makridakis e Hibon, 2000) e a competição de turismo (Athanasopoulos et al., 2011), consolidando-se como um dos modelos de referência na área. A seguir, conforme a exposição detalhada de Fiorucci et al. (2016), apresenta-se a formulação matemática do método Theta, originalmente proposto por Assimakopoulos e Nikolopoulos (2000).

Os autores definem a linha theta  $\mathbf{Z}(\theta) = (Z_1(\theta), \dots, Z_n(\theta))^\top$  como a solução da equação:

$$\nabla^2 Z_t(\theta) = \theta \nabla^2 Y_t, \quad t = 3, \dots, n, \quad (2.19)$$

em que  $Y_1, \dots, Y_n$  representa a série temporal sem sazonalidade, e  $\nabla$  é o operador diferença, definido por  $\nabla X_t = X_t - X_{t-1}$ . Os valores iniciais  $Z_1(\theta)$  e  $Z_2(\theta)$  são determinados pela minimização da soma dos quadrados dos resíduos  $\sum_{t=1}^n (Y_t - Z_t(\theta))^2$ .

Posteriormente, Hyndman e Billah (2003) apresentaram uma solução analítica para a linha theta, expressa como:

$$Z_t(\theta) = \theta Y_t + (1 - \theta)(A_n + B_n t), \quad t = 1, \dots, n,$$

em que  $A_n$  e  $B_n$  são os coeficientes de mínimos quadrados de uma regressão linear simples da série, sem sazonalidade,  $Y_1, \dots, Y_n$  sobre os tempos  $t = 1, \dots, n$ , dados por:

$$A_n = \frac{1}{n} \sum_{t=1}^n Y_t - \frac{n+1}{2} B_n, \quad B_n = \frac{6}{n^2 - 1} \left( \frac{2}{n} \sum_{t=1}^n t Y_t - \frac{n+1}{n} \sum_{t=1}^n Y_t \right).$$

Sob essa formulação, as linhas theta podem ser interpretadas como transformações da série sem sazonalidade que preservam seu nível médio e tendência linear, ao mesmo tempo em que modulam suas curvaturas locais. É importante destacar que  $A_n$  e  $B_n$  são funções exclusivamente dos dados da série sem sazonalidade e não constituem parâmetros ajustáveis do método.

As previsões produzidas pelo método Theta são obtidas diretamente para a série sem sazonalidade. Para um horizonte de  $h$  passos à frente do tempo  $n$ , calcula-se a média das extrapolações individuais das duas linhas theta:  $\tilde{Z}_{n+h|n}(0)$ , obtida por regressão linear, e  $\tilde{Z}_{n+h|n}(2)$ , obtida por alisamento exponencial simples (SES), conforme detalhado na Seção 2.2.1. Essa combinação é feita com pesos fixos de 50% para cada componente.

O algoritmo completo do método Theta de Assimakopoulos e Nikolopoulos (2000),

para previsão de uma *série temporal de interesse*, segue as etapas descritas abaixo:

1. **Ajuste da sazonalidade:** Avalia-se a presença de sazonalidade estatisticamente significativa na série temporal de interesse. Utiliza-se o critério:

$$|r_m| > q_{1-\alpha/2} \sqrt{\frac{1 + 2 \sum_{i=1}^{m-1} r_i^2}{n}},$$

em que  $r_k$  representa a autocorrelação de ordem  $k$ ,  $m$  é o número de períodos por ciclo sazonal (por exemplo, 12 para séries mensais),  $n$  é o tamanho da amostra,  $q$  é a função quantil da normal padrão, e  $(1 - \alpha)\%$  é o nível de confiança — sendo adotado, originalmente, o valor  $\alpha = 0,10$ .

Caso identificada sazonalidade, realiza-se seu ajuste por decomposição clássica, assumindo componente sazonal multiplicativo. Nesse caso, denomina-se por série sem sazonalidade aquela obtida após esse ajuste. Caso contrário, a série sem sazonalidade é tida como a própria série de interesse.

2. **Decomposição:** A série sem sazonalidade é decomposta em duas linhas theta:  $Z(0)$ , que captura a tendência de longo prazo, e  $Z(2)$ , que enfatiza os componentes de curto prazo.
3. **Extrapolação:** A linha  $Z(0)$  é extrapolada por regressão linear, e  $Z(2)$ , por SES.
4. **Combinação:** As previsões para a série sem sazonalidade são obtidas pela média simples das projeções de  $Z(0)$  e  $Z(2)$ .
5. **Ressazonalização:** Se houve ajuste sazonal na Etapa 1, para gerar as previsões da série de interesse, aplica-se a reversão multiplicando os valores previstos na Etapa 4 pelos índices sazonais correspondentes aos observados no último ciclo sazonal. Caso contrário, as previsões da série de interesse são as obtidas na Etapa 4.

Esse procedimento com dois valores fixos de  $\theta$ , extrapolações específicas e combinação com pesos iguais foi o utilizado na competição M3 (Makridakis e Hibon, 2000), onde ele apresentou desempenho superior aos demais concorrentes, consolidando-se como um modelo de referência em previsão de séries temporais.

## 2.4 Modelos para Otimização do Método Theta

Nesta seção, apresentam-se os modelos propostos por Fiorucci et al. (2016) como extensões do método Theta. Tais modelos visam aprimorar a escolha das linhas theta a partir de critérios de otimização, em contraste com os valores fixos adotados por Assimakopoulos e Nikolopoulos (2000). A formulação é desenvolvida dentro de um arcabouço de modelos de espaço de estados, permitindo maior flexibilidade e fundamentação estatística.

Um aspecto chave do método Theta é que, por definição, ele permite a combinação de diferentes linhas theta com pesos que podem ser fixos ou otimizados. No entanto, na formulação clássica adotada por Assimakopoulos e Nikolopoulos (2000), os coeficientes  $\theta$  são definidos *a priori*, restringindo o modelo a uma estrutura específica (cf. Seção 2.3).

Seja  $Y_1, \dots, Y_n$  uma série temporal não sazonal ou com sazonalidade previamente ajustada por decomposição clássica multiplicativa. Define-se a combinação linear de duas linhas theta como:

$$X_t = \omega Z_t(\theta_1) + (1 - \omega) Z_t(\theta_2), \quad (2.20)$$

em que  $\omega \in [0, 1]$  é o peso da combinação, e os coeficientes  $\theta_1 < 1$ ,  $\theta_2 \geq 1$ . Fiorucci et al. (2016) mostram que, para que  $X_t = Y_t$  para todo  $t = 1, \dots, n$ , o peso ótimo deve ser:

$$\omega := \omega(\theta_1, \theta_2) = \frac{\theta_2 - 1}{\theta_2 - \theta_1}. \quad (2.21)$$

Essa solução é única e garante a recomposição exata da série original a partir das linhas theta. O método Theta é recuperado ao se fixar  $\theta_1 = 0$  e  $\theta_2 = 2$ , resultando em

$\omega = 0,5$ . Assim, as equações (2.20) e (2.21) fundamentam uma generalização do método Theta que mantém a propriedade de reconstrução exata da série, independentemente dos valores escolhidos para  $\theta_1$  e  $\theta_2$ .

### 2.4.1 Modelo Theta Otimizado

Sejam  $A_n$  e  $B_n$  os coeficientes de uma regressão linear simples da série  $\{Y_t, t = 1, \dots, n\}$  sobre  $t$ , conforme definidos na Seção 2.3. O *Modelo Theta Otimizado* (OTM) é formulado como um modelo de espaço de estados estocástico, para uma série sem sazonalidade, composto pelas seguintes equações:

$$Y_t = \mu_t + \varepsilon_t, \quad (2.22)$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left( \frac{1 - (1 - \alpha)^t}{\alpha} \right) B_n \right\}, \quad (2.23)$$

$$\ell_t = \alpha Y_t + (1 - \alpha) \ell_{t-1}, \quad (2.24)$$

com parâmetros  $\ell_0 \in \mathbb{R}$ ,  $\alpha \in (0, 1)$  e  $\theta \in [1, \infty)$ . Os parâmetros  $\theta$ ,  $\alpha$  e  $\ell_0$  são estimados a partir dos dados.

A previsão pontual  $h$  passos à frente, calculada na origem  $n$ , é dada por:

$$\begin{aligned} \hat{Y}_{n+h|n} &= \mathbb{E}[Y_{n+h} \mid Y_1, \dots, Y_n] \\ &= \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + \left[ (h - 1) + \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] B_n \right\}. \end{aligned}$$

A variância condicional associada é:

$$\text{Var}[Y_{n+h} \mid Y_1, \dots, Y_n] = [1 + (h - 1)\alpha^2] \sigma^2.$$

O intervalo de previsão de  $(1 - a)\%$  para  $Y_{n+h}$  é construído como:

$$\hat{Y}_{n+h|n} \pm q_{1-a/2} \sqrt{[1 + (h - 1)\alpha^2] \sigma^2}.$$

Considerando a série sem sazonalidade e quando  $\theta = 2$ , o OTM reproduz as previsões do método Theta. Para mais detalhes sobre a equivalência entre OTM e o modelo SES-d, ver Fiorucci et al. (2016).

#### 2.4.2 Modelo Theta Otimizado Dinâmico

Para ampliar a flexibilidade do modelo, os coeficientes  $A_n$  e  $B_n$  podem ser tratados como funções dinâmicas do tempo. Ao atualizar o estado de  $t$  para  $t + 1$ , consideram-se apenas os dados disponíveis (sem sazonalidade) até o tempo  $t$ , ou seja,  $Y_1, \dots, Y_t$ . Com isso, as equações do modelo OTM são reescritas substituindo-se  $A_n$  e  $B_n$  por  $A_t$  e  $B_t$ .

A previsão de um passo à frente no tempo  $t$  é dada por:

$$\hat{Y}_{t+1|t} = \ell_t + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^t A_t + \left[ \frac{1 - (1 - \alpha)^{t+1}}{\alpha} \right] B_t \right\}.$$

O sistema dinâmico é descrito pelas equações de espaço de estados:

$$Y_t = \mu_t + \varepsilon_t, \tag{2.25}$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_{t-1} + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_{t-1} \right\}, \tag{2.26}$$

$$\ell_t = \alpha Y_t + (1 - \alpha) \ell_{t-1}, \tag{2.27}$$

com os coeficientes dinâmicos atualizados recursivamente por:

$$\begin{aligned} A_t &= \bar{Y}_t - \frac{t+1}{2} B_t, \\ B_t &= \frac{1}{t+1} \left[ (t-2)B_{t-1} + \frac{6}{t}(Y_t - \bar{Y}_{t-1}) \right], \\ \bar{Y}_t &= \frac{1}{t} \left[ (t-1)\bar{Y}_{t-1} + Y_t \right], \end{aligned}$$

com condições iniciais  $A_0 = B_0 = B_1 = \bar{Y}_0 = 0$ . Os parâmetros do modelo continuam sendo  $\ell_0 \in \mathbb{R}$ ,  $\alpha \in (0, 1)$  e  $\theta \in [1, \infty)$ .

Esse modelo é denominado *Modelo Theta Otimizado Dinâmico* (DOTM), e constitui uma extensão do OTM, permitindo que a componente de tendência se adapte ao comportamento recente da série.

A previsão 1 passo à frente de  $n$  é:

$$\hat{Y}_{n+1|n} = \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1-\alpha)^n A_n + \left[ \frac{1 - (1-\alpha)^{n+1}}{\alpha} \right] B_n \right\}.$$

Para horizontes maiores ( $h \geq 2$ ), as previsões são obtidas recursivamente, substituindo os valores futuros  $Y_{n+1}, \dots, Y_{n+h-1}$  por suas respectivas previsões. A variância condicional  $\text{Var}[Y_{n+h} \mid Y_1, \dots, Y_n]$  não possui forma fechada neste modelo. Sua estimação pode ser realizada por reamostragem (e.g., *bootstrap*) a partir das simulações de  $Y_{n+h}$ , considerando o modelo ajustado.

Diferentemente de modelos lineares como o OTM ou o SES-d, o DOTM permite componentes de tendência não lineares, tanto no ajuste in-sample quanto na projeção out-of-sample, conferindo maior flexibilidade ao modelo (Fiorucci et al., 2016).

### 2.4.3 Estimação

A estimação dos parâmetros é feita por meio da minimização da soma dos quadrados dos erros (SSE, do inglês *Sum of Squared Errors*):

$$(\hat{\ell}_0, \hat{\alpha}, \hat{\theta}) = \arg \min_{\ell_0, \alpha, \theta} \sum_{t=t_0}^n (Y_t - \mu_t)^2.$$

Para o modelo DOTM, Fiorucci et al. (2016) recomendam iniciar a soma a partir de  $t_0 = 3$ , pois  $A_t$  e  $B_t$  são coeficientes de regressão linear e requerem ao menos duas observações.

Assumindo que  $\{\varepsilon_t\}$  é um processo i.i.d.  $N(0, \sigma^2)$ , a estimativa de SSE coincide com o estimador de máxima verossimilhança. Ao substituir  $\sigma^2$  por  $\hat{\sigma}^2 = \text{SSE}/n$ , a log-verossimilhança é:

$$\ell(\ell_0, \alpha, \theta) = -\frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2} (1 + \log 2\pi).$$

**Nota:** A formulação matemática e os algoritmos descritos nesta seção seguem de forma adaptada o trabalho original de Fiorucci et al. (2016), com reorganização textual e exposições complementares para fins didáticos nesta dissertação.

# Capítulo 3

## Bagging

A técnica conhecida como *Bagging* (do inglês *Bootstrap Aggregating*), proposta inicialmente por Breiman ([1996](#)), constitui um marco no desenvolvimento de métodos estatísticos e de aprendizado de máquina baseados em *ensembles*. A ideia central deste método consiste em gerar múltiplas versões de um estimador por meio de reamostragens bootstrap dos dados originais, combinando posteriormente esses resultados para obter uma única estimativa mais robusta e estável (Hastie et al., [2009](#); Dietterich, [2000](#)).

A motivação original por trás do Bagging está associada ao problema dos chamados estimadores instáveis, cuja variabilidade pode levar a previsões pouco confiáveis. Tais estimadores, como árvores de decisão profundas ou redes neurais, apresentam alta sensibilidade a pequenas perturbações nos dados de treinamento (James et al., [2021](#)). O Bagging, ao explorar reamostragens e agregações, atua como um mecanismo eficaz para reduzir essa variância, estabilizar os resultados e aumentar a precisão preditiva (Zhou, [2012](#)).

De maneira intuitiva, o Bagging se apoia no princípio conhecido como “sabedoria das multidões” (*wisdom of the crowds*) (Surowiecki, [2004](#); Kuncheva, [2014](#)). Em outras palavras, o método constrói uma “multidão” de modelos individuais ligeiramente diferentes uns dos outros e agrega suas previsões ou estimativas. Como esses modelos individuais

são construídos a partir de diferentes subconjuntos dos dados originais, suas variações tendem a se compensar mutuamente, produzindo um resultado agregado que frequentemente supera, em termos de desempenho preditivo, aquele produzido por um único modelo.

Dessa forma, este capítulo pretende explorar detalhadamente o método Bagging, abordando sua implementação prática, adaptações para problemas específicos (especialmente séries temporais, foco desta dissertação), vantagens e limitações, além de apresentar aplicações concretas que ilustram sua utilidade no contexto de análise e previsão de séries temporais.

### 3.1 Formulação do Método

Formalmente, o método de *Bootstrap Aggregating* (Bagging) pode ser descrito como uma técnica que combina múltiplas versões de um estimador, obtidas por meio da reamostragem bootstrap dos dados originais.

Considere um conjunto de dados observado

$$\mathcal{D} = \{(\vec{x}_i, y_i)\}_{i=1}^n,$$

em que  $\vec{x}_i$  representa o vetor de variáveis preditoras e  $y_i$  a variável resposta associada à  $i$ -ésima unidade amostral.

Dado um estimador ou preditor instável  $\hat{f}(\cdot)$ , construído a partir do conjunto de dados original  $\mathcal{D}$ , o procedimento geral do Bagging pode ser descrito pelos seguintes passos:

1. Gerar  $B$  novas amostras bootstrap independentes  $\mathcal{D}^{*(1)}, \mathcal{D}^{*(2)}, \dots, \mathcal{D}^{*(B)}$ , cada uma contendo  $n$  observações obtidas por amostragem com reposição a partir de  $\mathcal{D}$ .
2. Para cada amostra bootstrap  $\mathcal{D}^{*(b)}$ , construir um estimador individual  $\hat{f}^{*(b)}(\cdot)$ , com  $b = 1, 2, \dots, B$ .
3. Obter o estimador agregado  $\hat{f}_{\text{bag}}(\cdot)$  por meio da combinação simples dos estimadores

individuais. Em problemas de regressão, utiliza-se a média:

$$\hat{f}_{\text{bag}}(\vec{x}) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*(b)}(\vec{x}),$$

enquanto que, em problemas de classificação, utiliza-se a moda:

$$\hat{f}_{\text{bag}}(\vec{x}) = \text{moda} \left( \hat{f}^{*(1)}(\vec{x}), \dots, \hat{f}^{*(B)}(\vec{x}) \right).$$

Em suma, como descrito no Algoritmo 1, o método Bagging utiliza reamostragens bootstrap para gerar diversas versões do conjunto de dados, aplicando o estimador base a cada uma delas e combinando as estimativas resultantes, com o objetivo de reduzir a variância e aumentar a estabilidade do modelo preditivo.

Esse método pode ser interpretado como um caso especial de técnicas de agregação baseadas em bootstrap, amplamente utilizadas para reduzir a variância e aumentar a estabilidade de estimadores que, individualmente, exibem alta sensibilidade a variações nos dados (Breiman, 1996; Hastie et al., 2009).

A eficiência do Bagging, no que diz respeito à melhoria do desempenho preditivo, está intimamente relacionada à propriedade estatística conhecida como *redução de variância* (James et al., 2021; Zhou, 2012). De forma intuitiva, supondo que cada estimador individual possua alta variabilidade, a média agregada de múltiplas versões independentes tende a apresentar menor variância do que cada estimador individual tomado isoladamente. A magnitude da redução dependerá diretamente da correlação entre os estimadores obtidos via bootstrap: quanto menor essa correlação, maior tende a ser a redução de variância alcançada (Kuncheva, 2014).

Por outro lado, é importante destacar que o Bagging se mostra especialmente eficaz para estimadores instáveis, tais como árvores de decisão profundas, redes neurais complexas ou modelos não paramétricos flexíveis, cuja alta variabilidade pode ser compensada

pela agregação (Dietterich, 2000). Já em modelos naturalmente estáveis, como modelos lineares ou aditivos simples, a técnica tende a apresentar ganhos menos expressivos, ou até mesmo nulos, em termos de desempenho preditivo (James et al., 2021).

---

**Algoritmo 1:** Bagging (Bootstrap Aggregating)

---

**Entrada:** Conjunto de dados original  $\mathcal{D} = \{(\vec{x}_i, y_i)\}_{i=1}^n$ ; número de modelos  $B$ ;

estimador base  $f$

**Saída** : Estimador agregado  $\hat{f}_{\text{bag}}$

```

1 para  $b = 1$  até  $B$  faça
2   | Gerar uma amostra bootstrap  $\mathcal{D}^{*(b)}$  com reposição, de tamanho  $n$ , a partir de
   |  $\mathcal{D}$ 
3   | Ajustar o estimador base  $f$  aos dados  $\mathcal{D}^{*(b)}$ , obtendo  $\hat{f}^{*(b)}$ 
4 fim
5 para cada nova observação  $\vec{x}$  faça
6   | Obter as  $B$  predições individuais  $\hat{f}^{*(1)}(\vec{x}), \dots, \hat{f}^{*(B)}(\vec{x})$ 
7   | if tarefa de regressão then
8   |   |  $\hat{f}_{\text{bag}}(\vec{x}) \leftarrow \frac{1}{B} \sum_{b=1}^B \hat{f}^{*(b)}(\vec{x})$ 
9   | else
10  |   |  $\hat{f}_{\text{bag}}(\vec{x}) \leftarrow \text{moda}(\hat{f}^{*(1)}(\vec{x}), \dots, \hat{f}^{*(B)}(\vec{x}))$ 
11  | end
12 fim

```

---

### 3.2 Uso em Séries Temporais

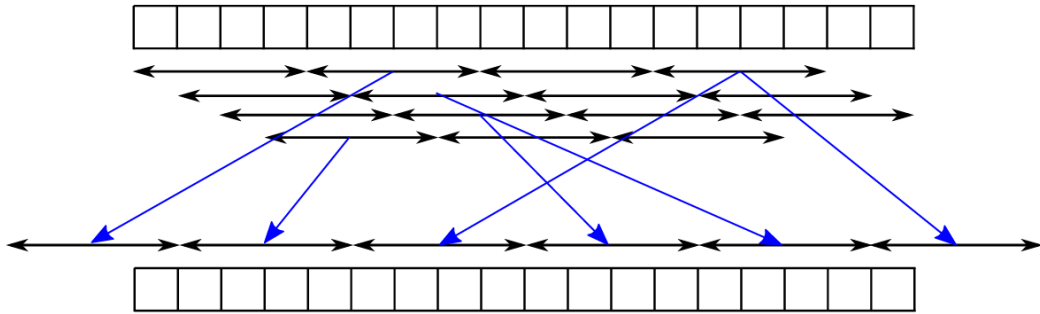
A aplicação direta de métodos estatísticos e de aprendizado de máquina às séries temporais apresenta desafios específicos, decorrentes fundamentalmente da estrutura temporal dos dados. Diferentemente de outras áreas, em séries temporais as observações são naturalmente ordenadas no tempo e apresentam dependência temporal. Essa dependência implica que o valor de uma observação está fortemente associado às observações anteriores,

violando a suposição de independência frequentemente assumida em métodos estatísticos clássicos.

Métodos baseados em *Bootstrap Aggregating* (Bagging), como originalmente propostos por Breiman (1996), partem da premissa de que as observações são independentes e identicamente distribuídas. No contexto temporal, entretanto, essa hipótese não se mantém, impondo limitações ao uso direto do bootstrap tradicional e exigindo técnicas específicas que possam preservar a estrutura de autocorrelação presente nos dados.

Adicionalmente, séries temporais reais frequentemente exibem outros fenômenos, como não estacionariedade, tendências de longo prazo, sazonalidade e heterocedasticidade (variância não constante ao longo do tempo). Tais fenômenos demandam cuidados adicionais na reamostragem e agregação dos modelos individuais, para evitar que padrões estruturais importantes sejam destruídos ou mal representados nas amostras geradas.

Neste sentido, a aplicação eficaz do Bagging em séries temporais exige adaptações metodológicas rigorosas. Uma das principais estratégias adotadas é o uso de técnicas especializadas de reamostragem por bootstrap, especialmente o método conhecido como *Moving Block Bootstrap* (MBB, bootstrap de blocos móveis), originalmente proposto por Künsch (1989). O MBB consiste na geração de amostras bootstrap por meio da seleção aleatória de blocos contíguos de observações, preservando, assim, a dependência temporal dentro de cada bloco. Desse modo, ao invés de sortear observações individuais, o método sorteia blocos de tamanho fixo, garantindo que a estrutura autocorrelacionada dos dados seja mantida nas amostras reamostradas (ver Figura 3.1).



**Figura 3.1:** Ilustração esquemática do método de bootstrap em blocos móveis (Moving Block Bootstrap – MBB). Na figura, uma série temporal com comprimento  $n = 18$  e tamanho de bloco  $\ell = 4$  é apresentada. Note que existem  $n - \ell + 1 = 15$  possíveis blocos a serem sorteados. O procedimento seleciona aleatoriamente  $\lfloor n/\ell \rfloor + 2$  blocos e, adicionalmente, sorteia aleatoriamente um ponto inicial para a geração da nova série bootstrap dentro do primeiro bloco selecionado. *Fonte: adaptado de Petropoulos, Hyndman et al. (2018).*

Recentemente, uma abordagem eficaz foi proposta por Bergmeir et al. (2016), na qual o método Bagging é combinado com técnicas adicionais de pré-processamento dos dados, incluindo a transformação de Box–Cox (Box e Cox, 1964) e a decomposição STL (*Seasonal and Trend decomposition using Loess*) (Cleveland, Cleveland et al., 1990). O procedimento geral pode ser resumido nas seguintes etapas:

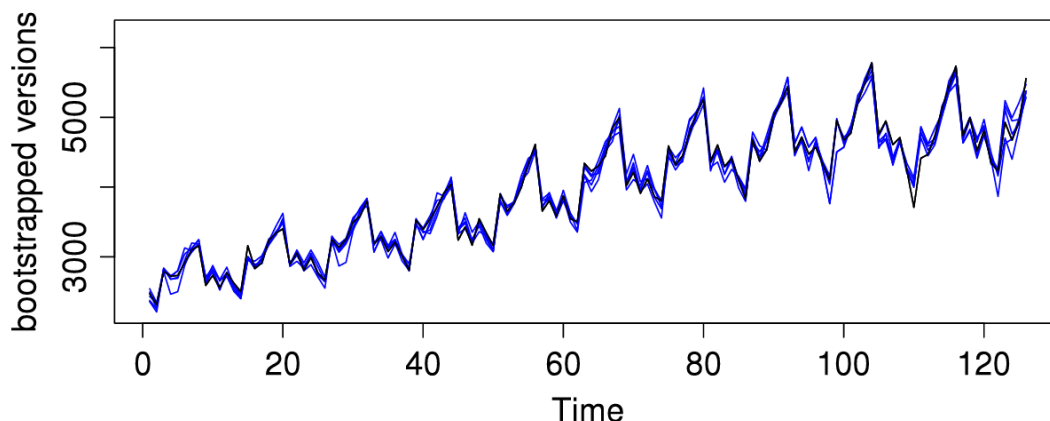
1. **Transformação de Box–Cox:** os dados originais são primeiramente submetidos à transformação de Box–Cox, com o objetivo de estabilizar a variância das observações, mitigando o problema da heterocedasticidade frequentemente presente em séries temporais reais.
2. **Decomposição STL:** após a transformação, a série é decomposta em componentes de tendência, sazonalidade e resíduo. A decomposição STL (*Seasonal and Trend decomposition using Loess*) utiliza suavizações locais (*loess*) e é amplamente empregada por sua flexibilidade e robustez na separação de componentes sazonais e de tendência, mesmo em séries com variações não lineares ao longo do tempo. Caso a série não apresente componente sazonal, utiliza-se a decomposição baseada somente

em suavizações locais (*Loess decomposition*) (Cleveland, Grosse et al., 1992).

3. **Bootstrap dos resíduos (MBB):** o resíduo obtido da decomposição é reamostrado utilizando-se o MBB, gerando novas séries que preservam a estrutura autocorrelacionada original.
  
4. **Reconstrução das séries bootstrap:** as séries são reconstruídas combinando-se os resíduos reamostrados com as componentes de tendência e sazonalidade previamente obtidas. A inversa da transformação de Box–Cox é então aplicada para retornar os dados à sua escala original.

Cada série temporal reconstruída por meio desse processo fornece uma nova versão da série original, permitindo o ajuste independente de modelos individuais em cada amostra bootstrap. Por fim, as previsões obtidas desses modelos individuais são agregadas, geralmente por meio da média ou mediana, produzindo previsões finais mais robustas e menos sensíveis a perturbações nos dados originais. Esse procedimento adaptado de Bagging tem mostrado resultados promissores, especialmente quando aplicado em métodos tradicionais como o alisamento exponencial (Bergmeir et al., 2016; Petropoulos, Hyndman et al., 2018). Em particular, estudos indicam que essas adaptações reduzem a variância das previsões e melhoram consistentemente a acurácia em diversas frequências temporais (mensal, trimestral e anual) (Petropoulos, Hyndman et al., 2018).

A Figura 3.2 mostra o resultado ao aplicar o método bootstrap adaptado, de Bergmeir et al. (2016), na série M495 do bando de dados da M3 (Makridakis e Hibon, 2000). Nota-se que as séries geradas preservam adequadamente o comportamento geral da série original.



**Figura 3.2:** Versões *bootstrap* (em azul) da série temporal original M495 (em preto). São apresentadas cinco séries geradas pelo método adaptado, no qual resíduos são primeiramente reamostrados por meio do *Moving Block Bootstrap* (MBB), para em seguida serem recombinados com as componentes de tendência e sazonalidade previamente obtidas pela decomposição STL. Por fim, aplica-se a inversa da transformação de Box–Cox para obter as séries na escala original. *Fonte: adaptado de Bergmeir et al. (2016).*

### 3.3 Limitações

Apesar das vantagens e adaptações descritas anteriormente, o método Bagging apresenta limitações importantes. Em particular, por ser uma técnica projetada para reduzir a variância de estimadores instáveis, ele não é capaz de solucionar problemas relacionados ao viés estrutural de modelos inadequados para a estrutura dos dados (Breiman, 1996; Hastie et al., 2009).

Se o estimador base escolhido não consegue representar corretamente as relações presentes nos dados, combinar previsões obtidas por Bagging não corrigirá essa limitação fundamental (James et al., 2021). Em outras palavras, a agregação via Bagging não resolve problemas relacionados à falta de flexibilidade do modelo base. Portanto, embora o Bagging seja altamente eficaz na redução da variabilidade das previsões, é importante reconhecer que sua utilização não supera limitações intrínsecas relacionadas à escolha do modelo base (Zhou, 2012).

## Capítulo 4

# Modelo Theta Otimizado Sazonal

Fiorucci et al. (2016) mostram que o Modelo SES-d é equivalente ao Modelo Theta Otimizado (OTM), sob certas condições. Por outro lado, Winters (1960) propõe uma modificação para lidar com sazonalidade no método linear de Holt, que em um caso particular faz previsão da mesma forma que o SES-d. Baseado nisso, apresentamos o *Modelo Theta Otimizado Sazonal* como uma extensão do OTM para lidar internamente com sazonalidade. Em seguida propomos uma versão dinâmica para esse modelo.

O OTM é dado pelas equações:

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, \\ \mu_t &= \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_n \right\}, \\ \ell_t &= \alpha y_t + (1 - \alpha) \ell_{t-1}.\end{aligned}$$

Quando tomamos  $J_{t-1} = \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_n \right\}$ , as equações acima

se tornam:

$$\begin{aligned} y_t &= \ell_{t-1} + J_{t-1} + \varepsilon_t, \\ \ell_t &= \alpha y_t + (1 - \alpha) \ell_{t-1}, \\ J_t &= (1 - \alpha) J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n, \end{aligned}$$

pois,  $J_t = \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^t A_n + \left[ \frac{1 - (1 - \alpha)^{t+1}}{\alpha} \right] B_n \right\}$ , então:

$$\begin{aligned} (1 - \alpha) J_{t-1} &= \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_n \right\} \\ &= \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^t A_n + \left[ \frac{(1 - \alpha) - (1 - \alpha)^{t+1}}{\alpha} \right] B_n \right\} \\ &= \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^t A_n + \left[ \frac{1 - (1 - \alpha)^{t+1}}{\alpha} \right] B_n - B_n \right\}, \end{aligned}$$

o que implica que  $J_t = (1 - \alpha) J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n$ . Seja  $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n$  — para que  $y_t = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \varepsilon_t$  —, então  $\ell_t = \ell_t^{**} - J_t + \left(1 - \frac{1}{\theta}\right) B_n$ . Por outro lado,  $\ell_t = \alpha y_t + (1 - \alpha) \ell_{t-1}$ . Assim:

$$\ell_t^{**} - J_t + \left(1 - \frac{1}{\theta}\right) B_n = \alpha \left[ \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \varepsilon_t \right] + (1 - \alpha) \left[ \ell_{t-1}^{**} - J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n \right],$$

o que implica que

$$\begin{aligned} \ell_t^{**} &= \alpha \varepsilon_t + -(1 - \alpha) J_{t-1} + \ell_{t-1}^{**} + J_t \\ &= \alpha \varepsilon_t + \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n. \end{aligned}$$

Em suma, conseguimos escrever o  $\text{OTM}(\ell_0, \alpha, \theta)$  usando

$$y_t = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \varepsilon_t, \quad (4.1)$$

$$\ell_t^{**} = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \alpha \varepsilon_t, \quad (4.2)$$

que por sua vez, são as mesmas equações do Modelo, que pode ser visto em (2.9) e (2.10),  $\text{SES-d}(\ell_0^{**}, \alpha, b)$  quando  $b = \left(1 - \frac{1}{\theta}\right) B_n$ . Essa relação de equivalência do SES-d com o OTM é apresentada como um teorema em Fiorucci et al. (2016) e, de fato, ela acontece se  $b = \left(1 - \frac{1}{\theta}\right) B_n$  e  $\ell_o^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right) A_n$ . Toda construção acima é uma reedição desse teorema.

Esse resultado mostra que para adicionar um termo sazonal no Modelo Theta Otimizado, não precisamos fazer isso diretamente nas equações originais do modelo. Podemos fazer isso no SES-d equivalente, em (4.1) e (4.2), e depois colocar a restrição  $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n$ .

Para isso, vamos utilizar da estratégia, já conhecida e validada, que o Modelo de Holt-Winters usa para adicionar o termo sazonal no Modelo de Holt, visto que o SES-d é um caso particular do mesmo. Isso implica em duas abordagens diferentes: a inclusão baseada no Holt-Winters aditivo ou baseada no Holt-Winters multiplicativo.

#### 4.1 Inclusão do termo sazonal de modo aditivo

Começamos com o modelo  $\text{SES-d}(\ell_0^{**}, \alpha, b)$  equivalente ao  $\text{OTM}(\ell_0, \alpha, \theta)$ :

$$y_t = \ell_{t-1}^{**} + b + \varepsilon_t \quad (4.3)$$

$$\ell_t^{**} = \ell_{t-1}^{**} + b + \alpha \varepsilon_t, \quad (4.4)$$

em que a relação de equivalência é baseada em  $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n$ . Nesse caso  $b = \left(1 - \frac{1}{\theta}\right) B_n$ , como foi mostrado anteriormente.

Ao incluir o termo sazonal de forma aditiva no modelo acima — da mesma forma que é feita no Modelo Holt-Winters aditivo, partindo do Modelo de Holt —, ele se torna:

$$y_t = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + s_{t-m} + \varepsilon_t \quad (4.5)$$

$$\ell_t^{**} = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \alpha \varepsilon_t, \quad (4.6)$$

$$s_t = s_{t-m} + \gamma \varepsilon_t. \quad (4.7)$$

Aplicando a restrição  $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n$ , chegamos na expressão para o  $y_t$ . Ela segue de:

$$\begin{aligned} y_t &= \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + s_{t-m} + \varepsilon_t \\ &= \left\{ \ell_{t-1} + J_{t-1} - \left(1 - \frac{1}{\theta}\right) B_n \right\} + \left(1 - \frac{1}{\theta}\right) B_n + s_{t-m} + \varepsilon_t \\ &= \{ \ell_{t-1} + J_{t-1} + s_{t-m} \} + \varepsilon_t = \mu_t + \varepsilon_t. \end{aligned}$$

Fazemos a mesma coisa para chegarmos na expressão para o termo de *level*. Ele segue de:

$$\begin{aligned} \ell_t^{**} &= \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \alpha \varepsilon_t \\ &= \left\{ \ell_{t-1} + J_{t-1} - \left(1 - \frac{1}{\theta}\right) B_n \right\} + \left(1 - \frac{1}{\theta}\right) B_n + \alpha \varepsilon_t \\ &= \ell_{t-1} + J_{t-1} + \alpha (y_t - \ell_{t-1} - J_{t-1} - s_{t-m}) \\ &= \alpha (y_t - s_{t-m}) + (1 - \alpha) (\ell_{t-1} + J_{t-1}), \end{aligned}$$

o que equivale a:

$$\begin{aligned} \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n &= \alpha (y_t - s_{t-m}) + (1 - \alpha) (\ell_{t-1} + J_{t-1}) \\ \Rightarrow \ell_t &= \alpha (y_t - s_{t-m}) + (1 - \alpha) (\ell_{t-1} + J_{t-1}) - J_t + \left(1 - \frac{1}{\theta}\right) B_n, \end{aligned}$$

como  $J_t = (1-\alpha)J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n$ , então, o *level* pode ser escrito como  $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ . Por fim, a componente sazonal é escrito como:

$$\begin{aligned} s_t &= s_{t-m} + \gamma\varepsilon_t = s_{t-m} + \gamma(Y_t - \ell_{t-1} - J_{t-1} - s_{t-m}) \\ &= \gamma(y_t - \ell_{t-1} - J_{t-1}) + (1-\gamma)s_{t-m}. \end{aligned}$$

Então, apresentamos uma versão sazonal aditiva para o OTM, na forma:

$$Y_t = \mu_t + \varepsilon_t \quad (4.8)$$

$$\mu_t = \ell_{t-1} + \mathcal{J} + s_{t-m} \quad (4.9)$$

$$\ell_t = \alpha(Y_t - s_{t-m}) + (1-\alpha)\ell_{t-1} \quad (4.10)$$

$$s_t = \gamma(Y_t - \ell_{t-1} - \mathcal{J}) + (1-\gamma)s_{t-m} \quad (4.11)$$

em que  $\mathcal{J} = \left(1 - \frac{1}{\theta}\right) \left\{ (1-\alpha)^{t-1} A_n + \left[ \frac{1-(1-\alpha)^t}{\alpha} \right] B_n \right\}$ ;  $\alpha, \gamma \in (0, 1)$  parâmetros de alisamento;  $\ell_0, s_{1-m}, \dots, s_0 \in \mathbb{R}$  parâmetros de inicialização; e  $\theta \in [1, \infty)$ . Esse modelo será chamado de *Modelo Theta Otimizado Sazonal aditivo* ou SOTM<sub>a</sub>. Veja que quando  $\gamma = 0$  e  $(s_{1-m}, \dots, s_0)^\top = \vec{0}$ , o SOTM<sub>a</sub>( $\ell_0, s_{1-m}, \dots, s_0, \alpha, \gamma, \theta$ ) é equivalente ao OTM( $\ell_0, \alpha, \theta$ ).

**Teorema 1.** *O modelo HW<sub>a</sub>( $\ell_0^{**}, b_0, s_{1-m}, \dots, s_0, \alpha, \beta, \gamma$ ), com  $\ell_0^{**}, b_0 \in \mathbb{R}$ ,  $s_{1-m}, \dots, s_0 \in \mathbb{R}$  e  $\alpha, \beta, \gamma \in (0, 1)$ , é equivalente ao modelo SOTM<sub>a</sub>( $\ell_0, s_{1-m}, \dots, s_0, \alpha, \gamma, \theta$ ), com  $\ell_0 \in \mathbb{R}$ ,  $s_{1-m}, \dots, s_0 \in \mathbb{R}$ ,  $\alpha, \gamma \in (0, 1)$  e  $\theta \geq 1$ , se*

$$\ell_0^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right) A_n \quad e \quad b_t = b_0 = \left(1 - \frac{1}{\theta}\right) B_n.$$

**Demonstração.** *Seja*

$$J_{t-1} = \left(1 - \frac{1}{\theta}\right) \left\{ (1-\alpha)^{t-1} A_n + \left[ \frac{1-(1-\alpha)^t}{\alpha} \right] B_n \right\}.$$

O SOTM<sub>a</sub> pode ser reescrito como:

$$\begin{aligned} Y_t &= \ell_{t-1} + J_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t &= \alpha(Y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}, \\ s_t &= \gamma(Y_t - \ell_{t-1} - J_{t-1}) + (1 - \gamma)s_{t-m}, \\ J_t &= (1 - \alpha)J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n, \end{aligned}$$

em que  $J_0 = \left(1 - \frac{1}{\theta}\right) (A_n + B_n)$ . Assim, ao tomar  $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n$ , podemos reescrever as equações anteriores como:

$$\begin{aligned} Y_t &= \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + s_{t-m} + \varepsilon_t, \\ \ell_t^{**} &= \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \alpha\varepsilon_t, \\ s_t &= s_{t-m} + \gamma\varepsilon_t, \end{aligned}$$

que correspondem ao modelo HW<sub>a</sub> com  $\ell_0^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right) A_n$  e  $b_t = b_0 = \left(1 - \frac{1}{\theta}\right) B_n$ , como se queria demonstrar.  $\square$

Conforme o Teorema 1, podemos concluir que a previsão pontual do SOTM<sub>a</sub> segue a forma que é herdada do HW<sub>a</sub>:

$$\begin{aligned} \hat{Y}_{n+h|n} &= E[Y_{n+h} \mid Y_1, \dots, Y_n] \\ &= \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + \left[ \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] B_n \right\} \\ &\quad - \left(1 - \frac{1}{\theta}\right) B_n + h \left(1 - \frac{1}{\theta}\right) B_n + s_{n-m+h_m^+} \\ &= \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + B_n \left[ (h - 1) + \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] \right\} + s_{n-m+h_m^+}, \end{aligned}$$

em que  $h_m^+ = [(h - 1) \bmod m] + 1$ . O mesmo vale para a variância condicional. Assim:

$$\begin{aligned} \text{Var}[Y_{n+h} \mid Y_1, \dots, Y_n] &= \left[ 1 + (h - 1)\alpha^2 + (k - 1)\gamma^2 + 2\gamma(k - 1)\alpha \right] \sigma^2 \\ &= \left[ 1 + (h - 1)\alpha^2 + (k - 1)(\gamma^2 + 2 + \gamma\alpha) \right] \sigma^2, \end{aligned}$$

em que  $k$  é o menor número inteiro positivo tal que  $n - km + h \leq n$ .

## 4.2 Inclusão de modo multiplicativo

Novamente, começamos com o modelo  $\text{SES-d}(\ell_0^{**}, \alpha, b)$  equivalente ao  $\text{OTM}(\ell_0, \alpha, \theta)$  (4.3) e (4.4). Ao incluir o termo sazonal de forma multiplicativa — da mesma forma que é feita no Modelo Holt-Winters multiplicativo, partindo do Modelo de Holt —, ele se torna:

$$y_t = \left[ \ell_{t-1}^{**} + \left( 1 - \frac{1}{\theta} \right) B_n \right] s_{t-m} + \varepsilon_t \quad (4.12)$$

$$\ell_t^{**} = \ell_{t-1}^{**} + \left( 1 - \frac{1}{\theta} \right) B_n + \frac{\alpha \varepsilon_t}{s_{t-m}}, \quad (4.13)$$

$$s_t = s_{t-m} + \frac{\gamma \varepsilon_t}{\ell_{t-1}^{**} + \left( 1 - \frac{1}{\theta} \right) B_n}. \quad (4.14)$$

Aplicando a restrição  $\ell_t^{**} = \ell_t + J_t - \left( 1 - \frac{1}{\theta} \right) B_n$ , chegamos na expressão para o  $y_t$ . Ela segue de:

$$\begin{aligned} y_t &= \left[ \ell_{t-1}^{**} + \left( 1 - \frac{1}{\theta} \right) B_n \right] s_{t-m} + \varepsilon_t \\ &= \left\{ \ell_{t-1} + J_{t-1} - \left( 1 - \frac{1}{\theta} \right) B_n + \left( 1 - \frac{1}{\theta} \right) B_n \right\} s_{t-m} + \varepsilon_t \\ &= (\ell_{t-1} + J_{t-1}) s_{t-m} + \varepsilon_t. \end{aligned}$$

Fazemos a mesma coisa para chegarmos na expressão para o termo de *level*. Ele segue

de:

$$\begin{aligned}
\ell_t^{**} &= \ell_{t-1}^* + \left(1 - \frac{1}{\theta}\right) B_n + \frac{\alpha \varepsilon_t}{s_{t-m}} \\
&= \left\{ \ell_{t-1} + J_{t-1} - \left(1 - \frac{1}{\theta}\right) B_n \right\} + \left(1 - \frac{1}{\theta}\right) B_n + \frac{\alpha \varepsilon_t}{s_{t-m}} \\
&= \ell_{t-1} + J_{t-1} + \frac{\alpha \varepsilon_t}{s_{t-m}} \\
&= \frac{\alpha Y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + J_{t-1}),
\end{aligned}$$

o que equivale a:

$$\begin{aligned}
\ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n &= \frac{\alpha Y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + J_{t-1}) \\
\Rightarrow \ell_t &= \frac{\alpha Y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + J_{t-1}) + \left(1 - \frac{1}{\theta}\right) B_n - J_t,
\end{aligned}$$

como  $J_t = (1 - \alpha)J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n$ , então, o *level* pode ser escrito como  $\ell_t = \frac{\alpha Y_t}{s_{t-m}} + (1 - \alpha)\ell_{t-1}$ . Por fim e após algumas manipulações algébricas, a componente sazonal segue de:

$$\begin{aligned}
s_t &= s_{t-m} + \frac{\gamma \varepsilon_t}{\ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n} \\
&= \frac{\gamma Y_t}{\ell_{t-1} + J_{t-1}} + (1 - \gamma)s_{t-m}.
\end{aligned}$$

Então, apresentamos uma versão sazonal multiplicativa para o OTM, na forma:

$$Y_t = \mu_t + \varepsilon_t \quad (4.15)$$

$$\mu_t = (\ell_{t-1} + \mathcal{J})s_{t-m} \quad (4.16)$$

$$\ell_t = \frac{\alpha Y_t}{s_{t-m}} + (1 - \alpha)\ell_{t-1} \quad (4.17)$$

$$s_t = \frac{\gamma Y_t}{\ell_{t-1} + \mathcal{J}} + (1 - \gamma)s_{t-m} \quad (4.18)$$

em que  $\mathcal{J} = \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_n \right\}$ ;  $\alpha, \gamma \in (0, 1)$  parâmetros de alisamento;  $\ell_0 \in \mathbb{R}$  e  $s_{1-m}, \dots, s_0 \in \mathbb{R} \setminus \{0\}$  parâmetros de inicialização; e  $\theta \in [1, \infty)$ . Esse modelo será chamado de *Modelo Theta Otimizado Sazonal multiplicativo* ou SOTM<sub>m</sub>. Veja que quando  $\gamma = 0$  e  $(s_{1-m}, \dots, s_0)^\top = \vec{1}$ , o SOTM<sub>m</sub>( $\ell_0, s_{1-m}, \dots, s_0, \alpha, \gamma, \theta$ ) é equivalente ao OTM( $\ell_0, \alpha, \theta$ ).

**Teorema 2.** *O modelo HW<sub>m</sub>( $\ell_0^{**}, b_0, s_{1-m}, \dots, s_0, \alpha, \beta, \gamma$ ), com  $\ell_0^{**}, b_0 \in \mathbb{R}$ ,  $s_{1-m}, \dots, s_0 \in \mathbb{R} \setminus \{0\}$  e  $\alpha, \beta, \gamma \in (0, 1)$ , é equivalente ao modelo SOTM<sub>m</sub>( $\ell_0, s_{1-m}, \dots, s_0, \alpha, \gamma, \theta$ ), com  $\ell_0 \in \mathbb{R}$ ,  $s_{1-m}, \dots, s_0 \in \mathbb{R} \setminus \{0\}$ ,  $\alpha, \gamma \in (0, 1)$  e  $\theta \geq 1$ , se*

$$\ell_0^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right) A_n \quad e \quad b_t = b_0 = \left(1 - \frac{1}{\theta}\right) B_n.$$

**Demonstração.** *Seja*

$$J_{t-1} = \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_n \right\}.$$

O SOTM<sub>m</sub> pode ser reescrito como:

$$\begin{aligned} Y_t &= \ell_{t-1} + J_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t &= \alpha(Y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}, \\ s_t &= \gamma(Y_t - \ell_{t-1} - J_{t-1}) + (1 - \gamma)s_{t-m}, \\ J_t &= (1 - \alpha)J_{t-1} + \left(1 - \frac{1}{\theta}\right) B_n, \end{aligned}$$

em que  $J_0 = \left(1 - \frac{1}{\theta}\right) (A_n + B_n)$ . Assim, ao tomar  $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right) B_n$ , podemos

reescrever as equações anteriores como:

$$\begin{aligned} Y_t &= \left[ \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n \right] s_{t-m} + \varepsilon_t, \\ \ell_t^{**} &= \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n + \frac{\alpha \varepsilon_t}{s_{t-m}}, \\ s_t &= s_{t-m} + \frac{\gamma \varepsilon_t}{\ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right) B_n}, \end{aligned}$$

que correspondem ao modelo  $\text{HW}_m$  com  $\ell_0^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right) A_n$  e  $b_t = b_0 = \left(1 - \frac{1}{\theta}\right) B_n$ , como se queria demonstrar.  $\square$

Conforme o Teorema 2, podemos concluir que a previsão pontual do  $\text{SOTM}_m$  segue a forma que é herdada do  $\text{HW}_m$ :

$$\begin{aligned} \hat{Y}_{n+h|n} &= E[Y_{n+h} \mid Y_1, \dots, Y_n] \\ &= \left\{ \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + \left[ \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] B_n \right\} + (h - 1) \left(1 - \frac{1}{\theta}\right) B_n \right\} s_{n-m+h_m^+} \\ &= \left\{ \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + B_n \left[ (h - 1) + \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] \right\} \right\} s_{n-m+h_m^+}, \end{aligned}$$

em que  $h_m^+ = [(h - 1) \bmod m] + 1$ . A variância condicional ainda não é conhecida, mas ainda é possível que ela seja estimada via reamostragem (*bootstrap*) a partir das simulações de  $Y_{n+h}$ , considerando o modelo ajustado.

### 4.3 Estimação

Pelo fato de que tanto o  $\text{SOTM}_a$ , quanto o  $\text{SOTM}_m$  serem Modelos de Espaço de Estado — como verificado nos Teoremas 1 e 2 —, a estimação dos parâmetros segue conforme apresentada na Seção 2.2.4. Ou seja, sob suposição de normalidade dos erros:

$$\hat{\vec{x}}_0, \hat{\vec{\theta}}, \hat{\theta} = \arg \min_{\vec{x}_0, \vec{\theta}, \theta} n \log \left( \sum_{t=1}^n \varepsilon_t^2 \right),$$

em que  $\varepsilon_t = Y_t - \mu_t$ , pelo fato de que nos dois modelos os erros são aditivos (característica herdada do OTM). Os parâmetros são divididos em inicialização  $\vec{x}_0 = (\ell_0, s_{1-m}, \dots, s_0)^\top$ , alisamento  $\vec{\theta} = (\alpha, \gamma)^\top$  e  $\theta$ . O logaritmo da verossimilhança é resgatado de (2.16).

$$\log L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2.$$

Como comentado na Seção 2.2.4, as estimativas só podem ser alcançadas via uma minimização de forma numérica, o que implica na necessidade de darmos um chute inicial para esses algoritmos.

Um chute intuitivo para  $\theta$  é 2, o que remete à relação do OTM com o método Theta, para séries sem sazonalidade, descrita em Fiorucci et al. (2016), em que o modelo replica as previsões do método. Visto que ainda no SOTM conseguimos voltar para o OTM, mantemos o chute inicial  $\theta = 2$ . Para  $\vec{\theta}$ , Hyndman, Koehler, Ord et al. (2008) recomendam  $(0.1, 0.01)^\top$ , mas não precisa ser assim. Para a inicialização da componente de sazonalidade  $(s_{1-m}, \dots, s_0)^\top$ , basta seguir a heurística apresentada em 2.2.4 para calculo do chute inicial para esse vetor de tamanho  $m$ . Como o  $\ell_0$  parte da restrição  $\ell_0^{**} = \ell_0 + J_0 - \left(1 - \frac{1}{\theta}\right) B_n$  — imposta ao modelo HW<sub>a</sub> ou HW<sub>m</sub> para garantirmos a equivalência com o SOTM correspondente —, e como usualmente  $\ell_0^{**} = Y_1$ , concluímos que um chute inicial adequado é  $\ell_0 = \frac{Y_1}{2}$ .

A estimação de  $\sigma^2$  segue conforme em 2.17:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2.$$

#### 4.4 Versões dinâmicas

Com base na construção da versão dinâmica para o OTM (Fiorucci et al., 2016), seguimos a mesma lógica para não considerarmos os coeficientes  $A_n$  e  $B_n$  fixos no SOTM<sub>a</sub> e, também, no SOTM<sub>m</sub>. Assim, propomos uma versão dinâmica para cada um deles, em que a

atualização desses coeficientes — agora,  $A_t$  e  $B_t$  — do estado  $t$  para  $t + 1$ , utilizará somente as informações anteriores disponíveis, ou seja, apenas os valores  $Y_1, \dots, Y_t$ .

Nesse cálculo, utilizamos  $Y_t^*$  (série sem sazonalidade) em vez de  $Y_t$  para assegurar que os coeficientes calculados capturem exclusivamente a componente de crescimento subjacente, removendo assim a influência da sazonalidade, que iria contaminar e introduzir distorções nas estimativas dinâmicas.

Com base no SOTM<sub>a</sub>, apresentamos o *Modelo Theta Otimizado Sazonal Dinâmico aditivo* ou SOTM-D<sub>a</sub>:

$$Y_t = \mu_t + \varepsilon_t$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_{t-1} + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_{t-1} \right\} + s_{t-m}$$

$$\ell_t = \alpha (Y_t - s_{t-m}) + (1 - \alpha) \ell_{t-1}$$

$$s_t = \gamma \left\{ Y_t - \ell_{t-1} - \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_{t-1} + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_{t-1} \right\} \right\} + (1 - \gamma) s_{t-m}$$

A sua previsão para 1 passo a frente de  $n$  é dada por:

$$\begin{aligned} \hat{Y}_{n+h|n} &= E[Y_{n+h} \mid Y_1, \dots, Y_n] \\ &= \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + B_n \left[ \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] \right\} + s_{n-m+h_m^+}, \end{aligned}$$

em que  $h_m^+ = [(h - 1) \bmod m] + 1$ .

A partir do SOTM<sub>m</sub>, apresentamos o *Modelo Theta Otimizado Sazonal Dinâmico mul-*

*tiplicativo* ou SOTM-D<sub>m</sub>:

$$\begin{aligned}
Y_t &= \mu_t + \varepsilon_t \\
\mu_t &= \left\{ \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_{t-1} + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_{t-1} \right\} \right\} s_{t-m} \\
\ell_t &= \frac{\alpha Y_t}{s_{t-m}} + (1 - \alpha) \ell_{t-1} \\
s_t &= \frac{\gamma Y_t}{\ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_{t-1} + \left[ \frac{1 - (1 - \alpha)^t}{\alpha} \right] B_{t-1} \right\}} + (1 - \gamma) s_{t-m}
\end{aligned}$$

A sua previsão para 1 passo a frente de  $n$  é dada por:

$$\begin{aligned}
\hat{Y}_{n+1|n} &= E[Y_{n+1} \mid Y_1, \dots, Y_n] \\
&= \left\{ \ell_n + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^n A_n + B_n \left[ \frac{1 - (1 - \alpha)^{n+1}}{\alpha} \right] \right\} \right\} s_{n-m+h_m^+},
\end{aligned}$$

em que  $h_m^+ = [(h - 1) \bmod m] + 1$ .

Para os dois sistemas dinâmicos acima, a atualização dos coeficientes  $A_t$  e  $B_t$  segue a estrutura usada em Fiorucci et al. (2016):

$$\begin{aligned}
A_t &= \bar{Y}_t - \frac{t+1}{2} B_t, \\
B_t &= \frac{1}{t+1} \left[ (t-2) B_{t-1} + \frac{6}{t} (Y_t^* - \bar{Y}_{t-1}) \right], \\
\bar{Y}_t &= \frac{1}{t} \left[ (t-1) \bar{Y}_{t-1} + Y_t^* \right], \\
Y_t^* &= I_a (Y_t - s_{t-m}) + (1 - I_a) \frac{Y_t}{s_{t-m}},
\end{aligned}$$

em que  $I_a$  é uma variável indicadora que vale 1 quando estivermos usando o SOTM-D<sub>a</sub> e 0 quando estivermos usando o SOTM-D<sub>m</sub>. As condições iniciais são  $A_0 = B_0 = B_1 = \bar{Y}_0 = 0$ .

Quando o horizonte for maior ( $h \geq 2$ ), as previsões são obtidas recursivamente substituindo os valores futuros  $Y_{n+1}, \dots, Y_{n+h-1}$  por suas respectivas previsões. É difícil de

escrever de forma analítica a variância condicional  $Var[Y_{n+h} \mid Y_1, \dots, Y_n]$  para esses dois últimos modelos, principalmente para o SOTM-D<sub>m</sub>. Ainda assim, é possível que ela seja estimada via reamostragem (*bootstrap*) a partir das simulações de  $Y_{n+h}$ , considerando o modelo ajustado.

Os parâmetros dos modelos SOTM são essencialmente os mesmo de suas versões dinâmicas e a estimação é feita da mesma forma que anteriormente, com um pequeno detalhe:

$$\hat{\vec{x}}_0, \hat{\vec{\theta}}, \hat{\theta} = \arg \min_{\vec{x}_0, \vec{\theta}, \theta} \sum_{t=3}^n (Y_t - \mu_t)^2.$$

Como  $A_t$  e  $B_t$  são coeficientes de uma regressão linear, eles precisam de ao menos dois pontos para serem bem definidos (Fiorucci et al., 2016), por isso a soma começa em  $t = 3$  — o que não acontecia antes. Os chutes iniciais dos parâmetros seguem da mesma forma que foi discutida anteriormente e a estimação para  $\sigma^2$  é dada por:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=3}^n \varepsilon_t^2.$$

# Capítulo 5

## Aplicação na competição M3

Este capítulo tem como principal objetivo apresentar uma aplicação empírica dos modelos desenvolvidos nesta dissertação no contexto da competição internacional de previsão de séries temporais conhecida como *M3-Competition* (Makridakis e Hibon, 2000). Além disso, busca-se investigar o impacto do uso de técnicas de agregação, como o Bagging, aplicadas aos modelos OTM e DOTM, ampliando o escopo de avaliação preditiva. A competição M3 é amplamente reconhecida como uma das mais importantes referências empíricas na área de previsão, sendo amplamente utilizada como benchmark na comparação de diferentes métodos estatísticos e técnicas de aprendizado de máquina.

### 5.1 Contextualização

As chamadas *M competitions*, organizadas por Makridakis e colaboradores desde a década de 1980, foram concebidas com o intuito de avaliar de forma sistemática e imparcial o desempenho preditivo de diferentes abordagens. Ao longo das décadas, essas competições tiveram profundo impacto tanto na literatura acadêmica quanto na prática empresarial, contribuindo para uma compreensão mais clara sobre a eficácia relativa dos métodos em diferentes contextos. A terceira edição, a M3-Competition, destaca-se por sua abrangên-

cia, tanto em termos do número de séries envolvidas quanto da diversidade de frequências, categorias e horizontes de previsão, além de contar com uma ampla base de resultados disponíveis para comparação.

A relevância da M3-Competition também se reflete na abrangência do conjunto de dados empregado: ao todo, são 3.003 séries temporais de diferentes frequências — anuais, trimestrais, mensais e de frequência irregular —, distribuídas em diversas categorias como séries microeconômicas, macroeconômicas, industriais, demográficas e financeiras. Os horizontes de previsão variam conforme a frequência da série, sendo definidos de forma padronizada pelos organizadores da competição. Essa diversidade de estruturas permite avaliar o desempenho dos métodos sob múltiplos contextos, simulando aplicações do mundo real.

A Tabela 5.1 apresenta um resumo da composição do conjunto de dados por frequência e respectivo horizonte de previsão. A categoria denominada *Outras* engloba séries com periodicidades menos comuns, incluindo dados semanais, diários e séries irregulares. Essas séries são tratadas como não sazonais e o pacote `Mcomp` (Kourentzes, 2021), que fornece os dados no R, registra sua frequência como igual a 1. Já a Tabela 5.2 apresenta um resumo mais geral da composição do conjunto de dados, isto é, agregado por nicho e frequência.

**Tabela 5.1:** Descrição do conjunto de dados utilizado na Competição M3

| Frequência   | Horizonte de previsão ( $h$ ) | Número de séries |
|--------------|-------------------------------|------------------|
| Anual        | 6                             | 645              |
| Trimestral   | 8                             | 756              |
| Mensal       | 18                            | 1428             |
| Outras       | 8                             | 174              |
| <b>Total</b> |                               | <b>3003</b>      |

**Tabela 5.2:** Distribuição das séries da competição M3 por nicho e frequência

| Frequência   | Demográfica | Financeira | Industrial | Macroeconômica | Microeconômica | Outras     | Total       |
|--------------|-------------|------------|------------|----------------|----------------|------------|-------------|
| Anual        | 245         | 58         | 102        | 83             | 146            | 11         | 645         |
| Trimestral   | 57          | 76         | 83         | 336            | 204            | 0          | 756         |
| Mensal       | 111         | 145        | 334        | 312            | 474            | 52         | 1428        |
| Outras       | 0           | 29         | 0          | 0              | 4              | 141        | 174         |
| <b>Total</b> | <b>413</b>  | <b>308</b> | <b>519</b> | <b>731</b>     | <b>828</b>     | <b>204</b> | <b>3003</b> |

## 5.2 Desenho da aplicação

Para fins de comparação, selecionamos um conjunto de modelos e métodos amplamente utilizados na literatura de séries temporais e também empregados na competição M3. A Tabela 5.3 apresenta os modelos de referência considerados neste estudo, incluindo alternativas mais simples como o método *Naive* — em que sua previsão é o último valor observado da série, para qualquer horizonte —, modelos clássicos de alisamento exponencial como o *SES* (Brown, 1956) e o *Damped* (Gardner et al., 1985), bem como outras abordagens estatísticas, como o modelo *ARIMA* (Box, Jenkins et al., 2015).

Além desses, também incluímos o método Theta de Assimakopoulos e Nikolopoulos (2000) e suas versões otimizadas OTM e DOTM (Fiorucci et al., 2016), com e sem Bagging, bem como as extensões sazonais propostas nesta dissertação. Essa seleção permite avaliar o desempenho dos modelos desenvolvidos frente a métodos e modelos consolidados na literatura, tanto em termos de acurácia preditiva quanto de robustez computacional.

**Tabela 5.3:** Métodos e modelos de benchmark utilizados

| Método/ Modelo | Referência                           | Seção |
|----------------|--------------------------------------|-------|
| Naive          | -                                    | -     |
| SES            | Brown (1956)                         | 2.2.1 |
| Damped         | Gardner et al. (1985)                | -     |
| ARIMA          | Hyndman e Khandakar (2008)           | -     |
| STheta         | Assimakopoulos e Nikolopoulos (2000) | 2.3   |
| OTM            | Fiorucci et al. (2016)               | 2.4.1 |
| DOTM           | Fiorucci et al. (2016)               | 2.4.2 |

Para a aplicação aos dados da M3-Competition, os modelos (que foram tratados na dissertação) são ajustados com base nos critérios de otimização definidos previamente, utilizando chutes iniciais para os parâmetros, nos respectivos algoritmos de otimização, estabelecidos conforme a Tabela 5.4, que é motivada pelos argumentos nas Seções 2.2.4 e 4.3. Nas versões sazonais do OTM e DOTM, os parâmetros  $s_{1-m}, \dots, s_0$  tiveram chutes iniciais com base no procedimento descrito na Seção 2.2.4, porém para séries não sazonais, eles são fixados como zero ou um (como explicado no Capítulo 4), juntamente com o parâmetro  $\gamma$  — que também, nesse caso, é fixo igual a zero. Adicionalmente, exploramos o uso da técnica de Bagging aplicada aos modelos OTM e DOTM, a fim de investigar possíveis ganhos em estabilidade e desempenho preditivo. Para isso, foram geradas 100 séries bootstrap para cada série original da M3, por meio do método *BLD.MBB* proposto por Bergmeir et al. (2016), conforme discutido no Capítulo 3.

**Tabela 5.4:** Modelos e método Theta considerados na avaliação empírica

| Modelo/ Método      | Equações/ Seção | Valores iniciais para otimização                         |
|---------------------|-----------------|----------------------------------------------------------|
| STheta              | 2.3             | $\ell_0^* = A_{10}, \alpha = 0,5$                        |
| OTM                 | 2.4             | $\ell_0 = Y_1/2, \alpha = 0,5, \theta = 2$               |
| DOTM                | 2.4             | $\ell_0 = Y_1/2, \alpha = 0,5, \theta = 2$               |
| SOTM <sub>a</sub>   | (4.8)–(4.11)    | $\ell_0 = Y_1/2, \alpha = 0,5, \gamma = 0.5, \theta = 2$ |
| SOTM <sub>m</sub>   | (4.15)–(4.18)   | $\ell_0 = Y_1/2, \alpha = 0,5, \gamma = 0.5, \theta = 2$ |
| SOTM-D <sub>a</sub> | 4.4             | $\ell_0 = Y_1/2, \alpha = 0,5, \gamma = 0.5, \theta = 2$ |
| SOTM-D <sub>m</sub> | 4.4             | $\ell_0 = Y_1/2, \alpha = 0,5, \gamma = 0.5, \theta = 2$ |

Os resultados serão apresentados conforme agregação das séries nas diferentes frequências, como visto na Tabela 5.1. Para uma avaliação empírica considerando os diferentes nichos vistos na Tabela 5.2, veja o Apêndice A.

### 5.3 Resultados

A implementação dos modelos utilizados neste estudo foi realizada na linguagem R, por meio de funções desenvolvidas especificamente para esta dissertação, em conjunto com

o pacote [forecTheta 3.0: Forecasting Time Series by Theta Models](#) (acesso em: 13/07/2025). Os modelos OTM e DOTM, bem como suas variantes com Bagging, estão plenamente implementados no pacote, com documentação e funções acessíveis. Já as versões sazonais propostas nesta dissertação foram desenvolvidas com base na mesma estrutura do `forecTheta`, garantindo compatibilidade e facilidade de integração.

Para assegurar a reprodutibilidade dos resultados obtidos com os dados da M3-Competition, disponibilizamos um repositório no GitHub contendo todos os scripts necessários.

O código-fonte está acessível no [repositório no GitHub](#) (acesso em: 13/07/2025).

Para replicar os resultados, siga os passos a seguir:

1. Clone o repositório;
2. Abra o projeto `.Rproj` no RStudio (certifique que está na *branch* chamada `devep_igor`);
3. Execute o comando `devtools::install()` para instalar a versão local (que foi clonada e está na *branch* `devep_igor`) do pacote `forecTheta`;
4. Por fim, execute o script `./benchmarks/m3_mase_smape.R`, que realiza a geração das previsões e o cálculo das métricas. Se quiser replicar as Tabelas [5.7–5.10](#), execute o script `./benchmarks/percentage_improv.R`.

Neste estudo, utilizamos duas métricas de acurácia para avaliar o desempenho dos modelos preditivos: *Symmetric Mean Absolute Percentage Error* — sMAPE e *Mean Absolute Scaled Error* — MASE. A escolha dessas métricas foi motivada pelo objetivo de garantir comparabilidade entre os resultados deste trabalho e os encontrados na literatura (Bergmeir et al., 2016; Spiliotis et al., 2019; Petropoulos, Hyndman et al., 2018), especialmente em Fiorucci et al. (2016). É importante observar, no entanto, que a sMAPE tende a penalizar erros positivos (quando a previsão supera o valor real) de forma mais severa do que erros negativos, com essa discrepância se acentuando à medida que o erro aumenta.

Esse comportamento foi discutido por Goodwin et al. (1999), que demonstraram essa assimetria crescente de penalização.

A sMAPE se baseia no erro percentual absoluto simétrico, calculado da seguinte forma:

$$\text{sAPE} = 200 \frac{|Y_{n+i} - \hat{Y}_{n+i|n}|}{|Y_{n+i}| + |\hat{Y}_{n+i|n}|}, \quad (5.1)$$

A MASE (Hyndman e Koehler, 2006), por sua vez, é baseada em um erro absoluto escalonado, definido como:

$$\text{ASE} = \frac{|Y_{n+i} - \hat{Y}_{n+i|n}|}{(n-m)^{-1} \sum_{t=m+1}^n |Y_t - Y_{t-m}|}, \quad (5.2)$$

em que o denominador corresponde à média das diferenças absolutas sazonais da série observada.

Tanto na sMAPE quanto na MASE, os erros são calculados ponto a ponto; em seguida, computa-se a média simples considerando todos os horizontes de previsão e, por fim, todas as séries. Isso implica que cada erro individual — independentemente de sua origem em uma série anual, trimestral ou mensal — recebe o mesmo peso na média final. Em outras palavras, erros provenientes de séries mais curtas, como as anuais, têm o mesmo impacto no valor médio das métricas que os erros de séries mais longas, como as mensais.

As Tabelas 5.5 e 5.6 apresentam os resultados obtidos pelos diferentes métodos e modelos aplicadas às séries temporais da M3-Competition, considerando as métricas sMAPE e MASE, respectivamente. Em ambas as tabelas, os métodos foram avaliados separadamente com base nos dados originais e nos dados previamente dessazonalizados. Essa dessazonalização é realizada via decomposição clássica multiplicativa, e após a realização das previsões (pelo método ou modelo específico) para essa série com sazonalidade ajustada, há um passo de ressazonalização com a multiplicação dessas previsões pelos índices sazonais correspondentes aos observados no último ciclo sazonal. Os resultados

estão organizados por frequência da série (anual, trimestral, mensal e outras), além de uma média geral indicada na coluna *All*. Essa média geral segue conforme Makridakis e Hibon (2000).

A última coluna de cada tabela informa o tempo total de execução, em minutos, necessário para gerar as previsões de todas as 3.003 séries para cada método ou modelo. Os tempos marcados com asterisco indicam valores extraídos da literatura, em particular de Fiorucci et al. (2016). Os demais valores foram obtidos a partir da execução dos experimentos realizados nesta dissertação, por meio de um computador pessoal com processador Intel(R) Core(TM) i5-7300HQ CPU @ 2.50GHz (4 núcleos, 4 threads), 32 GB de memória RAM, sistema operacional Microsoft Windows 11 Home Single Language (64 bits). Todos os experimentos foram executados com uso de paralelização explícita.

**Tabela 5.5:** Avaliação empírica utilizando a métrica sMAPE (%)

| Modelo/Método          | Anual        | Trimestral  | Mensal       | Outras      | All          | Tempo (min) |
|------------------------|--------------|-------------|--------------|-------------|--------------|-------------|
| Dados originais        |              |             |              |             |              |             |
| Naive                  | 17,88        | 11,32       | 18,18        | 6,30        | 16,58        | 0,31*       |
| SES                    | 17,78        | 10,83       | 16,14        | 6,30        | 15,07        | 0,22*       |
| Damped                 | 17,07        | 10,96       | 16,25        | <b>4,30</b> | 15,02        | 1.05*       |
| ARIMA                  | 17,62        | 9,99        | 15,30        | 4,54        | 14,27        | 23.48*      |
| SOTM <sub>a</sub>      | 16,60        | <b>9,10</b> | 14,59        | 4,85        | 13,54        | 2,08        |
| SOTM <sub>m</sub>      | 16,60        | 9,13        | 13,87        | 4,85        | 13,04        | 2,22        |
| SOTM-D <sub>a</sub>    | <b>15,94</b> | 9,12        | 14,26        | 4,58        | 13,23        | 2,15        |
| SOTM-D <sub>m</sub>    | <b>15,94</b> | 9,16        | 13,75        | 4,58        | 12,89        | 2.28        |
| Dados dessazonalizados |              |             |              |             |              |             |
| Naive                  | 17,88        | 10,02       | 16,76        | 6,30        | 15,38        | 0,62*       |
| SES                    | 17,78        | 9,77        | 14,17        | 6,30        | 13,53        | 0,55*       |
| Damped                 | 17,07        | 9,79        | 13,96        | <b>4,30</b> | 13,24        | 1.52*       |
| ARIMA                  | 17,62        | 9,74        | 15,68        | 4,54        | 14,49        | 21.40*      |
| STheta                 | 16,74        | 9,23        | 13,83        | 4,93        | 13,05        | 1,23        |
| OTM                    | 16,60        | 9,17        | 14,17        | 4,85        | 13,26        | 1,40        |
| DOTM                   | <b>15,94</b> | 9.28        | 13.74        | 4.58        | 12.90        | 1,39        |
| Bagged OTM             | 16,66        | 9,18        | 13,92        | 5,01        | 13,10        | 57,69       |
| Bagged DOTM            | 16,23        | 9,28        | <b>13,45</b> | 4,75        | <b>12,73</b> | 57,30       |

**Tabela 5.6:** Avaliação empírica utilizando a métrica MASE

| Modelo/Método          | Anual       | Trimestral  | Mensal      | Outras      | All         | Tempo (min) |
|------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Dados originais        |             |             |             |             |             |             |
| Naive                  | 3,17        | 1,46        | 1,17        | 3,09        | 1,50        | 0,31*       |
| SES                    | 3,17        | 1,41        | 1,09        | 3,10        | 1,43        | 0,22*       |
| Damped                 | 2,92        | 1,37        | 1,10        | <b>1,81</b> | 1,36        | 1,05*       |
| ARIMA                  | 2,99        | 1,17        | 0,88        | 1,87        | 1,19        | 23,48*      |
| SOTM <sub>a</sub>      | 2,71        | <b>1,10</b> | 0,87        | 2,23        | 1,15        | 2,08        |
| SOTM <sub>m</sub>      | 2,71        | 1,11        | 0,86        | 2,23        | 1,15        | 2,22        |
| SOTM-D <sub>a</sub>    | <b>2,59</b> | 1,11        | 0,86        | 1,94        | <b>1,12</b> | 2,15        |
| SOTM-D <sub>m</sub>    | <b>2,59</b> | 1,11        | 0,86        | 1,94        | <b>1,12</b> | 2,28        |
| Dados dessazonalizados |             |             |             |             |             |             |
| Naive                  | 3,17        | 1,25        | 1,04        | 3,09        | 1,37        | 0,62*       |
| SES                    | 3,17        | 1,24        | 0,93        | 3,10        | 1,29        | 0,55*       |
| Damped                 | 2,92        | 1,17        | 0,88        | <b>1,81</b> | 1,17        | 1,52*       |
| ARIMA                  | 2,99        | 1,15        | 0,89        | 1,87        | 1,19        | 21,40*      |
| STheta                 | 2,77        | 1,12        | 0,86        | 2,28        | 1,16        | 1,23        |
| OTM                    | 2,71        | <b>1,10</b> | 0,86        | 2,23        | 1,14        | 1,40        |
| DOTM                   | <b>2,59</b> | 1,12        | 0,85        | 1,94        | <b>1,12</b> | 1,39        |
| Bagged OTM             | 2,72        | <b>1,10</b> | 0,86        | 2,26        | 1,15        | 57,69       |
| Bagged DOTM            | 2,64        | 1,12        | <b>0,84</b> | 2,00        | <b>1,12</b> | 57,30       |

as Tabelas 5.7–5.10 são geradas a partir da verificação da presença de tendência e sazonalidade nas séries temporais da M3. Para tendência, utilizamos o teste de Mann-Kendall, um teste não paramétrico que avalia a existência de tendência monótona (crescente ou decrescente) sem exigir suposições sobre a distribuição dos dados. Já para sazonalidade, foi utilizada uma abordagem baseada na autocorrelação da série em sua defasagem sazonal. O procedimento envolve o cálculo do correlograma e a verificação se o coeficiente de autocorrelação sazonal excede um limite crítico, indicando padrão sazonal estatisticamente significativo, como detalhado na Seção 2.3. Como a técnica depende da frequência da série, ela só é aplicada quando a frequência é maior ou igual a 3; para séries com frequência menor que 3, considera-se que não há sazonalidade.

**Tabela 5.7:** Melhora percentual do modelo SOTM<sub>m</sub> em relação ao OTM em termos de sMAPE, de acordo com a presença de tendência (os números entre parênteses referem-se aos tamanhos das amostras).

| Frequência | Todas ( $n$ ) | Com tendência ( $n$ ) | Sem tendência ( $n$ ) |
|------------|---------------|-----------------------|-----------------------|
| Anual      | 0,0% (645)    | 0,0% (543)            | 0,0% (102)            |
| Trimestral | 4,17% (756)   | 4,94% (662)           | -1,29% (94)           |
| Mensal     | 2,7% (1428)   | 3,2% (1164)           | 0,46% (264)           |
| Outras     | 0,0% (174)    | 0,0% (166)            | 0,0% (8)              |

**Tabela 5.8:** Melhora percentual do modelo SOTM<sub>m</sub> em relação ao OTM em termos de sMAPE, de acordo com a presença de sazonalidade (os números entre parênteses referem-se aos tamanhos das amostras).

| Frequência | Todas ( $n$ ) | Sazonal ( $n$ ) | Não sazonal ( $n$ ) |
|------------|---------------|-----------------|---------------------|
| Anual      | 0,0% (645)    | –               | 0,0% (645)          |
| Trimestral | 4,17% (756)   | 5,68% (555)     | 0,0% (201)          |
| Mensal     | 2,7% (1428)   | 4,93% (788)     | 0,0% (640)          |
| Outras     | 0,0% (174)    | –               | 0,0% (174)          |

**Tabela 5.9:** Melhora percentual do modelo SOTM-D<sub>m</sub> em relação ao DOTM em termos de sMAPE, de acordo com a presença de tendência (os números entre parênteses referem-se aos tamanhos das amostras).

| Frequência | Todas ( $n$ ) | Com tendência ( $n$ ) | Sem tendência ( $n$ ) |
|------------|---------------|-----------------------|-----------------------|
| Anual      | 0,0% (645)    | 0,0% (543)            | 0,0% (102)            |
| Trimestral | 6,31% (756)   | 7,38% (662)           | -1,27% (94)           |
| Mensal     | 4,28% (1428)  | 5,18% (1164)          | 0,3% (264)            |
| Outras     | 0,0% (174)    | 0,0% (166)            | 0,0% (8)              |

**Tabela 5.10:** Melhora percentual do modelo SOTM-D<sub>m</sub> em relação ao DOTM em termos de sMAPE, de acordo com a presença de sazonalidade (os números entre parênteses referem-se aos tamanhos das amostras).

| Frequência | Todas ( $n$ ) | Sazonal ( $n$ ) | Não sazonal ( $n$ ) |
|------------|---------------|-----------------|---------------------|
| Anual      | 0,0% (645)    | –               | 0,0% (645)          |
| Trimestral | 6,31% (756)   | 8,59% (555)     | 0,0% (201)          |
| Mensal     | 4,28% (1428)  | 7,83% (788)     | 0,0% (640)          |
| Outras     | 0,0% (174)    | –               | 0,0% (174)          |

## 5.4 Resumo dos resultados

### Avaliação das Previsões (sMAPE e MASE)

Os modelos SOTM-D (SOTM-D<sub>a</sub> e SOTM-D<sub>m</sub>) apresentaram consistentemente as menores médias de erro nas séries originais, com destaque para o modelo SOTM-D<sub>m</sub>, que obteve o menor sMAPE geral de **12,89%** e MASE geral de **1,12**. Os métodos ou modelos tradicionais (Naive, SES e ARIMA) tiveram desempenhos inferiores em todas as frequências avaliadas.

Para dados dessazonalizados, os modelos com *Bagging* (Bagged DOTM) tiveram o melhor desempenho geral, alcançando um sMAPE de **12,73%** e um MASE de **1,12**. Apesar disso, o aumento significativo no tempo de processamento (57 minutos contra cerca de 2 minutos dos modelos não-*bagged*) deve ser considerado.

O modelo ARIMA, embora competitivo, apresentou tempos de processamento muito elevados (aproximadamente 21 a 23 minutos) quando comparado aos modelos e método Theta, que ficaram entre 1 a 3 minutos em média.

### Melhora Percentual por Presença de Tendência

- **Modelo SOTM<sub>m</sub> vs OTM:**

- **Trimestral:** Melhora geral de **4,17%**, maior em séries com tendência (**4,94%**), mas negativa (**-1,29%**) sem tendência.
- **Mensal:** Melhora geral de **2,7%**, sendo maior para séries com tendência (**3,2%**) do que sem tendência (**0,46%**).
- Não houve diferença para dados anuais ou em “outros”.

- **Modelo SOTM-D<sub>m</sub> vs DOTM:**

- **Trimestral:** Melhora geral de **6,31%**, mais expressiva em séries com tendência (**7,38%**), mas negativa (**-1,27%**) sem tendência.
- **Mensal:** Melhora geral de **4,28%**, especialmente superior em séries com tendência (**5,18%**), sendo mais tímida sem tendência (**0,3%**).
- Não houve diferença para dados anuais ou em “outros”.

### Melhora Percentual por Presença de Sazonalidade

- **Modelo SOTM<sub>m</sub> vs OTM:**

- **Trimestral:** Melhor desempenho nas séries sazonais (**5,68%**) e não houve diferença em séries não sazonais.
- **Mensal:** Melhora evidente em séries sazonais (**4,93%**) e não houve diferença em séries não sazonais.
- Não houve diferença para dados anuais ou em “outros”.

- **Modelo SOTM-D<sub>m</sub> vs DOTM:**

- **Trimestral:** Maior ganho nas séries sazonais (**8,59%**) e não houve diferença nas séries não sazonais.
- **Mensal:** Melhora significativa nas séries sazonais (**7,83%**) e não houve diferença nas séries não sazonais.
- Não houve diferença para dados anuais ou em “outros”.

### Conclusão Resumida

As versões sazonais dos modelos Theta Otimizado, especialmente com dinâmica multiplicativa (SOTM-D<sub>m</sub>), apresentaram os melhores desempenhos nas métricas sMAPE e MASE, principalmente em séries com presença de tendência e sazonalidade. Como nesse

experimento foram garantidas as condições, descritas no Capítulo 4, de forma que houvesse equivalência entre os modelos Theta Otimizados e suas versões sazonais, o desempenho entre eles foi idêntico para séries não sazonais. Modelos tradicionais como ARIMA, apesar de competitivos nas previsões, foram menos eficientes computacionalmente, enquanto técnicas com *Bagging* proporcionaram melhorias adicionais, porém com elevado custo computacional.

## Capítulo 6

# Considerações Finais e Conclusão

Este trabalho teve como objetivo principal propor uma versão sazonal para o Modelo Theta Otimizado (OTM), visando modelar adequadamente séries temporais com sazonalidade, especialmente aquelas com sazonalidade dinâmica, além de avaliar o uso de *Bagging* para o Modelo Theta Otimizado. As inovações propostas, SOTM e sua versão dinâmica, superam a limitação do OTM, herdada do método Theta, de não modelar a componente de sazonalidade em séries sazonais.

No Capítulo 5, realizamos uma aplicação empírica utilizando o conjunto de dados da competição M3, avaliando a performance preditiva dos modelos propostos e comparando-os com outros modelos tradicionais na literatura de previsão de séries temporais. Utilizando métricas como sMAPE e MASE, observamos que as versões dinâmicas e multiplicativas dos modelos propostos tiveram desempenho superior, destacando-se especialmente o Modelo Theta Otimizado Sazonal Dinâmico Multiplicativo (SOTM dinâmico multiplicativo).

Observamos uma melhora nos resultados preditivos do Modelo Theta Otimizado com a aplicação do Bagging na competição M3. Além disso, propusemos um novo modelo para séries temporais, denominado Modelo Theta Otimizado Sazonal. Os resultados empíricos indicam que essa extensão sazonal apresenta desempenho superior ao do modelo do qual

se originou, superando-o nas previsões realizadas para os dados da competição M3.

## 6.1 Possibilidade de trabalhos futuros

Como sugestões para trabalhos futuros, propomos:

1. Aplicação do *Bagging* para o Modelo Theta Otimizado Sazonal;
2. Inclusão das versões sazonais do Modelo Theta Otimizado no pacote para o R chamado `forecTheta`;
3. Desenvolvimento de um algoritmo de seleção automática para as versões aditiva e multiplicativa do Modelo Theta Otimizado Sazonal, considerando critérios de informação como AIC ou AICc.

A complexidade dessas tarefas está em evoluir [os códigos](#) (acesso em: 13/07/2025) afim de abrir um *pull request* (PR) para o [repositório oficial](#) (acesso em: 13/07/2025) do pacote `forecTheta`. As tarefas 1 e 2 são relativamente simples, basta adaptar o código da função já criada `bagged_twoTL` e refinar a `seasonal_twoTL` — afim de facilitar manutenção. Já a tarefa 3 exige a implementação do cálculo dos determinados critérios de informação. O que também não é muito complexo, já que o logaritmo da verossimilhança está bem desenvolvido na Seção [4.3](#).

# Referências Bibliográficas

- Anderson, B. D. O. e Moore, J. B. (1979). *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall (ver p. 4).
- Aoki, M. (1990). *State Space Modeling of Time Series*. Berlin: Springer-Verlag (ver p. 4).
- Assimakopoulos, V. e Nikolopoulos, K. (2000). “The theta model: a decomposition approach to forecasting”. *International Journal of Forecasting* 16.4. The M3- Competition, pp. 521–530. ISSN: 0169-2070. DOI: [https://doi.org/10.1016/S0169-2070\(00\)00066-2](https://doi.org/10.1016/S0169-2070(00)00066-2). URL: <https://www.sciencedirect.com/science/article/pii/S0169207000000662> (ver pp. 1, 4, 17, 18, 20, 49).
- Assimakopoulos, V. (1995). “A successive filtering technique for identifying long-term trends”. *Journal of Forecasting* 14.1, pp. 35–43. DOI: [10.1002/for.3980140104](https://doi.org/10.1002/for.3980140104). eprint: <https://doi.org/10.1002/for.3980140104>. URL: <https://doi.org/10.1002/for.3980140104> (ver p. 17).
- Athanasopoulos, G., Hyndman, R. J., Song, H. e Wu, D. C. (2011). “The tourism forecasting competition”. *International Journal of Forecasting* 27.3. Special Section 1: Forecasting with Artificial Neural Networks and Computational Intelligence Special Section 2: Tourism Forecasting, pp. 822–844. ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2010.04.009>. URL: <https://www.sciencedirect.com/science/article/pii/S016920701000107X> (ver p. 17).
- Bergmeir, C., Hyndman, R. J. e Benítez, J. M. (2016). “Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation”. *International Jour-*

- nal of Forecasting* 32.2, pp. 303–312. DOI: [10.1016/j.ijforecast.2015.07.002](https://doi.org/10.1016/j.ijforecast.2015.07.002) (ver pp. 2, 30–32, 50, 51).
- Box, G. E. P. e Cox, D. R. (1964). “An analysis of transformations”. *Journal of the Royal Statistical Society: Series B* 26.2, pp. 211–243 (ver p. 30).
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. 5<sup>a</sup> ed. Hoboken, NJ: John Wiley & Sons (ver p. 49).
- Breiman, L. (1996). “Bagging predictors”. *Machine Learning* 24.2, pp. 123–140. DOI: [10.1007/BF00058655](https://doi.org/10.1007/BF00058655) (ver pp. 2, 25, 27, 29, 32).
- Brown, R. (1956). *Exponential Smoothing for Predicting Demand*. Little. URL: [https://books.google.com.br/books?id=Eo\\_rMgEACAAJ](https://books.google.com.br/books?id=Eo_rMgEACAAJ) (ver p. 49).
- Clemen, R. T. (1989). “Combining forecasts: A review and annotated bibliography”. *International Journal of Forecasting* 5.4, pp. 559–583. DOI: [10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5). eprint: [https://doi.org/10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5). URL: [https://doi.org/10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5) (ver p. 17).
- Cleveland, R. B., Cleveland, W. S., McRae, J. E. e Terpenning, I. (1990). “STL: A seasonal-trend decomposition procedure based on loess”. *Journal of Official Statistics* 6.1, pp. 3–73 (ver p. 30).
- Cleveland, W. S., Grosse, E. e Shyu, W. M. (1992). “Local regression models”. Em: *Statistical Models in S*. Ed. por J. M. Chambers e T. J. Hastie. Chapman e Hall/CRC, pp. 309–376 (ver p. 31).
- Dietterich, T. G. (2000). “Ensemble Methods in Machine Learning”. *International Workshop on Multiple Classifier Systems*, pp. 1–15 (ver pp. 25, 28).
- Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F. e Koehler, A. B. (2016). “Models for optimising the theta method and their relationship to state space models”. *International Journal of Forecasting* 32.4, pp. 1151–1161. ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2016.02.005>. URL: <https://www.>

- [sciencedirect.com/science/article/pii/S0169207016300243](https://www.sciencedirect.com/science/article/pii/S0169207016300243) (ver pp. 1, 4, 17, 20, 22–24, 33, 35, 43, 45, 46, 49, 51, 53).
- Gardner, E. S. e McKenzie, E. (1985). “Forecasting Trends in Time Series”. *Management Science* 31.10, pp. 1237–1246. DOI: [10.1287/mnsc.31.10.1237](https://doi.org/10.1287/mnsc.31.10.1237). eprint: <https://doi.org/10.1287/mnsc.31.10.1237>. URL: <https://doi.org/10.1287/mnsc.31.10.1237> (ver p. 49).
- Goodwin, P. e Lawton, R. (1999). “On the asymmetry of the symmetric MAPE”. *International Journal of Forecasting* 15.4, pp. 405–408. ISSN: 0169-2070. DOI: [https://doi.org/10.1016/S0169-2070\(99\)00007-2](https://doi.org/10.1016/S0169-2070(99)00007-2). URL: <https://www.sciencedirect.com/science/article/pii/S0169207099000072> (ver p. 52).
- Hannan, E. J. e Deistler, M. (1988). *The Statistical Theory of Linear Systems*. New York: Wiley (ver p. 4).
- Hastie, T., Tibshirani, R. e Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer (ver pp. 25, 27, 32).
- Hyndman, R. J. e Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. 2nd. Online book: <https://otexts.com/fpp2/>. OTexts (ver p. 1).
- 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* 18.3, pp. 439–454. ISSN: 0169-2070. DOI: [https://doi.org/10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8). URL: <https://www.sciencedirect.com/science/article/pii/S0169207001001108> (ver p. 16).
- Hyndman, R. J. e Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3<sup>a</sup> ed. Melbourne, Australia: OTexts. URL: <https://otexts.com/fpp3> (acesso em 10/07/2025) (ver p. 13).
- Hyndman, R. J. e Billah, B. (2003). “Unmasking the Theta method”. *International Journal of Forecasting* 19.2, pp. 287–290. ISSN: 0169-2070. DOI: [https://doi.org/10.1016/S0169-2070\(03\)00007-2](https://doi.org/10.1016/S0169-2070(03)00007-2).

- 1016/S0169-2070(01)00143-1. URL: <https://www.sciencedirect.com/science/article/pii/S0169207001001431> (ver p. 18).
- Hyndman, R. J. e Khandakar, Y. (2008). “Automatic Time Series Forecasting: The forecast Package for R”. *Journal of Statistical Software* 27.3, pp. 1–22. DOI: [10.18637/jss.v027.i03](https://doi.org/10.18637/jss.v027.i03). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v027i03> (ver p. 49).
- Hyndman, R. J. e Koehler, A. B. (2006). “Another look at measures of forecast accuracy”. *International Journal of Forecasting* 22.4, pp. 679–688. ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2006.03.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0169207006000239> (ver p. 52).
- Hyndman, R. J., Koehler, A. B., Ord, J. K. e Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin, Heidelberg: Springer (ver pp. 4, 6, 15, 16, 43).
- James, G., Witten, D., Hastie, T. e Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer (ver pp. 25, 27, 28, 32).
- Koning, A. J., Franses, P. H., Hibon, M. e Stekler, H. (2005). “The M3 competition: Statistical tests of the results”. *International Journal of Forecasting* 21.3, pp. 397–409. ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2004.10.003>. URL: <https://www.sciencedirect.com/science/article/pii/S0169207004000810> (ver p. 17).
- Kourentzes, N. (2021). *Mcomp: Data from the M-Competitions*. R package version 2.9. URL: <https://CRAN.R-project.org/package=Mcomp> (ver p. 48).
- Kuncheva, L. I. (2014). *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons (ver pp. 25, 27).
- Künsch, H. R. (1989). “The jackknife and the bootstrap for general stationary observations”. *Annals of Statistics* 17.3, pp. 1217–1241. DOI: [10.1214/aos/1176347265](https://doi.org/10.1214/aos/1176347265) (ver p. 29).

- Makridakis, S. e Hibon, M. (2000). “The M3-Competition: results, conclusions and implications”. *International Journal of Forecasting* 16.4. The M3- Competition, pp. 451–476. ISSN: 0169-2070. DOI: [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1). URL: <https://www.sciencedirect.com/science/article/pii/S0169207000000571> (ver pp. 17, 20, 31, 47, 53, 67).
- Makridakis, S. e Winkler, R. L. (1983). “Averages of forecasts: Some empirical results”. *Management Science* 29.9, pp. 987–996. DOI: [10.1287/mnsc.29.9.987](https://doi.org/10.1287/mnsc.29.9.987). eprint: <https://doi.org/10.1287/mnsc.29.9.987>. URL: <https://doi.org/10.1287/mnsc.29.9.987> (ver p. 17).
- Petropoulos, F., Hyndman, R. J. e Bergmeir, C. (2018). “Exploring the sources of uncertainty: Why does bagging for time series forecasting work?” *European Journal of Operational Research* 268.2, pp. 545–554. ISSN: 0377-2217. DOI: <https://doi.org/10.1016/j.ejor.2018.01.045>. URL: <https://www.sciencedirect.com/science/article/pii/S037722171830081X> (ver pp. 2, 30, 31, 51).
- Petropoulos, F., Makridakis, S., Assimakopoulos, V. e Nikolopoulos, K. (2014). “Horses for Courses’ in demand forecasting”. *European Journal of Operational Research* 237.1, pp. 152–163. ISSN: 0377-2217. DOI: <https://doi.org/10.1016/j.ejor.2014.02.036>. URL: <https://www.sciencedirect.com/science/article/pii/S0377221714001714> (ver p. 17).
- Shumway, R. H. e Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer (ver p. 1).
- Spiliotis, E., Assimakopoulos, V. e Nikolopoulos, K. (2019). “Forecasting with a hybrid method utilizing data smoothing, a variation of the Theta method and shrinkage of seasonal factors”. *International Journal of Production Economics* 209. The Proceedings of the 19th International Symposium on Inventories, pp. 92–102. ISSN: 0925-5273. DOI: <https://doi.org/10.1016/j.ijpe.2018.01.020>. URL: <https://www.sciencedirect.com/science/article/pii/S0925527318300501> (ver p. 51).

Surowiecki, J. (2004). *The Wisdom of Crowds*. Doubleday (ver p. 25).

Winters, P. R. (1960). “Forecasting Sales by Exponentially Weighted Moving Averages”. *Management Science* 6.3, pp. 324–342. DOI: [10.1287/mnsc.6.3.324](https://doi.org/10.1287/mnsc.6.3.324). eprint: <https://doi.org/10.1287/mnsc.6.3.324>. URL: <https://doi.org/10.1287/mnsc.6.3.324> (ver pp. 1, 33).

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. CRC Press (ver pp. 25, 27, 32).

# Apêndice A

## Resultados Adicionais

Este apêndice apresenta resultados complementares àqueles discutidos no Capítulo 5, incluindo variações das Tabelas 5.5 e 5.6, com avaliação considerando agregação das séries da M3 (Makridakis e Hibon, 2000) por nicho como visto na Tabela 5.2.

Tabela A.1: Avaliação empírica utilizando a métrica sMAPE (%)

| Modelo/Método          | Demográfica | Financeira   | Industrial   | Macroeconômica | Microeconômica | Outras      | All          | Tempo (min) |
|------------------------|-------------|--------------|--------------|----------------|----------------|-------------|--------------|-------------|
| Dados originais        |             |              |              |                |                |             |              |             |
| SOTM <sub>a</sub>      | 9,78        | 14,39        | 12,91        | 6,36           | 21,79          | 7,86        | 13,54        | 2,13        |
| SOTM <sub>m</sub>      | 9,60        | 14,39        | <b>12,24</b> | 6,36           | 20,63          | 7,86        | 13,04        | 2,23        |
| SOTM-D <sub>a</sub>    | 8,90        | 14,13        | 12,85        | 6,22           | 21,30          | 7,90        | 13,23        | 2,27        |
| SOTM-D <sub>m</sub>    | 8,79        | 14,18        | 12,25        | 6,22           | 20,54          | 8,04        | 12,89        | 2,60        |
| Dados dessazonalizados |             |              |              |                |                |             |              |             |
| STheta                 | 10,03       | 14,45        | 12,38        | 6,35           | 20,42          | 7,79        | 13,05        | 1,44        |
| OTM                    | 9,56        | 14,36        | 12,42        | 6,19           | 21,38          | 7,90        | 13,26        | 1,47        |
| DOTM                   | <b>8,69</b> | 14,04        | 12,34        | 6,13           | 20,70          | 7,99        | 12,90        | 1,51        |
| Bagged OTM             | 9,93        | 14,18        | 12,54        | 6,13           | 20,70          | 8,04        | 13,09        | 58,15       |
| Bagged DOTM            | 9,32        | <b>13,95</b> | 12,40        | <b>6,06</b>    | <b>20,16</b>   | <b>7,45</b> | <b>12,76</b> | 56,24       |

**Tabela A.2:** Avaliação empírica utilizando a métrica MASE

| Modelo/Método          | Demográfica | Financeira  | Industrial  | Macroeconômica | Microeconômica | Outras      | All         | Tempo (min) |
|------------------------|-------------|-------------|-------------|----------------|----------------|-------------|-------------|-------------|
| Dados originais        |             |             |             |                |                |             |             |             |
| SOTM <sub>a</sub>      | 1,33        | 1,34        | 1,09        | 1,11           | 1,03           | 1,48        | 1,15        | 2,13        |
| SOTM <sub>m</sub>      | 1,34        | 1,35        | <b>1,07</b> | 1,12           | 1,02           | 1,48        | 1,15        | 2,23        |
| SOTM-D <sub>a</sub>    | 1,27        | 1,32        | 1,09        | 1,09           | 1,02           | <b>1,29</b> | <b>1,12</b> | 2,27        |
| SOTM-D <sub>m</sub>    | 1,28        | 1,33        | 1,08        | 1,09           | 1,01           | 1,30        | <b>1,12</b> | 2,60        |
| Dados dessazonalizados |             |             |             |                |                |             |             |             |
| STheta                 | 1,36        | 1,35        | 1,10        | 1,12           | 1,02           | 1,50        | 1,16        | 1,44        |
| OTM                    | 1,31        | 1,34        | 1,09        | 1,08           | 1,03           | 1,48        | 1,14        | 1,47        |
| DOTM                   | <b>1,26</b> | <b>1,31</b> | 1,10        | 1,07           | 1,02           | 1,30        | <b>1,12</b> | 1,51        |
| Bagged OTM             | 1,43        | 1,33        | 1,10        | <b>1,05</b>    | 1,01           | 1,48        | 1,14        | 58,15       |
| Bagged DOTM            | 1,35        | 1,32        | 1,10        | 1,06           | <b>1,00</b>    | 1,31        | <b>1,12</b> | 56,24       |