

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Método *bagging* para aprimoramento de previsões de séries temporais**

**Juliana Shibaki Camargo**

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Juliana Shibaki Camargo**

## Método *bagging* para aprimoramento de previsões de séries temporais

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestra em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Carlos Alberto Ribeiro Diniz  
Coorientador: Prof. Dr. Marinho Gomes de Andrade Filho

**USP – São Carlos**  
**Novembro de 2021**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

C172m Camargo, Juliana Shibaki  
Método bagging para aprimoramento de previsões  
de séries temporais / Juliana Shibaki Camargo;  
orientador Carlos Alberto Ribeiro Diniz;  
coorientador Marinho Gomes de Andrade Filho. -- São  
Carlos, 2021.  
79 p.

Dissertação (Mestrado - Programa  
Interinstitucional de Pós-graduação em Estatística) --  
Instituto de Ciências Matemáticas e de Computação,  
Universidade de São Paulo, 2021.

1. Bagging. 2. Bootstrap. 3. Moving Block  
Bootstrap. 4. Previsão. 5. Séries temporais. I.  
Diniz, Carlos Alberto Ribeiro, orient. II. Filho,  
Marinho Gomes de Andrade, coorient. III. Título.

**Juliana Shibaki Camargo**

## Bagging method for improving time series forecasts

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.  
*FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Carlos Alberto Ribeiro Diniz

Co-advisor: Prof. Dr. Marinho Gomes de Andrade Filho

**USP – São Carlos**  
**November 2021**





# UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

## Folha de Aprovação

---

Defesa de Dissertação de Mestrado da candidata Juliana Shibaki Camargo, realizada em 22/10/2021.

### Comissão Julgadora:

Prof. Dr. Carlos Alberto Ribeiro Diniz (UFSCar)

Prof. Dr. Anderson Luiz Ara Souza (UFPR)

Prof. Dr. José Augusto Fiorucci (UnB)

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

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.





# AGRADECIMENTOS

---

---

Em primeiro lugar, agradeço ao meu orientador Carlos Alberto Ribeiro Diniz e meu coorientador Marinho Gomes de Andrade Filho pelas sugestões e correções que contribuíram para o meu aprendizado e para o desenvolvimento desta dissertação.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela tão valiosa bolsa de estudo cedida, que possibilitou e foi fundamental para o desenvolvimento deste trabalho.

Por fim, agradeço à minha família e amigos, que sempre estiveram presentes, apoiando minhas escolhas e acreditando em meu potencial.



# RESUMO

CAMARGO, J. S. **Método *bagging* para aprimoramento de previsões de séries temporais.** 2021. 79 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Diferentes metodologias são propostas e exploradas com o intuito de reduzir o erro de previsão de séries temporais. Uma estratégia que vem se apresentando bastante promissora consiste em combinar diferentes previsões de diferentes modelos a fim de se obter uma melhor acurácia, ou seja, um menor erro de previsão. Este trabalho teve como objetivo realizar um estudo e aplicação do método *bootstrap aggregating*, mais conhecido como *bagging*, para aprimorar previsões de séries temporais. Primeiramente, cada série temporal foi separada em série de treinamento e série de teste, e então utilizou-se a metodologia *moving block bootstrap* aplicada à série de treinamento para gerar diferentes séries reamostradas, realizar a previsão de cada uma delas e combiná-las, obtendo-se assim uma previsão final combinada. Posteriormente, a série de teste foi utilizada para calcular a acurácia dos modelos, individual e combinado. Foram realizados um estudo com séries simuladas e uma aplicação com séries temporais reais mensais. O modelo escolhido e ajustado para cada uma das séries foi o auto-regressivo integrado de médias móveis (ARIMA). As medidas de acurácia utilizadas foram o erro quadrático médio e sua raiz, o erro percentual absoluto médio arcotangente e o erro percentual absoluto médio simétrico. Ao final do estudo, explorou-se o impacto que a variação dos parâmetros da reamostragem do modelo combinado causa na previsão e foram realizadas comparações entre os métodos de previsão combinado e individual.

**Palavras-chave:** *Bagging, Bootstrap, Moving Block Bootstrap, Previsão, Séries temporais.*



# ABSTRACT

CAMARGO, J. S. **Bagging method for improving time series forecasts**. 2021. 79 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Different methodologies are proposed and explored aiming to reduce time series forecasting error. A promising approach consists in combining different forecasts from different models in order to get a better accuracy, i.e., a smaller forecast error. This work aims to review and apply the bootstrap aggregating method, also known as bagging, in order to improve time series forecasting. First, each time series is divided into training and testing time series, and then the moving block bootstrap methodology is applied to the training series to generate different resampled time series, and then forecasting for each one of the series is performed and combined, thus obtaining the final combined forecast. The test data set is used to calculate the accuracy of the models, individual and combined. A simulation study of time series and application to a real time series data sets are presented. The chosen and fitted model for each of the time series was an autoregressive integrated moving average (ARIMA). The accuracy measurements used were the mean square error and its root, mean arctangent absolute percentage error and the symmetric mean absolute percentage error. Finally, the impact on the forecasts of the combined model by varying the resampling method parameters was explored and comparisons between the combined and individual forecasting methods were also carried out.

**Keywords:** Bagging, Bootstrap, Forecast, Moving Block Bootstrap, Time Series.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Série temporal periódica. . . . .	32
Figura 2 – MSE da autocorrelação de ordem 1 até 4 para diferentes tamanhos de blocos. . . . .	35
Figura 3 – MSE dos parâmetros estimados $\hat{\phi}_1$ e $\hat{\phi}_2$ para diferentes tamanhos de blocos. . . . .	36
Figura 4 – MSE da variância estimada para diferentes tamanhos de blocos. . . . .	36
Figura 5 – A primeira imagem mostra a série temporal mensal número 350 da competição M3 original e a segunda imagem mostra a mesma série depois de se realizar a transformação Box-Cox. . . . .	41
Figura 6 – Regressão local LOESS aplicada a um conjunto de dados que contém a temperatura média anual em graus Fahrenheit em New Haven, Connecticut, de 1912 a 1971. Utilizou-se três diferentes valores para o parâmetro de suavização $\alpha$ . . . . .	45
Figura 7 – Decomposição STL aplicada no conjunto de dados de números mensais de passageiros em voos internacionais, de 1949 a 1960, utilizando a série original. . . . .	47
Figura 8 – Decomposição STL aplicada no conjunto de dados de números mensais de passageiros em voos internacionais, de 1949 a 1960, utilizando a série transformada (Box-Cox). . . . .	48
Figura 9 – Versões <i>bootstrap</i> (azul) da série temporal mensal número 350 da competição M3 (preto). São mostradas $B = 100$ séries <i>bootstrap</i> , com tamanho de bloco $l = 24$ . . . . .	49
Figura 10 – Fluxograma <i>bagging</i> . . . . .	52
Figura 11 – Séries simuladas a partir de um modelo autorregressivo de ordem dois, com parâmetros fixados, e suas respectivas variâncias. . . . .	56
Figura 12 – Séries simuladas a partir de um modelo de médias móveis de ordem três, com parâmetros fixados, e suas respectivas variâncias. . . . .	56
Figura 13 – Séries simuladas a partir da soma de uma série proveniente de um modelo autorregressivo de ordem dois com uma série proveniente de um modelo de médias móveis de ordem três, ambas com parâmetros fixados, e suas respectivas variâncias. . . . .	57
Figura 14 – Gráfico do comportamento do MSE e do MAAPE, utilizando a série gerada a partir da semente de número 100, quando o número de amostras <i>bootstrap</i> e o tamanho do bloco escolhido variam. . . . .	60

Figura 15 – Gráfico do comportamento do MSE e do MAAPE, utilizando a série gerada a partir da semente de número 95, quando o número de amostras <i>bootstrap</i> e o tamanho do bloco escolhido variam. . . . .	61
Figura 16 – Gráfico da diferença de médias do RMSE da previsão da série original e o RMSE da previsão da série combinada utilizando tamanho de bloco $l = 12$ e amostras <i>bootstrap</i> $B = 100$ . . . . .	65
Figura 17 – Gráfico da diferença de médias do MAAPE da previsão da série original e o MAAPE da previsão da série combinada utilizando tamanho de bloco $l = 12$ e amostras <i>bootstrap</i> $B = 100$ . . . . .	65
Figura 18 – Gráfico da diferença de médias do sMAPE da previsão da série original e o sMAPE da previsão da série combinada utilizando tamanho de bloco $l = 12$ e amostras <i>bootstrap</i> $B = 100$ . . . . .	66
Figura 19 – Gráfico do comportamento do RMSE, do MAAPE e do sMAPE, utilizando a série temporal mensal número 350 da competição M3, quando o número de amostras <i>bootstrap</i> e o tamanho do bloco escolhido variam. . . . .	67
Figura 20 – Gráfico do comportamento do RMSE, do MAAPE e do sMAPE, utilizando a série temporal mensal número 360 da competição M3, quando o número de amostras <i>bootstrap</i> e o tamanho do bloco escolhido variam. . . . .	68



# LISTA DE ALGORITMOS

---

---

Algoritmo 1 – Método de Guerrero para a escolha do valor de $\lambda$ . . . . .	41
Algoritmo 2 – Algoritmo STL . . . . .	46
Algoritmo 3 – Método <i>bagging</i> para previsão de séries temporais . . . . .	53



# LISTA DE TABELAS

---

---

Tabela 1	– Casos especiais da transformação Box-Cox. . . . .	40
Tabela 2	– Número de séries simuladas de diferentes maneiras que obtiveram desempenho superior em cada modelo, assumindo $B = 10$ séries <i>bootstrap</i> , diferentes medidas de acurácia e diferentes tamanhos de blocos utilizados no processo de reamostragem. . . . .	56
Tabela 3	– Número de séries simuladas de diferentes maneiras que obtiveram desempenho superior em cada modelo, assumindo $B = 100$ séries <i>bootstrap</i> , diferentes medidas de acurácia e diferentes tamanhos de blocos utilizados no processo de reamostragem. . . . .	57
Tabela 4	– Média e mediana das medidas de acurácia das séries, simuladas de diferentes maneiras, assumindo $B = 10$ séries <i>bootstrap</i> e diferentes tamanhos de blocos utilizados no processo de reamostragem. . . . .	58
Tabela 5	– Média e mediana das medidas de acurácia das séries, simuladas de diferentes maneiras, assumindo $B = 100$ séries <i>bootstrap</i> e diferentes tamanhos de blocos utilizados no processo de reamostragem. . . . .	58
Tabela 6	– Porcentagem e número de séries que obtiveram desempenho superior em cada modelo, para diferentes medidas de acurácia e diferentes tamanhos de blocos utilizados na reamostragem. . . . .	64
Tabela 7	– Média e mediana das medidas de acurácia de cada modelo, separadas por diferentes tamanhos de blocos utilizados na reamostragem. . . . .	64
Tabela 8	– Intervalos de confiança para a média da diferença das medidas de acurácia calculadas, fixando tamanho de bloco como $l = 12$ e número de amostras <i>bootstrap</i> $B = 100$ para o modelo combinado. . . . .	66
Tabela 9	– Número de séries presentes em cada segmento do banco de dados mensais da competição M3 de previsão. . . . .	69
Tabela 10	– Porcentagem e número de séries que obtiveram desempenho superior em cada modelo de acordo com o RMSE, separadas por diferentes segmentos e tamanhos de blocos utilizados na reamostragem. . . . .	69
Tabela 11	– Porcentagem e número de séries que obtiveram desempenho superior em cada modelo sob diferentes métricas, utilizando 145 séries temporais financeiras mensais. . . . .	70
Tabela 12	– Média e mediana das medidas de acurácia de cada modelo, sob diferentes métricas, utilizando 145 séries temporais financeiras mensais. . . . .	70

Tabela 13 – Porcentagem e número de séries que obtiveram desempenho superior em cada modelo sob diferentes métricas, utilizando 58 séries temporais financeiras anuais. . . . .	71
Tabela 14 – Média e mediana das medidas de acurácia de cada modelo, sob diferentes métricas, utilizando 58 séries temporais financeiras anuais. . . . .	71

# LISTA DE ABREVIATURAS E SIGLAS

---

---

AIC	Critério de Informação de Akaike
AICc	Critério de Informação de Akaike corrigido
AR	Modelo Autorregressivo
ARIMA	Modelo Autorregressivo, Integrado e de Médias Móveis
ARMA	Modelo Autorregressivo e de Médias Móveis
ETS	<i>Error Trend Seasonal</i>
FAC	Função de Autocorrelação
IID	Independentes e Identicamente Distribuídos
KNN	<i>K-nearest neighbor</i>
LOESS	<i>Locally Estimated Scatterplot Smoothing</i>
MA	Modelo de médias móveis
MAAPE	Erro Percentual Absoluto Médio Arcotangente
MAPE	Erro Percentual Absoluto Médio
MSE	<i>Mean Squared Error</i>
sMAPE	Erro Percentual Absoluto Médio Simétrico
STL	<i>Seasonal and Trend decomposition using LOESS</i>



# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	23
2	REAMOSTRAGEM DE SÉRIES TEMPORAIS . . . . .	27
2.1	Método <i>bootstrap</i> . . . . .	27
2.2	Dados não IID . . . . .	28
2.3	Reamostragem <i>bootstrap</i> em blocos . . . . .	29
2.3.1	<i>Moving Block Bootstrap</i> . . . . .	29
2.3.2	<i>Non-overlapping Block Bootstrap</i> . . . . .	30
2.3.3	<i>Generalized Block Bootstrap</i> . . . . .	31
2.3.3.1	<i>Circular Block Bootstrap (CBB)</i> . . . . .	32
2.4	Escolha da melhor técnica de reamostragem . . . . .	33
2.4.1	<i>Modelo autorregressivo</i> . . . . .	33
2.4.2	<i>Função de autocorrelação</i> . . . . .	33
2.4.3	<i>Erro quadrático médio</i> . . . . .	34
2.4.4	<i>Simulação</i> . . . . .	34
2.4.5	<i>Considerações finais</i> . . . . .	36
3	MÉTODO <i>BAGGING</i> PARA PREVISÃO DE SÉRIES TEMPORAIS	39
3.1	Transformação Box-Cox . . . . .	40
3.2	Regressão local . . . . .	41
3.2.1	<i>Parâmetro de suavização</i> . . . . .	43
3.2.2	<i>Grau do polinômio local</i> . . . . .	43
3.2.3	<i>Função de ponderação</i> . . . . .	44
3.3	Decomposição STL . . . . .	45
3.4	Reamostragem em blocos . . . . .	46
3.5	Modelos de séries temporais . . . . .	47
3.5.1	<i>Modelo autorregressivo</i> . . . . .	47
3.5.2	<i>Modelo de médias móveis</i> . . . . .	48
3.5.3	<i>Modelo autorregressivo e de médias móveis</i> . . . . .	49
3.5.4	<i>Modelo autorregressivo, integrado e de médias móveis</i> . . . . .	49
3.6	Previsão de modelos ARIMA sazonais . . . . .	50
3.7	Critério de informação de Akaike corrigido . . . . .	51
3.8	Método <i>bagging</i> para previsão de séries temporais . . . . .	51

3.9	Medidas de acurácia . . . . .	53
3.9.1	<i>Erro quadrático médio</i> . . . . .	53
3.9.2	<i>Erro percentual absoluto médio arcotangente</i> . . . . .	53
3.9.3	<i>Erro percentual absoluto médio simétrico</i> . . . . .	54
4	<b>SIMULAÇÃO</b> . . . . .	55
4.1	Estudo do comportamento dos parâmetros do processo de reamostragem . . . . .	59
5	<b>APLICAÇÃO</b> . . . . .	63
5.1	Séries Temporais Financeiras . . . . .	70
6	<b>CONCLUSÃO</b> . . . . .	73
	<b>REFERÊNCIAS</b> . . . . .	77



---

# INTRODUÇÃO

---

Previsão de séries temporais é um tema importante e amplamente discutido na literatura. Inúmeros métodos foram e ainda são propostos com o objetivo de se obter uma melhor acurácia, ou seja, um menor erro de previsão. Métodos que consistem em combinar previsões de diferentes modelos estão se destacando ao longo dos anos. Algumas das primeiras contribuições sobre o assunto foram o estudo realizado por [Bates e Granger \(1969\)](#), que combinou previsões de diferentes modelos de séries temporais e obteve resultados bastante promissores; o trabalho de [Dickinson \(1973\)](#), que considera as distribuições amostrais dos pesos a serem atribuídos aos componentes e da variância do erro da previsão combinada; e a ideia abordada por [Bunn \(1975\)](#), que considerou uma abordagem Bayesiana para a combinação de previsões. A partir de então, inúmeros outros métodos foram e ainda são propostos e explorados a fim de se obter um menor erro de previsão utilizando previsões combinadas. [Clemen \(1989\)](#), [Mancuso e Werner \(2013\)](#) fazem revisões extensivas desses procedimentos.

Uma estratégia que vem se destacando nos dias atuais e que será abordada neste trabalho consiste em utilizar o método *bootstrap aggregating*, mais conhecido como *bagging*, para aprimorar a acurácia das previsões de modelos de séries temporais individuais. Resumidamente, o procedimento consiste na divisão da série temporal original em série de treinamento e série de teste. Utilizando a série de treinamento, aplica-se a transformação Box-Cox, realiza-se a decomposição da série em tendência, sazonalidade e componente aleatório e aplica-se o método de reamostragem em blocos escolhido no componente aleatório extraído da série. Então, adiciona-se a tendência e/ou sazonalidade de volta para a série, caso exista, e também o inverso da transformação Box-Cox, resultando assim nas séries reamostradas. Em seguida, realiza-se a previsão para as  $B$  séries, as  $B - 1$  séries reamostradas mais a série de treinamento original, e a média delas será a previsão combinada. Posteriormente, a série de teste é utilizada para calcular a acurácia dos modelos, individual e combinado. Todas as técnicas e procedimentos utilizados serão descritos mais detalhadamente no [Capítulo 3](#).

Cordeiro e Neves (2009), Bergmeir, Hyndman e Benítez (2016) foram pioneiros em explorar e aplicar o *bagging* no contexto de séries temporais. O primeiro trabalho combinou o uso do método de suavização exponencial com a metodologia *bootstrap* para realizar previsões de séries temporais. O segundo mostrou que este procedimento, utilizando modelos *Error Trend Seasonal* (ETS), pode gerar bons resultados. O trabalho realizado por Dantas, Oliveira e Repolho (2017), utilizou os métodos *bagging* e Holt-Winters para aprimorar previsões de demanda de séries da indústria de transporte aéreo. Outro estudo importante, que cita e compara a acurácia de diferentes métodos de combinação de modelos para gerar previsões, incluindo o *bagging*, foi realizado por Petropoulos, Hyndman e Bergmeir (2018). O método também foi utilizado na tentativa de aprimorar previsões de demanda do setor de energia elétrica no estudo realizado por Oliveira e Oliveira (2018). Além dos trabalhos citados anteriormente, existem outros diversos estudos que exploram o método *bagging* aplicado em diferentes setores, com o objetivo de reduzir erros de previsões de modelos individuais de séries temporais.

Esta dissertação, além de aplicar o método *bagging* em séries temporais reais e simuladas a fim de verificar seu desempenho, busca explorar o impacto que a variação dos parâmetros utilizados no processo de reamostragem das séries (tamanho do bloco escolhido e número de séries reamostradas) tem na previsão final combinada. O processo de reamostragem é uma etapa fundamental e de extrema importância no método *bagging* e, para que o desempenho do mesmo seja o melhor possível, deve-se explorar métodos adequados para a utilização em séries temporais, que possuem uma estrutura de dependência entre as observações.

Métodos de reamostragem são muito utilizados para quantificar incertezas, calculando-se erros padrão e intervalos de confiança, por exemplo, e suas vantagens são a exigência de menos suposições e que geralmente fornecem respostas mais precisas do que os métodos tradicionais (MOORE *et al.*, 2008). Utilizando reamostragem não é necessário supor que as distribuições dos dados sejam normais e nem que as amostras sejam muito grandes, pois os resultados podem ser obtidos assintoticamente por pequenas amostras. Esse tipo de método procura substituir a análise estatística teórica por procedimentos computacionais, que estão cada vez mais acessíveis e otimizados.

Uma das técnicas mais famosas de reamostragem é o método *bootstrap*, que exige que os dados a serem reamostrados sejam provenientes de variáveis aleatórias independentes e identicamente distribuídas, o que limita sua aplicação direta no caso de séries temporais, em que existe uma dependência entre as observações, e como consequência o método *bootstrap* original proposto por Efron (1979) não pode ser utilizado neste contexto. Alguns métodos foram desenvolvidos como extensões do método *bootstrap* original para lidar com este tipo de situação. Três desses métodos de reamostragem *bootstrap* em blocos serão discutidos com detalhes no Capítulo 2, sendo eles o *Moving Block Bootstrap* (MBB), proposto por Kunsch (1989), *Non-overlapping Block Bootstrap* (NBB), baseado no trabalho de Carlstein (1986) e o *Circular Block Bootstrap* (CBB), proposto por Politis e Romano (1992), que buscam trabalhar com dados que

apresentam esse tipo de estrutura de dependência. O método identificado como mais eficaz será utilizado na implementação do método *bagging* para aprimoramento de previsões de séries temporais.

A fim de prosseguir com o estudo do método *bagging* para aprimoramento de previsões de séries temporais, foi definido que o modelo escolhido para cada uma das séries reamostradas será da classe dos modelos autorregressivos, integrados e de médias móveis (ARIMA) que melhor se ajusta aos dados com base no critério de informação de Akaike corrigido (AICc), muito utilizado para lidar com tamanhos amostrais pequenos. As medidas de acurácia utilizadas para definir o desempenho das previsões individuais e combinadas serão o erro quadrático médio e sua raiz, o erro percentual absoluto médio arcotangente e o erro percentual absoluto médio simétrico, em inglês, *mean squared error* (MSE), *root mean squared error* (RMSE), *mean arctangent absolute percentage error* (MAAPE) e *symmetric mean absolute percentage error* (sMAPE), respectivamente. Serão utilizadas neste trabalho as siglas em inglês citadas anteriormente para se referir às medidas de acurácia.

Inicialmente, serão estudadas séries geradas a partir de três simulações diferentes. A primeira apresenta séries provenientes de um modelo autorregressivo de ordem dois, a segunda apresenta séries provenientes de um modelo de médias móveis de ordem três e a terceira apresenta séries provenientes da soma de uma série simulada a partir de um modelo autorregressivo de ordem dois com uma série simulada a partir de um modelo de médias móveis de ordem três, todas com parâmetros fixados. A escolha das diferentes simulações se justifica pelo fato de cada uma delas apresentar média distinta para as variâncias amostrais das séries obtidas. As séries serão divididas em duas partes: série de treinamento e série de teste. A série de treinamento será utilizada para gerar as séries reamostradas. Para cada uma destas novas séries amostradas, será ajustado o modelo de séries temporais e serão obtidas as previsões. Estas previsões serão combinadas para o cálculo da previsão final. Já a série de teste será utilizada na determinação da acurácia das previsões, tanto do modelo combinado quanto do modelo individual.

Na finalização desta dissertação, o método será aplicado em 1.428 séries temporais reais mensais, provenientes de diferentes segmentos (séries financeiras, demográficas, industriais, etc.), disponibilizadas pela competição M3<sup>1</sup> de previsão, que podem ou não possuir tendência e/ou sazonalidade. O processo será realizado da mesma maneira do descrito para as séries simuladas e os resultados das séries temporais reais mensais e das séries simuladas serão comparados no final do estudo.

Resumidamente, esta dissertação se inicia com um breve estudo sobre reamostragem de séries temporais, um procedimento de extrema importância na implementação do método *bagging*. Posteriormente é apresentada toda a metodologia utilizada no método *bagging*, além da escolha do procedimento de reamostragem, que é definido de acordo com os resultados apresentados anteriormente. São realizados três estudos de simulação e um estudo com séries temporais

<sup>1</sup> <<https://forecasters.org/resources/time-series-data/m3-competition/>>, acessado em 17/10/2020.

reais, e seus resultados são apresentados. Na finalização da dissertação são apresentadas as principais conclusões e discutidos os possíveis trabalhos futuros.

---

# REAMOSTRAGEM DE SÉRIES TEMPORAIS

---

A reamostragem é um procedimento fundamental na implementação do método *bagging* para aprimoramento de previsões de séries temporais. Neste capítulo serão apresentados e comparados alguns métodos de reamostragem em blocos elaborados para lidar com a estrutura de dependência entre as observações presentes em séries temporais. A realização de estudos de comparação entre os métodos propostos para tal finalidade é importante, pois necessitamos determinar qual é o procedimento mais eficaz para preservar a estrutura de dependência da série, e para isso será verificado como a variação do tamanho do bloco refletiu nas estimativas da autocorrelação, variância e parâmetros estimados da série reamostrada em relação à série original, que será simulada de um processo autorregressivo de ordem dois. O método de reamostragem escolhido será implementado posteriormente no estudo proposto. Para a determinação deste método, serão apresentados ao longo do Capítulo uma breve explanação do método *bootstrap*, dados não IID, reamostragem *bootstrap* em blocos, modelo autorregressivo, função de autocorrelação e erro quadrático médio.

## 2.1 Método *bootstrap*

Reamostragem é o nome dado a um conjunto de métodos ou técnicas utilizados para calcular estimativas a partir de repetidas amostragens obtidas dentro de uma única amostra. Estas técnicas possuem o objetivo de quantificar incertezas relacionadas a obtenção de estatísticas de dados com distribuições amostrais desconhecidas, e também avaliar a significância de testes cujas estatísticas básicas não possuem suas propriedades bem estabelecidas ou cujas premissas não podem ser consideradas verdadeiras. Testes de Aleatorização (Testes de Permutação), Jackknife, Bootstrap e Validação Cruzada são alguns exemplos de técnicas de reamostragem.

O procedimento *bootstrap* é uma técnica computacional intensiva de reamostragem para dados Independentes e Identicamente Distribuídos (IID), inicialmente introduzida por [Efron \(1979\)](#), que permite aproximar a distribuição de uma função das observações pela distribuição

empírica dos dados baseado em uma amostra finita.

Neste método assume-se que  $(X_1, X_2, \dots)$  é uma sequência de variáveis aleatórias IID com uma distribuição comum  $F$ . Supondo que  $\mathbf{X}_n = (X_1, \dots, X_n)$  seja a amostra aleatória observada de tamanho  $n$ , e  $T_n = t_n(\mathbf{X}_n; F)$  seja a estatística de interesse para  $n \geq 1$ , o objetivo é estimar a distribuição amostral de  $T_n$  (desconhecida) ou alguma outra característica relacionada a ela, como seu desvio padrão, por exemplo. O método fornece uma maneira eficaz de resolver este problema sem assumir nenhum modelo para  $F$ .

No *bootstrap* não-paramétrico, a reamostragem é feita com reposição da amostra original, em que supõe-se que as observações são obtidas da função de distribuição empírica  $F$ , que designa uma massa de probabilidade igual a  $1/n$  para cada ponto amostral, ou seja,

$$P_*(X_i^* = X_i) = \frac{1}{n}, \quad (2.1)$$

em que  $1 \leq i \leq n$ ,  $P_*$  denota a probabilidade condicional dado  $\mathbf{X}_n$  e a distribuição dos  $X_i^*$ 's é a distribuição empírica  $F$ .

A técnica procura realizar o que seria desejável na prática para reduzir desvios e prover desvios padrão mais confiáveis, se tal fosse possível: repetir a experiência. As observações são escolhidas de forma aleatória e as estimativas são recalculadas para cada uma das reamostras e a distribuição amostral de  $T_n$  é aproximada utilizando simulação de Monte Carlo (ROBERT, 2004). Assim, a técnica *bootstrap* assume a amostra original como se ela representasse exatamente toda a população. Segundo Proença (1988), até agora o *bootstrap* chega aos mesmos resultados que o processo tradicional, baseado na máxima verossimilhança.

Devido ao fato de que a reamostragem *bootstrap* assume que os dados da amostra original sejam IID, o que não ocorre no caso de uma série temporal, utiliza-se extensões deste método para lidar com dados dependentes.

## 2.2 Dados não IID

Em diversas situações não é possível assumir que os dados da amostra original são IID, o que torna o procedimento *bootstrap* um pouco mais complicado pois a “população” não se caracteriza completamente e unicamente pela distribuição unidimensional marginal  $F$ , mas requer conhecimento da distribuição conjunta de toda a sequência  $(X_1, X_2, \dots, X_n)$ . Séries temporais são exemplos de dados que possuem essa característica.

Alguns métodos foram propostos para lidar com o contexto de dados dependentes, e consideraremos aqui alguns métodos *bootstrap* em blocos, que são comumente utilizados quando trabalhamos com séries temporais, e são extensões do método *bootstrap* original. Neste estudo serão abordados três destes métodos.

## 2.3 Reamostragem *bootstrap* em blocos

Dentre os métodos de reamostragem existentes, o método de reamostragem em blocos é proposto como sendo uns dos mais eficazes para lidar com séries temporais. Este procedimento divide a série temporal original em blocos de dados consecutivos que são reamostrados via *bootstrap* para formar a série reamostrada, e se justifica por manter a estrutura de dependência dos dados dentro de cada bloco, que podem ser escolhidos inicialmente como sendo sobrepostos ou não sobrepostos. É necessário ressaltar que este método tem como pressuposto os dados serem provenientes de uma série temporal fracamente estacionária e fracamente dependente.

Segundo Morettin e Tolo (2004), um processo estocástico  $X = \{X(t)\}_{t \in T}$ , em que  $T$  é o conjunto índice que controla a evolução do tempo, é denominado fracamente estacionário ou estacionário de segunda ordem (ou em sentido amplo), se e somente se:

1.  $E\{X(t)\} = \mu(t) = \mu$ , constante, para todo  $t \in T$ ;
2.  $E\{X^2(t)\} < \infty$ , para todo  $t \in T$ ;
3.  $\gamma(t_1, t_2) = Cov\{X(t_1), X(t_2)\}$  é uma função de  $|t_1 - t_2|$ .

Então, a média, variância e estrutura de autocorrelação da série não mudam no decorrer do tempo. Uma série temporal estacionária é fracamente dependente quando  $x_t$  e  $x_{t+h}$  se tornam progressivamente menos dependentes à medida em que a defasagem  $h$  aumenta.

### 2.3.1 Moving Block Bootstrap

Método de reamostragem formulado por Kunsch (1989) e posteriormente reformulado por Liu e Singh (1992), que pode ser utilizado em dados dependentes sem se assumir nenhum modelo paramétrico. Neste procedimento não reamostramos uma única observação de cada vez, mas sim blocos sobrepostos de observações consecutivas no tempo. Como resultado, temos que a estrutura de dependência dentro de cada bloco é preservada, assim, quando os dados são gerados por um processo fracamente dependente o MBB produz a estrutura subjacente de dependência do processo assintoticamente. O tamanho de cada bloco cresce de acordo com o tamanho da amostra original.

Seja  $\{X_t\}_{t \in \mathbb{N}}$  uma série temporal estacionária e fracamente dependente e  $\mathbf{X}_n = (X_1, \dots, X_n)$  denota a série observada. Define-se a versão MBB dos estimadores da forma  $\hat{\theta}_n = T(F_n)$ , em que  $F_n$  é a função de distribuição empírica da amostra aleatória e  $T(\cdot)$  é uma função de  $F_n$ . Assume-se também que  $l \equiv l_n \in [1 : n]$  é um número inteiro e para dados dependentes temos que  $l \rightarrow \infty$  e  $1/n \rightarrow 0$  quando  $n \rightarrow \infty$ .

Definimos os blocos sobrepostos  $(\mathbb{B}_1, \dots, \mathbb{B}_N)$ , cada um com tamanho  $l$ , como  $\mathbb{B}_i = (X_i, \dots, X_{i+l-1})$ , em que  $i = 1, \dots, N$  e  $N = n - l + 1$  representa o total de número de blocos possí-

veis a partir da amostra original. Ou seja,

$$\begin{aligned}\mathbb{B}_1 &= (X_1, X_2, \dots, X_l) \\ \mathbb{B}_2 &= (X_2, X_3, \dots, X_{l+1}) \\ &\dots \\ \mathbb{B}_N &= (X_{n-l+1}, \dots, X_n)\end{aligned}$$

Para obter a amostra MBB é necessário escolher aleatoriamente um certo número de blocos dentre os  $(\mathbb{B}_1, \dots, \mathbb{B}_N)$ . Seja  $(\mathbb{B}_1^*, \dots, \mathbb{B}_k^*)$  uma amostra aleatória simples, retirada com reposição de  $(\mathbb{B}_1, \dots, \mathbb{B}_N)$ , em que cada um dos blocos selecionados possui tamanho  $l$ . Denota-se os elementos em  $\mathbb{B}_i$  como  $(X_{(i-1)l+1}^*, \dots, X_{il}^*)$ , em que  $i = 1, \dots, k$  e  $k$  denota a quantidade de blocos escolhidos com reposição para formar a amostra *bootstrap*. Então,  $(X_1^*, \dots, X_m^*)$  constitui a amostra MBB de tamanho  $m \equiv kl$ . A versão MBB do parâmetro  $\hat{\theta}_n$  é denotada como  $\theta_{m,n}^*$  e é definida como  $\theta_{m,n}^* = T(F_{m,n}^*)$ , em que  $F_{m,n}^*$  denota a distribuição empírica de  $(X_1^*, \dots, X_m^*)$ .

Pode-se notar que selecionar aleatoriamente blocos  $\mathbb{B}_i^*$ 's de  $(\mathbb{B}_1, \dots, \mathbb{B}_N)$  é equivalente a selecionar  $k$  índices aleatórios de  $(1, \dots, N)$ . Tome  $(I_1, \dots, I_k)$  como sendo variáveis aleatórias IID com distribuição uniforme discreta em  $(1, \dots, N)$ . Se  $\mathbb{B}_i^* = \mathbb{B}_{I_i}$ , em que  $i = 1, \dots, k$ , então  $(\mathbb{B}_1^*, \dots, \mathbb{B}_k^*)$  representa uma amostra aleatória simples com reposição de  $(\mathbb{B}_1, \dots, \mathbb{B}_N)$ . A amostra *bootstrap*  $(X_1^*, \dots, X_m^*)$  pode ser definida usando os blocos reamostrados  $(\mathbb{B}_1^*, \dots, \mathbb{B}_k^*)$ . Observa-se que condicionados a  $\mathbf{X}_n$ , os blocos reamostrados  $(X_1^*, \dots, X_l^*)', (X_{l+1}^*, \dots, X_{2l}^*)', \dots, (X_{(k-1)l+1}^*, \dots, X_{kl}^*)'$  são vetores aleatórios IID com

$$P_*((X_1^*, \dots, X_l^*)' = (X_j^*, \dots, X_{j+l-1}^*)') = P_*(I_1 = j) = N^{-1}, \quad (2.2)$$

em que  $1 \leq j \leq n$  e  $P_*$  denota a probabilidade condicional dado  $\mathbf{X}_n$ .

Quando o tamanho de bloco escolhido é igual a um ( $l = 1$ ), tem-se que o método MBB se reduz ao método *bootstrap* original. Para  $l > 1$ , a distribuição conjunta  $l$ -dimensional do processo  $\{X_n\}_{n \geq 1}$  é preservada dentro dos blocos reamostrados. Desde que o tamanho do bloco cresça à medida em que o tamanho da amostra aumenta, qualquer distribuição de probabilidade conjunta finita-dimensional pode ser recuperada a partir das observações reamostradas.

### 2.3.2 Non-overlapping Block Bootstrap

Na literatura existem muitas variações de métodos de reamostragem *bootstrap* em blocos. Outra versão conhecida é a proposta por [Carlstein \(1986\)](#), chamada *Non-overlapping Block Bootstrap*, ou simplesmente NBB. Como o próprio nome já diz, a principal diferença em relação ao método MBB está no fato de que os blocos agora serão retirados de maneira não sobreposta da amostra original. Assume-se que  $l \equiv l_n \in [1 : n]$  é um número inteiro, que representa o tamanho de cada bloco, e  $b \geq 1$  é o maior número inteiro que satisfaz  $lb \leq n$ . As amostras NBB



são geradas selecionando  $b$  blocos aleatoriamente e com reposição de  $(\tilde{\mathbb{B}}_1, \dots, \tilde{\mathbb{B}}_b)$ . Define-se  $\tilde{\mathbb{B}}_i = (X_{(i-1)l+1}, \dots, X_{il})'$ , em que  $i = 1, \dots, b$ . Ou seja,

$$\begin{aligned} \tilde{\mathbb{B}}_1 &= (X_1, X_2, \dots, X_l)' \\ \tilde{\mathbb{B}}_2 &= \qquad \qquad \qquad (X_{l+1}, \dots, X_{2l})' \\ &\dots \qquad \qquad \qquad \dots \\ \tilde{\mathbb{B}}_b &= \qquad \qquad \qquad (X_{(b-1)l+1}, \dots, X_n)' \end{aligned}$$

A implementação do NBB é exatamente a mesma da proposta pelo método MBB. Seleciona-se uma amostra aleatória simples com reposição dos blocos  $(\tilde{\mathbb{B}}_1^*, \dots, \tilde{\mathbb{B}}_k^*)$  de  $(\tilde{\mathbb{B}}_1, \dots, \tilde{\mathbb{B}}_b)$  para algum número inteiro  $k \geq 1$ , em que  $k$  denota a quantidade de blocos escolhidos com reposição para formar a amostra *bootstrap*. Então,  $(X_{2,1}^*, \dots, X_{2,m}^*)$  constitui a amostra NBB de tamanho  $m \equiv kl$ , em que o número dois é utilizado para indicar que a amostra foi gerada utilizando o método NBB.

A versão NBB do parâmetro  $\hat{\theta}_n$  é denotada como  $\theta_{m,n}^{*(2)}$  e é definida como  $\theta_{m,n}^{*(2)} = T(F_{m,n}^{*(2)})$ , em que  $F_{m,n}^{*(2)}$  denota a distribuição empírica de  $(X_{2,1}^*, \dots, X_{2,m}^*)$ . Apesar da definição dos estimadores da versão MBB e NBB serem similares, eles possuem propriedades distribucionais diferentes, como por exemplo, terem médias condicionais distintas.

### 2.3.3 Generalized Block Bootstrap

O método MBB possui um indesejável efeito de fronteira visto que atribui pesos menores às observações do início e do final da série já que, para  $l \leq j \leq n-l$ , a  $j$ -ésima observação da série  $X_j$  aparece exatamente em  $l$  blocos do conjunto  $(\mathbb{B}_1, \dots, \mathbb{B}_N)$ , enquanto que, para  $1 \leq j \leq l-1$ ,  $X_j$  e  $X_{n-j+i}$ , aparece em apenas  $j$  blocos. Já para o método NBB um problema similar acontece com as observações próximas ao final da série quando  $n$  não é múltiplo de  $l$ .

Uma possível solução para este problema surgiu com [Politis e Romano \(1992\)](#). A ideia consiste em agrupar os dados em um modelo de círculo, para que não exista um começo ou final na série temporal, ou seja, quando a série chega ao fim, logo recomeça. Esse procedimento foi chamado de *Generalized Block Bootstrap*, ou simplesmente GBB.

Seja  $\mathbf{X}_n = (X_1, \dots, X_n)$  a série observada. Primeiramente, definimos uma nova série temporal  $Y_{n,i}$ , em que  $i \geq 1$  e  $n$  representa o período. Para qualquer  $i \geq 1$ , existem números inteiros  $k_i \geq 0$  e  $j_i \in [1, n]$ , em que  $i = k_i n + j_i$ . Obtemos, então, a relação  $Y_{n,i} = X_{j_i}$ . A associação entre as variáveis  $Y$  e  $X$  pode ser facilmente visualizada na [Figura 1](#).

Os blocos são definidos como  $\mathbb{B}(i, j) = (Y_{n,i}, \dots, Y_{n,(i+j-1)})$ , com  $i \geq 1$  e  $j \geq 1$ . Seja  $\Gamma_n$  uma função de transição de probabilidade no conjunto  $\mathbb{R}^n \times \otimes_{t=1}^{\infty} (\{1, \dots, n\} \times \mathbb{N})$ , isto é, para cada  $x \in \mathbb{R}^n$ ,  $\Gamma_n(x; \cdot)$  é uma medida de probabilidade em  $\otimes_{t=1}^{\infty} (\{1, \dots, n\} \times \mathbb{N}) \equiv \{\{i_t, l_t\}_{t=1}^{\infty} : 1 \leq$

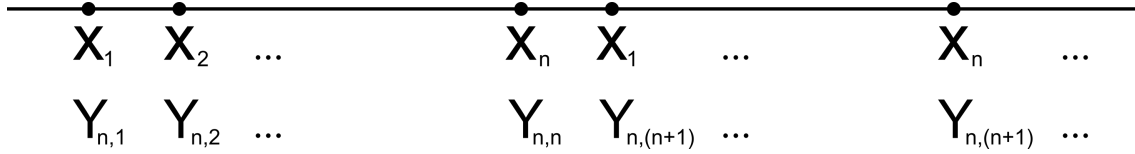


Figura 1 – Série temporal periódica.

Fonte – Elaborada pela autora.

$i_t \leq n, 1 \leq l_t \leq \infty, \forall t \leq 1\}$ , e para qualquer conjunto  $A \subset \otimes_{t=1}^{\infty} (\{1, \dots, n\} \times \mathbb{N})$ ,  $\Gamma_n(\cdot; A)$  é uma função mensurável de Borel do  $\mathbb{R}^n$  para  $[0, 1]$ , em que  $\otimes$  representa o produto tensorial.

Então, o método GBB reamostra blocos da coleção  $\{\mathbb{B}(i, j), i \geq 1, j \geq 1\}$  de acordo com a função de transição de probabilidade  $\Gamma_n$  definida anteriormente. Definimos  $\{(I_1, J_1), (I_2, J_2), \dots\}$  como sendo uma sequência de vetores aleatórios com distribuição condicional conjunta  $\Gamma_n(\mathbf{X}_n; \cdot)$ , dado  $\mathbf{X}_n$ . E, assim, os blocos selecionados pelo método GBB são dados por  $\{\mathbb{B}(I_1, J_1), \mathbb{B}(I_2, J_2), \dots\}$ , que podem não ser independentes, e  $(X_{G,1}^*, X_{G,2}^*, \dots)$  denota os elementos destes blocos reamostrados.

A versão GBB do estimador  $\hat{\theta}_n = T(F_n)$  é definida como  $\theta_{m,n}^{*(G)} = T(F_{m,n}^{*(G)})$  para um valor de  $m \geq 1$ , em que  $F_{m,n}^{*(G)}$  denota a distribuição empírica de  $X_{G,1}^*, \dots, X_{G,m}^*$ , com  $m \equiv kl$ , em que a letra G foi utilizada para indicar que a amostra foi gerada utilizando o método GBB.

### 2.3.3.1 Circular Block Bootstrap (CBB)

Proposto por [Politis e Romano \(1992\)](#), este método reamostra blocos sobrepostos e periódicos de tamanho  $l$ , em que  $1 \leq l \leq n$ , provenientes de  $\{\mathbb{B}(i, l), \dots, \mathbb{B}(n, l)\}$ . É um dos casos em que se utiliza a ideia do GBB. A função de transição  $\Gamma_n$  para o método CBB é dada por

$$\Gamma_n = \otimes_{i=1}^{\infty} \left( (n^{-1} \sum_{j=1}^n \delta_j) \times \delta_l \right), \quad x \in \mathbb{R}^n, \quad (2.3)$$

em que  $n$  é o tamanho da amostra de dados dependentes e  $\delta_y$  é uma medida de probabilidade com massa total igual a 1 em  $y$ .

Denota-se os índices dos blocos reamostrados via CBB (isto é, as variáveis  $I_i$ 's da coleção  $\{(I_1, J_1), (I_2, J_2), \dots\}$  cuja distribuição conjunta é especificada por  $\Gamma_n(\cdot; \cdot)$  por  $(I_{3,1}, I_{3,2}, \dots)$ . Então, a [Equação 2.3](#) implica que as variáveis  $(I_{3,1}, I_{3,2}, \dots)$  são condicionalmente IID com  $P_*(I_{3,1} = i) = n^{-1}$  e  $P_*(J_i = l) = 1, \forall i = 1, \dots, n$ . Desde que cada  $X_i$  apareça exatamente  $l$  vezes no conjunto de blocos  $(\mathbb{B}(i, l), \dots, \mathbb{B}(n, l))$  e que os blocos tenham probabilidade igual de pertencerem à amostra reamostrada, cada uma das observações originais  $(X_1, \dots, X_n)$  recebe o mesmo peso no método CBB, não sofrendo o efeito de fronteira existente nos métodos MBB e NBB. Os elementos dos blocos e das séries reamostradas e o estimador  $\hat{\theta}_n = T(F_n)$  são especificados na [Subseção 2.3.3](#), por ser um caso particular da reamostragem via GBB.

## 2.4 Escolha da melhor técnica de reamostragem

### 2.4.1 Modelo autorregressivo

De acordo com [Box et al. \(2015\)](#), o Modelo Autorregressivo (AR) é um modelo estocástico bastante útil na representação de diversas séries temporais. Os valores atuais da série são expressos em função de uma combinação linear de valores passados mais um choque aleatório  $\varepsilon_t$ .

Dizemos que  $X_t$  é um processo autorregressivo de ordem  $p$ , ou seja, é um  $AR(p)$  se pudermos escrever o processo da seguinte forma

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (2.4)$$

em que  $\mu$  é constante,  $(\phi_1, \dots, \phi_p)$  são parâmetros reais e  $\varepsilon_t$  é IID com  $E(\varepsilon_t) = 0$  e  $\text{Var}(\varepsilon_t) = \sigma^2$ .

O processo  $AR(p)$  também pode ser escrito em termos do operador de retardo como

$$\phi(B)X_t = \mu + \varepsilon_t, \quad (2.5)$$

em que  $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ .

Tem-se que um processo  $AR(p)$  é sempre invertível e é estacionário se a equação característica  $\phi(B) = 0$  tiver todas as raízes fora do círculo unitário. As estimativas de  $(\phi_1, \dots, \phi_p)$  foram obtidas neste trabalho via máxima verossimilhança.

### 2.4.2 Função de autocorrelação

A Função de Autocorrelação (FAC) é utilizada para medir o grau de correlação de uma variável, em um determinado instante, consigo mesma, em um instante de tempo anterior. As observações podem estar autocorrelacionadas em diversas ordens. A autocorrelação de primeira ordem ( $k = 1$ ) caracteriza séries em que uma observação está correlacionada com a observação imediatamente anterior, a de segunda ordem ( $k = 2$ ) caracteriza séries em que uma observação está correlacionada com a observação a duas unidades de tempo no passado, e assim por diante.

Assim, a FAC auxilia na caracterização do desenvolvimento da variável ao longo do tempo. É definida como a razão entre a autocovariância e a variância para um conjunto de dados (uma série temporal) com estacionariedade fraca

$$\rho(k) = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)} = \frac{\gamma(k)}{\gamma(0)} \quad (2.6)$$

O valor da FAC está entre 1 (correlação positiva perfeita) e -1 (correlação negativa perfeita). O valor 0 significa total ausência de correlação. Quando  $k = 0$ , a autocorrelação é calculada entre uma observação e ela mesma, portanto seu valor sempre será igual a 1.

### 2.4.3 Erro quadrático médio

O erro quadrático médio, ou *Mean Squared Error* (MSE) (em inglês), é uma forma de avaliar a diferença entre o estimador e o verdadeiro valor da quantidade estimada. Calcula a média do quadrado do erro, ou seja, a média do quadrado da diferença entre o valor estimado e o verdadeiro valor. O MSE entre os valores observados e os valores preditos de amostras de tamanho  $n$  é definido como

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)^2, \quad (2.7)$$

em que  $X_i$  é o valor verdadeiro da  $i$ -ésima observação e  $\hat{X}_i$  é o valor predito da  $i$ -ésima observação.

Obter um MSE igual a zero é o ideal, mas na maioria das situações isso não é possível. Um MSE igual a zero significa que o estimador prevê observações com precisão perfeita. É muito comum a utilização da raiz do erro quadrático médio (RMSE) como medida de acurácia e precisão

$$RMSE = \sqrt{MSE}. \quad (2.8)$$

### 2.4.4 Simulação

A série temporal utilizada para comparação dos métodos de reamostragem MBB, NBB e CBB foi simulada de um processo AR(2), com  $\phi_1 = 0,5$  e  $\phi_2 = 0,3$ , e por definição tem-se que a série é invertível e foi verificado que também é estacionária. Para cada um dos três métodos propostos, foram geradas  $B = 1.000$  séries reamostradas de tamanho  $n = 500$  cada. Então, calculou-se o MSE das estimativas da autocorrelação de ordem 1 até 4, das estimativas da variância e dos parâmetros estimados  $\hat{\phi}_1$  e  $\hat{\phi}_2$  do AR(2), quando possível, das séries reamostradas em relação à série original. Os parâmetros estimados,  $\hat{\phi}_1$  e  $\hat{\phi}_2$ , só foram calculados nas séries reamostradas para fins de comparação, pois como a série original é simulada, sabia-se exatamente de que processo ela foi gerada. O procedimento foi realizado para diferentes tamanhos de blocos ( $l \in [1, 50]$ ).

Observa-se pela [Figura 2](#) que para todas as ordens de autocorrelação calculadas, o MSE decresce à medida em que o tamanho do bloco aumenta. Os erros calculados pelo método NBB são os que mais oscilam conforme o tamanho do bloco cresce, e em geral, seu erro é superior aos outros métodos. Já os métodos MBB e CBB decrescem sem muita oscilação conforme o tamanho do bloco cresce. Neste exemplo, apesar de muito próximos, os erros calculados pelo método MBB são um pouco mais baixos do que os calculados pelo método CBB.

Em relação aos parâmetros estimados,  $\hat{\phi}_1$  e  $\hat{\phi}_2$ , percebe-se, pela [Figura 3](#), que o MSE de  $\phi_1$  à medida em que o tamanho do bloco cresce, apresenta uma pequena tendência de crescimento, enquanto que o de  $\phi_2$  decresce significativamente à medida em que o tamanho do bloco aumenta. Como observado nos gráficos de autocorrelação, o método NBB é o menos estável, oscilando bastante, enquanto os métodos MBB e CBB se comportam melhor. Mais uma vez, o método MBB aparece com o MSE um pouco menor do que quando comparado ao método CBB.

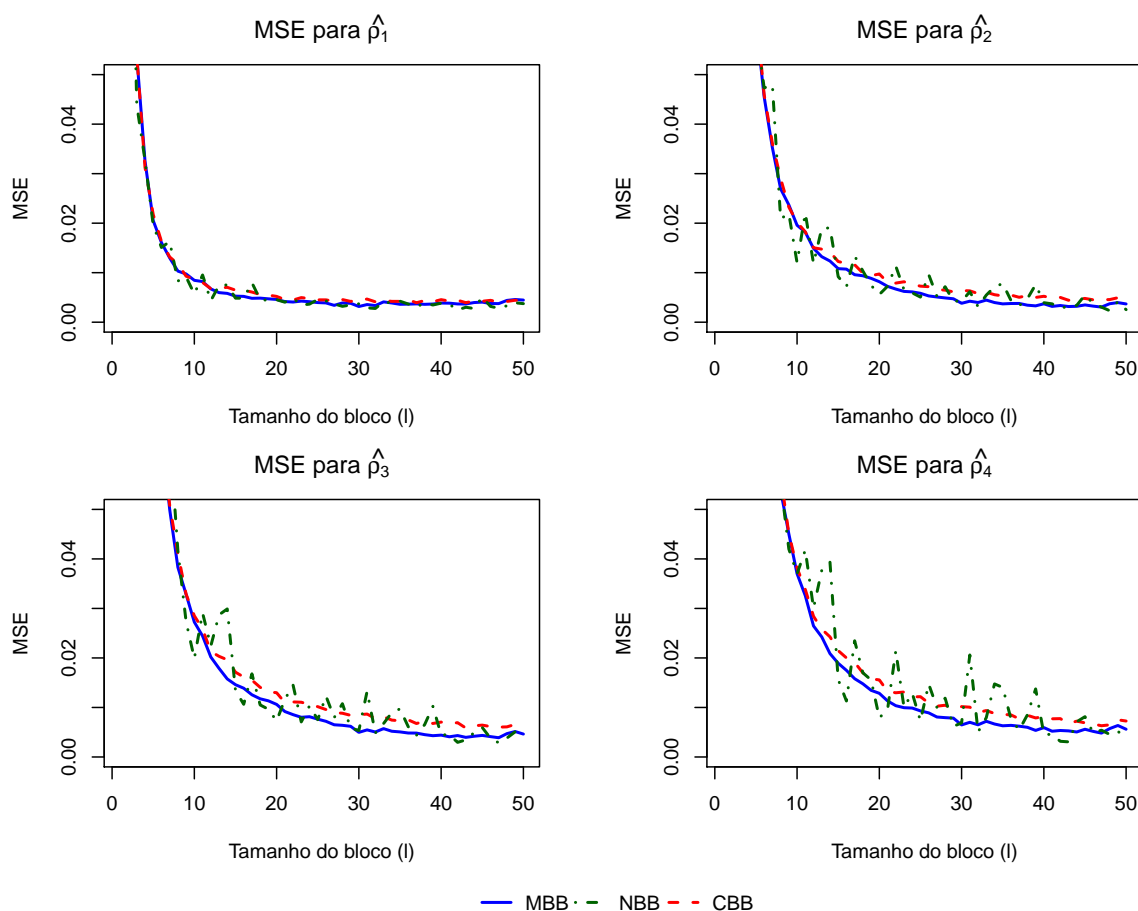


Figura 2 – MSE da autocorrelação de ordem 1 até 4 para diferentes tamanhos de blocos.

Fonte – Elaborada pela autora.

Observou-se também que em algumas das séries reamostradas o melhor ajuste foi o modelo AR(1), e não o AR(2). Essas amostras foram descartadas na elaboração do gráfico. A maioria dos casos em que isto ocorreu, foi em tamanhos de blocos muito pequenos, pois se perde as características do processo autorregressivo.

Percebe-se pelo gráfico da [Figura 4](#) que o MSE da variância das séries reamostradas cresce à medida que o tamanho do bloco cresce. Mais uma vez, os métodos MBB e CBB permanecem muito próximos quando comparados, e o método NBB oscila com maior frequência.

Para escolher o tamanho ideal de bloco a ser utilizado nas séries reamostradas devemos nos atentar que ele seja grande o suficiente para que se preserve a estrutura de autocorrelação dos dados, mas não tão grande, pois a variância da série cresce à medida em que o tamanho do bloco cresce. Após a observação dos gráficos, pode-se concluir que o método MBB é o que obteve melhor performance, ou seja, o MSE foi menor. Em relação ao tamanho ideal de cada bloco utilizado na reamostragem, tem-se que, para este caso (tomando a série original como um processo AR(2) com  $n = 500$ ), um tamanho de bloco próximo a 30 seria o ideal.

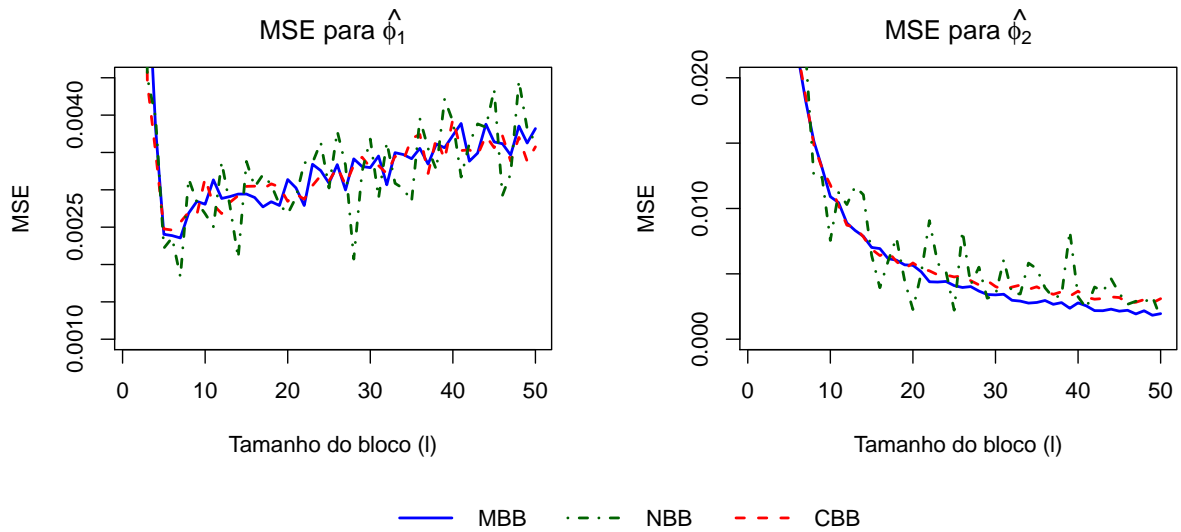


Figura 3 – MSE dos parâmetros estimados  $\hat{\phi}_1$  e  $\hat{\phi}_2$  para diferentes tamanhos de blocos.

Fonte – Elaborada pela autora.

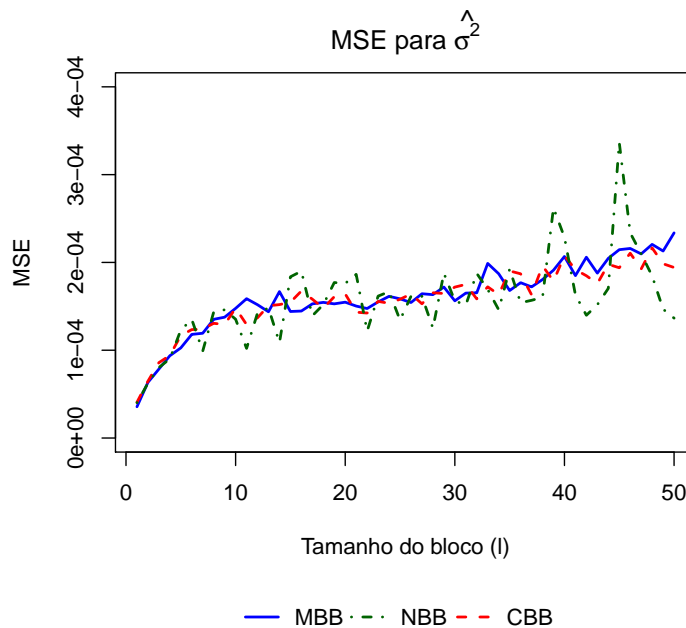


Figura 4 – MSE da variância estimada para diferentes tamanhos de blocos.

Fonte – Elaborada pela autora.

### 2.4.5 Considerações finais

Três métodos de reamostragem *bootstrap* para séries temporais foram estudados e implementados. Observou-se que o método mais eficiente para preservar a estrutura de autocorrelação, os parâmetros estimados e a variância da série foi o método *Moving Block Bootstrap*, ou MBB, e o tamanho ideal de cada bloco foi definido como sendo algo próximo de 30, para uma série de

tamanho 500 gerada de um AR(2). Assim, o método de reamostragem que será implementado no procedimento *bagging* para realizar previsões combinadas de séries temporais, será o MBB. É importante ressaltar que cada série original utilizada na criação das séries reamostradas *bootstrap* deve ser estudada e tratada de maneira individual, para que o tamanho escolhido do bloco seja coerente e razoável para manter a estrutura de dependência entre as observações presente na série original.





---

## MÉTODO *BAGGING* PARA PREVISÃO DE SÉRIES TEMPORAIS

---

O estudo realizado por [Bergmeir, Hyndman e Benítez \(2016\)](#) avaliou o desempenho do método *bagging* na acurácia de previsões de séries temporais utilizando modelos de suavização exponencial. O *bagging* foi comparado com diversos outros modelos propostos na literatura e apresentou resultados promissores ao realizar previsões. No estudo citado, inicialmente, cada série foi dividida em série de treinamento e série de teste. Utilizou-se a transformação Box-Cox seguida da decomposição STL para dividir a série de treinamento em tendência, sazonalidade e componente aleatório. O componente aleatório foi submetido a diversas reamostragens, e foram obtidas novas séries a partir da série de treinamento original. Para a série original e cada uma das séries reamostradas, foi ajustado um modelo de suavização exponencial e calculadas as previsões. A previsão resultante foi obtida calculando-se a média das previsões ponto a ponto. A série de teste foi utilizada para avaliar o desempenho da acurácia das previsões. O resultado do estudo mostrou que o método *bagging* supera consistentemente os modelos originais de suavização exponencial, e muitos outros propostos na literatura.

Neste capítulo são descritas todas as técnicas e procedimentos necessários para a implementação do método *bagging* para aprimoramento de previsões de séries temporais. Os métodos apresentados foram descritos e implementados de acordo com o estudo citado anteriormente, modificando-se apenas alguns aspectos, como o modelo ajustado e as métricas utilizadas para medir a acurácia das previsões. O método de reamostragem MBB, além de ser o utilizado no estudo em questão, foi escolhido no [Capítulo 2](#) por apresentar o melhor desempenho em manter a estrutura da série original quando comparado a outros métodos.

### 3.1 Transformação Box-Cox

A transformação Box-Cox foi formulada por [Box e Cox \(1964\)](#) e é bastante útil para estabilizar a variância de séries heteroscedásticas. Através dela, pode-se obter uma série mais simétrica, com variância constante e tornar os dados mais semelhantes à distribuição normal quando valores apropriados de  $\lambda$  são obtidos de acordo com a equação a seguir.

$$X_i(\lambda) = \begin{cases} \frac{X_i^{*\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \log(X_i^*), & \lambda = 0, \end{cases} \quad (3.1)$$

em que  $X_1^*, \dots, X_n^*$  denota a série original e  $X_1, \dots, X_n$  denota a série transformada, ambas de tamanho  $n$ . O logaritmo utilizado na transformação é sempre o natural, ou seja, de base  $e$ .

Então, se  $\lambda = 0$  a transformação aplicada aos dados é a logarítmica, caso contrário é aplicada a transformação descrita na [Equação 3.1](#). Dependendo do valor de  $\lambda$ , a transformação Box-Cox inclui alguns casos especiais, como mostrado na [Tabela 1](#).

Tabela 1 – Casos especiais da transformação Box-Cox.

Valor de $\lambda$	Transformação
1	Identidade
0,5	Raiz quadrada
0	Logarítmica
-0,5	Inverso da raiz quadrada
-1	Inverso da identidade

Fonte – Elaborada pela autora.

A utilização da transformação Box-Cox, de acordo com a [Equação 3.1](#), possui a exigência de que todos os valores da série sejam positivos e diferentes de zero. No *software* R, esta transformação pode ser realizada utilizando a função `BoxCox()` do pacote *forecast* ([HYNDMAN; KHANDAKAR, 2008](#)); caso a série não atenda a estes requisitos, o valor de  $\lambda$  é atribuído como sendo igual a 1, ou seja, a transformação utilizada é a identidade. O valor de  $\lambda$  precisa ser estimado, e é obtido através de uma varredura sobre um intervalo de valores possíveis, e o objetivo é identificar o valor que resulte em uma menor soma dos quadrados dos resíduos. Neste estudo,  $\lambda$  é restrito ao intervalo  $[0, 1]$  e o método utilizado para escolher seu valor é o proposto por [Guerrero \(1993\)](#).

O procedimento proposto por [Guerrero \(1993\)](#), tem como objetivo escolher o valor de  $\lambda$  para que a variância da série seja estabilizada. O [Algoritmo 1](#) detalha resumidamente este processo.

Para ilustrar o procedimento foi realizada a decomposição Box-Cox na série temporal mensal número 350 da competição M3. Para esta série, o valor de  $\lambda$  escolhido foi  $\lambda = 6,61 \times$

**Algoritmo 1** – Método de Guerrero para a escolha do valor de  $\lambda$ 

- 1: As  $n$  observações da série original são agrupadas em  $h$  subséries de modo que a estimativa local da média  $m$  e desvio padrão  $d$  em cada uma delas possa ser obtida;
- 2: O valor de  $\lambda$  é escolhido de forma que o coeficiente de variação de  $\frac{d}{m^{1-\lambda}}$ , calculado a partir das subséries, seja minimizado. O tamanho de cada uma das subséries é igual ao comprimento da sazonalidade da série original, e se a série não for sazonal, o tamanho das subséries é atribuído como sendo igual a dois.

$10^{-5}$ . A Figura 5 mostra a série temporal original e a série temporal transformada, obtida através do  $\lambda$  estimado.

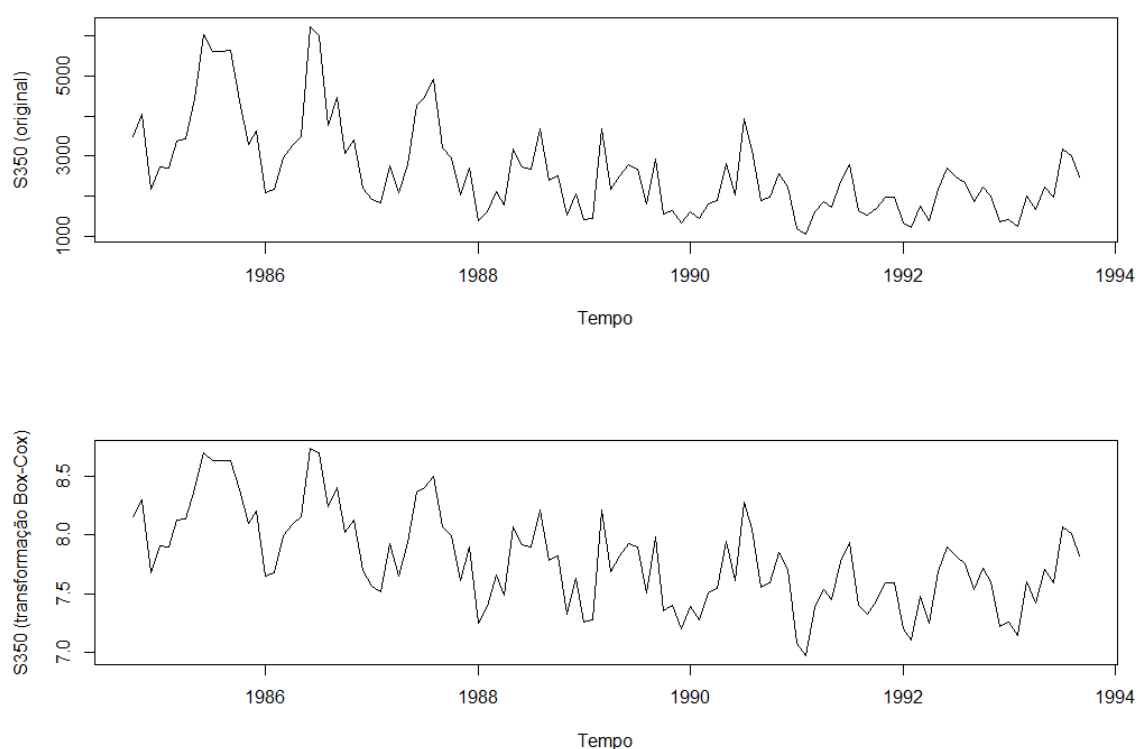


Figura 5 – A primeira imagem mostra a série temporal mensal número 350 da competição M3 original e a segunda imagem mostra a mesma série depois de se realizar a transformação Box-Cox.

Fonte – Elaborada pela autora.

## 3.2 Regressão local

A Regressão local *Locally Estimated Scatterplot Smoothing* (LOESS) é um método de suavização não-paramétrico proposto originalmente por Cleveland (1979) e aprimorado por Cleveland e Devlin (1988). É uma generalização da média móvel e da regressão polinomial, que combina o modelo de regressão múltipla com o algoritmo *K-nearest neighbor* (KNN). Combina a simplicidade das estimativas de mínimos quadrados utilizadas na regressão linear com a flexibilidade presente na regressão não-linear.

No contexto de séries temporais, este método é usado para modelar a relação entre a variável resposta  $X_i$  e a variável preditora  $t_i$ ,  $i = 1, 2, \dots, n$ , que são relacionadas pelo modelo mais simples de regressão, denotado por

$$X_i = f(t_i) + \varepsilon_i, \quad (3.2)$$

em que  $t_i$  é o tempo no instante  $i$ ,  $f(\cdot)$  é uma função desconhecida e  $\varepsilon_i$  denota o termo de erro IID,  $\varepsilon_i \sim N(0, \sigma^2)$ .

Para cada  $t_i$  é estimado o valor de  $f(t_i)$ , modelada por um polinômio de grau baixo, de acordo com um subgrupo de observações mais próximas de  $t_i$ . O polinômio é ajustado utilizando mínimos quadrados ponderados, dando mais peso para as observações mais próximas de  $t_i$  e menos peso para as observações mais distantes. Este processo é bastante similar ao método KNN ponderado (DUDANI, 1976), em que  $h$ , o tamanho de cada subgrupo, é um parâmetro ajustável, que determina a suavidade da estimativa resultante. Valores grandes de  $h$  produzem um maior viés, enquanto valores pequenos produzem uma maior variância.

Não é feita nenhuma suposição global sobre a função  $f(\cdot)$ , mas se assume que localmente ela pode ser bem aproximada pela classe de funções paramétricas, como uma constante ou uma linha reta. O Teorema de Taylor diz que qualquer função contínua pode ser aproximada por polinômios.

Além da formulação original do Teorema de Taylor, enunciado pelo matemático britânico Brook Taylor em 1712<sup>1</sup>, existem outras formas do mesmo que foram enunciadas posteriormente, que são úteis para justificar o procedimento utilizado na regressão local. A formulação original e outras duas versões, retiradas do material disponibilizado por Irizarry e Bravo (2010), são mostradas a seguir.

- Formulação original: Suponha que  $f(\cdot)$  seja um função real derivável  $K$  vezes no intervalo fechado  $[a, b]$  e  $K + 1$  vezes no intervalo aberto  $(a, b)$ . Se  $K \geq 0$ ,  $f^{(K-1)}(\cdot)$  é contínua em  $[a, b]$  e  $f^{(K)}(c)$  é limitada para  $c \in (a, b)$ , então para qualquer ponto distinto  $t_0 < t_1$  em  $[a, b]$  existe um ponto  $t$  entre  $t_0$  e  $t_1$ ,  $t_0 < t < t_1$ , em que

$$f(t_1) = f(t_0) + \sum_{k=1}^{K-1} \frac{f^{(k)}(t_0)}{k!} (t_1 - t_0)^k + \frac{f^{(K)}(t)}{K!} (t_1 - t_0)^K. \quad (3.3)$$

**Nota:** Se  $f(t_0) + \sum_{k=1}^{K-1} \frac{f^{(k)}(t_0)}{k!} (t_1 - t_0)^k$  for observado como uma função de  $t_1$ , observa-se um polinômio da família:  $P_{K+1} = \{f(t) = a_0 + a_1 t + \dots + a_K t^K, (a_0, \dots, a_K)' \in \mathbb{R}^{K+1}\}$ .

- Estatísticos muitas vezes utilizam a chamada forma de Young do Teorema de Taylor: Tome  $f(\cdot)$  tal que  $f^{(K)}(t_0)$  seja limitada por  $t_0$ , então

$$f(t) = f(t_0) + \sum_{k=1}^K \frac{f^{(k)}(t_0)}{k!} (t - t_0)^k + o(|t - t_0|^K), \text{ com } |t - t_0| \rightarrow 0. \quad (3.4)$$

<sup>1</sup> <[https://archive.org/details/UFIE003454\\_TO0324\\_PNI-2529\\_000000/page/24/mode/2up](https://archive.org/details/UFIE003454_TO0324_PNI-2529_000000/page/24/mode/2up)>, acessado em 17/10/2020.

**Nota:** Novamente, observa-se que os dois primeiros termos à direita estão em  $P_{K+1}$ .

- Outra versão bastante utilizada, interessante para visualizar a ideia por trás da regressão local, é a chamada Desigualdade de Jackson: Seja  $f()$  uma função real derivável  $K$  vezes no intervalo fechado  $[a,b]$ , então

$$\min_{g \in P_k} \sup_{t \in [a,b]} |g(t) - f(t)| \leq C \left( \frac{b-a}{2k} \right)^K, \quad (3.5)$$

em que  $P_k$  é o espaço linear de polinômios de grau  $k$  e  $C$  é uma constante qualquer.

### 3.2.1 Parâmetro de suavização

O parâmetro de suavização  $\alpha$ , também conhecido como *span*, é o responsável por controlar o tamanho da vizinhança  $h$  em torno do tempo em questão em que a função de ponderação será aplicada para posteriormente se ajustar o polinômio. Toma-se  $\alpha$  como sendo a porcentagem dos dados que serão utilizados na seleção dos vizinhos mais próximos de um determinado tempo  $t_0$ , ou seja, determina o valor da constante  $h$ . Esse parâmetro possui uma grande influência no viés e variância da estimativa, pois se  $h$  for pequeno o viés da estimativa será reduzido, mas a variância será elevada, e se  $h$  for grande, a estimativa terá um viés maior, mas uma pequena variabilidade.

Valores de  $h$  muito pequenos podem estar relacionados ao *overfitting*: quando o modelo se ajusta demais a uma amostra específica, mas possui baixo poder de generalização. E, valores de  $h$  muito grandes podem estar relacionados ao *underfitting*: caso em que o ajuste não é suficiente para explicar bem os dados. Ambos os casos são muito ruins para a generalização, pois no *overfitting*, o modelo tem um desempenho excelente nos dados de treino, porém quando utilizamos os dados de teste o resultado é ruim, e no *underfitting*, o modelo não consegue aprender o necessário sobre os dados, o que leva à um erro elevado tanto nos dados de treino quanto nos dados de teste (IZBICKI; SANTOS, 2020).

“O objetivo é produzir uma estimativa que seja a mais suave possível mas que não distorça a relação de dependência entre as variáveis em análise” (CLEVELAND; LOADER, 1996). O critério utilizado para a obtenção de  $h$  pode ser o Critério de Informação de Akaike (AIC) ou até mesmo métodos de *plug-in*, que consistem em escrever uma função inicialmente estimada,  $\hat{g}$ , como uma função  $g$  desconhecida e aproximada por uma expansão de Taylor ou outra expansão assintótica, de maneira a minimizar alguma medida de acurácia. No contexto de séries temporais, o valor de  $h$  precisa necessariamente ser par (HYNDMAN; KHANDAKAR, 2008; HYNDMAN *et al.*, 2020).

### 3.2.2 Grau do polinômio local

Assim como o parâmetro de suavização descrito na seção anterior, a escolha do grau  $d$  do polinômio ajustado afeta a relação entre a variância e o viés: quanto maior for o grau polinomial

menor será o viés e maior será a variância, quando fixado um parâmetro de suavização. Modelos mais complexos, com muitos parâmetros ( $d$  grande), podem estar relacionados ao *overfitting* e modelos mais simples, com poucos parâmetros ( $d$  pequeno), existe a possibilidade de *underfitting*.

Usualmente, os polinômios escolhidos na regressão local são sempre de primeiro ou segundo grau ( $d = 1$  ou  $d = 2$ ), ou seja, possuem forma linear ou quadrática. Se escolhido um polinômio de grau zero, o procedimento será nada mais do que uma média móvel ponderada. De forma geral, a presença de “picos” ou “vales” nos dados pode ser um indicativo que  $d$  deve ser tomado como sendo igual a dois ou três, enquanto que a presença de um comportamento único indica  $d = 1$ . Deve-se sempre lembrar que polinômios de grau  $d$  requerem pelo menos  $d + 1$  valores para que possa ser realizado o ajuste.

### 3.2.3 Função de ponderação

Esta função é utilizada para definir a ponderação (os pesos) que os  $h$  vizinhos mais próximos do tempo a ser estimado,  $t_0$ , terão. Ela atribui peso maior para tempos mais próximos de  $t_0$  e peso menor para as observações mais distantes. Utiliza-se a ideia de que quanto mais os tempos são próximos, mais eles são relacionados. Então, serão atribuídos pesos para os  $h$  tempos mais próximos de  $t_0$ , ou seja, as observações pertencentes ao intervalo  $[t_0 - h(t_0), t_0 + h(t_0)]$ , para se obter a estimativa de  $f(t_0)$ .

Sejam  $t_{01}, \dots, t_{0h}$  os tempos referentes aos  $h$  vizinhos mais próximos de  $t_0$ , em que  $t_{01}$  é o mais próximo,  $t_{02}$  o segundo mais próximo, e assim por diante, até chegar no mais distante dentre os  $h$  vizinhos mais próximos, o  $t_{0h}$ . Define-se a variável transformada  $u_i$  como

$$u_i = \frac{(t_{0i} - t_0)}{h(t)}, \quad (3.6)$$

em que  $h(t)$  é a distância máxima obtida na vizinhança, ou seja, a distância entre o tempo  $t_0$  e o tempo mais distante dentre os  $h$  vizinhos mais próximos a ele.

A função de ponderação tradicionalmente utilizada neste método, e que foi implementada neste estudo, é a tri-cúbica (GUTHRIE, 2012), definida como

$$W(u) = \begin{cases} (1 - |u|^3)^3, & |u| \leq 1, \\ 0, & |u| > 1. \end{cases} \quad (3.7)$$

A sequência de pesos, contidos no intervalo  $[0, 1]$ , é definida como

$$w_i = W(u_i). \quad (3.8)$$

Utilizando os pesos  $w_i$  obtidos, a regressão local LOESS utiliza o método dos mínimos quadrados ponderados para estimar  $\hat{f}(t_0)$ , que é aproximada por uma função polinomial. Por exemplo, no caso de um polinômio de grau 1 ajustado, deve-se encontrar os parâmetros  $\hat{\beta} = (\beta_0, \beta_1)'$  que minimizem

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^h w_i(t_0) [X_i - \beta_0 - \beta_1(t_i - t_0)]^2. \quad (3.9)$$

E obtém-se assim a reta estimada pela regressão local. Para polinômios de grau  $d$ , os parâmetros a serem encontrados são  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_d)'$ . No *software* R, a função `loess()` do pacote `stats` (R Core Team, 2020) realiza o procedimento descrito nesta seção.

Para ilustrar o procedimento, utilizou-se o conjunto de dados que contém a temperatura média anual em graus Fahrenheit em New Haven, Connecticut, de 1912 a 1971. Aplicou-se o método de regressão local LOESS, atribuindo diferentes valores para o parâmetro de suavização, sendo eles  $\alpha = (0,8, 0,5, 0,2)$ . Observa-se que quanto maior é o valor de  $\alpha$  mais suave é a curva e quanto menor é o valor de  $\alpha$  a curva se torna mais próxima de interpolar os dados. A Figura 6 apresenta os resultados obtidos.

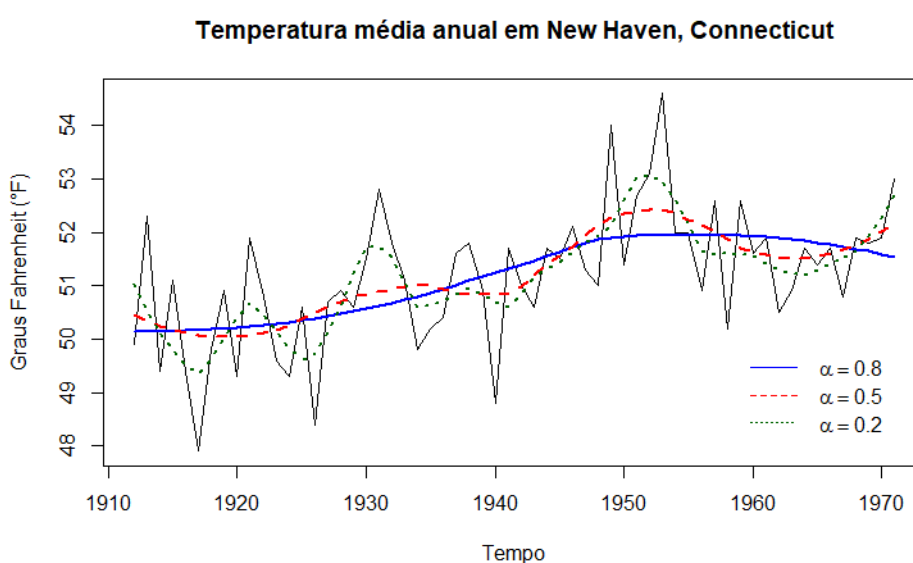


Figura 6 – Regressão local LOESS aplicada a um conjunto de dados que contém a temperatura média anual em graus Fahrenheit em New Haven, Connecticut, de 1912 a 1971. Utilizou-se três diferentes valores para o parâmetro de suavização  $\alpha$ .

Fonte – Figura elaborada pela autora.

### 3.3 Decomposição STL

*Seasonal and Trend decomposition using LOESS* (STL) é um método de decomposição de séries temporais, proposto por Cleveland *et al.* (1990), que decompõe a série temporal em três componentes: tendência, sazonalidade e componente aleatório.

A **tendência** de uma série temporal é definida como um padrão de crescimento ou decréscimo da variável de interesse em um período de tempo. A **sazonalidade** é a presença de variações que ocorrem em períodos regulares e específicos de tempo inferiores a 1 ano, podendo ser semanal, mensal, trimestral, etc. Já o **componente aleatório** consiste nos resíduos da série temporal depois que a tendência e a sazonalidade são extraídos.

A decomposição utilizada é a aditiva, ou seja, a variação sazonal é tomada como constante, e segue o seguinte modelo

$$X_t = T_t + S_t + a_t, \quad (3.10)$$

em que, para uma série temporal de tamanho  $n$ ,  $X_t$  representa as observações da série temporal,  $T_t$  a tendência,  $S_t$  a sazonalidade e  $a_t$  o componente aleatório, que possui média zero e variância  $\sigma_a^2$ , para  $t = 1, \dots, n$ .

Os passos do algoritmo STL são descritos por [Cleveland \*et al.\* \(1990\)](#). A ideia principal é descrita no [Algoritmo 2](#).

---

#### Algoritmo 2 – Algoritmo STL

---

- 1: Retira-se a tendência da série;
  - 2: É realizada a suavização, baseada em LOESS, para cada período da série, separadamente;
  - 3: Junta-se toda a série novamente e é realizada uma suavização que consiste em filtros baseados em médias móveis e LOESS. Utilizando a função `stl()`, no pacote *stats* ([R Core Team, 2020](#)) do *software* R, caso o parâmetro “*s.window*” esteja definido como “*periodic*”, então essa suavização é substituída pela média de cada ponto, utilizando todos os períodos extraídos da série;
  - 4: Retira-se novamente a tendência da série sazonal suavizada, caso exista;
  - 5: Retira-se a sazonalidade da série original, utilizando o componente sazonal calculado nos passos anteriores;
  - 6: A série sem sazonalidade é então suavizada através do método LOESS e a tendência é extraída.
- 

Para encontrar o componente aleatório da série, a tendência e a sazonalidade estimadas são retiradas da série original. Os graus do polinômio utilizados no método LOESS nos passos 3 e 4 são  $d = 1$  e  $d = 0$  no passo 2. A [Figura 7](#) ilustra o procedimento da aplicação da decomposição STL em um conjunto de dados que contém números mensais de passageiros em voos internacionais, de 1949 a 1960. A [Figura 8](#) ilustra o procedimento da aplicação da transformação Box-Cox, seguida da decomposição STL no mesmo conjunto de dados.

Outra função interessante que também pode ser utilizada para decompor séries temporais é a `mstl()`, do pacote *forecast* ([R Core Team, 2020](#)) do *software* R, uma extensão da função `stl()` definida anteriormente, que permite a decomposição de séries temporais com vários padrões sazonais.

## 3.4 Reamostragem em blocos

Conforme estudado, implementado e discutido no [Capítulo 2](#), o método de reamostragem utilizado neste estudo será o *moving block bootstrap*, ou MBB.

A [Figura 9](#) ilustra o procedimento de reamostragem MBB. Como a reamostragem é feita utilizando apenas o componente aleatório da série, primeiramente aplica-se a transformação Box-Cox, seguida da decomposição STL, e depois que as séries são reamostradas, adiciona-se



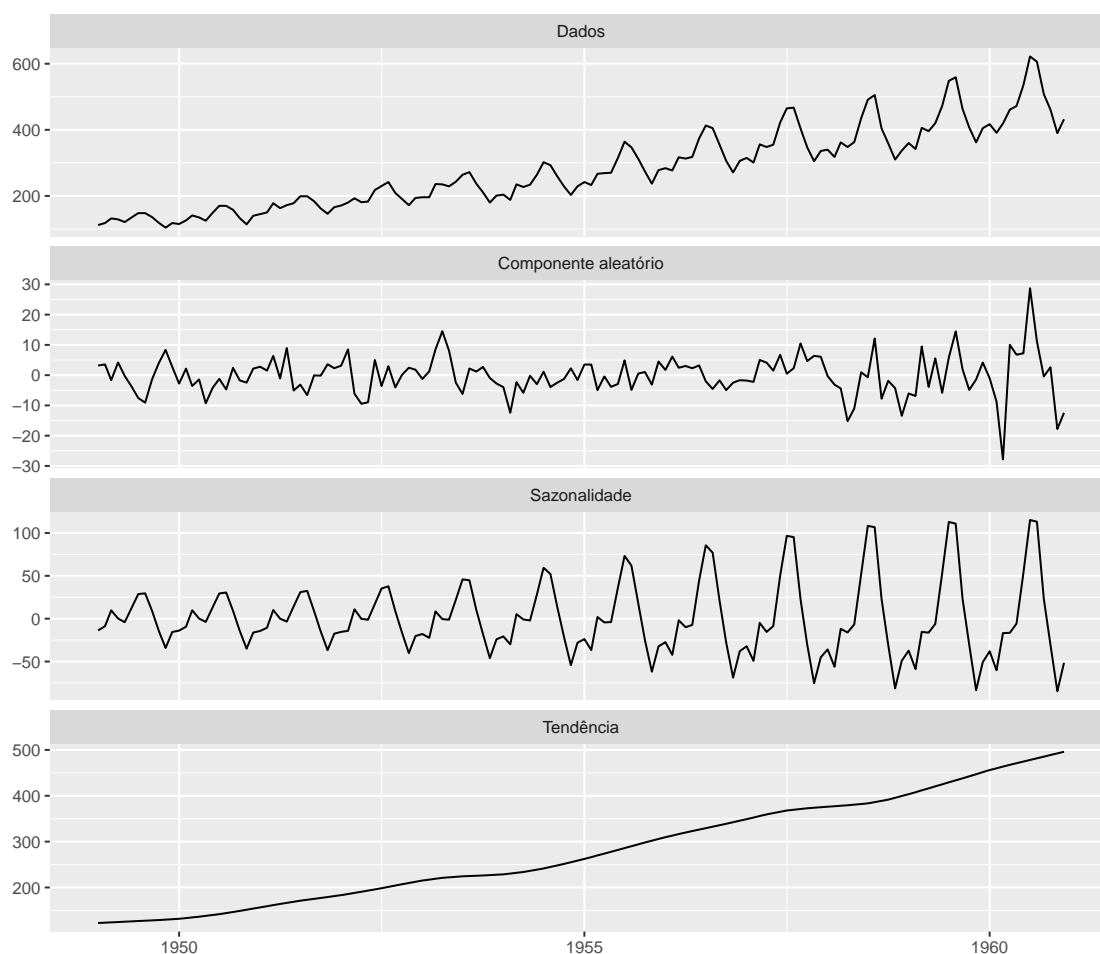


Figura 7 – Decomposição STL aplicada no conjunto de dados de números mensais de passageiros em voos internacionais, de 1949 a 1960, utilizando a série original.

Fonte – Elaborada pela autora.

a tendência e a sazonalidade e reverte-se a transformação Box-Cox. Foram geradas  $B = 100$  séries *bootstrap* com tamanho de bloco  $l = 24$  a partir da série temporal mensal número 350 da competição M3.

## 3.5 Modelos de séries temporais

### 3.5.1 Modelo autorregressivo

Os modelos autorregressivos, ou AR, são os definidos na [Subseção 2.4.1](#).

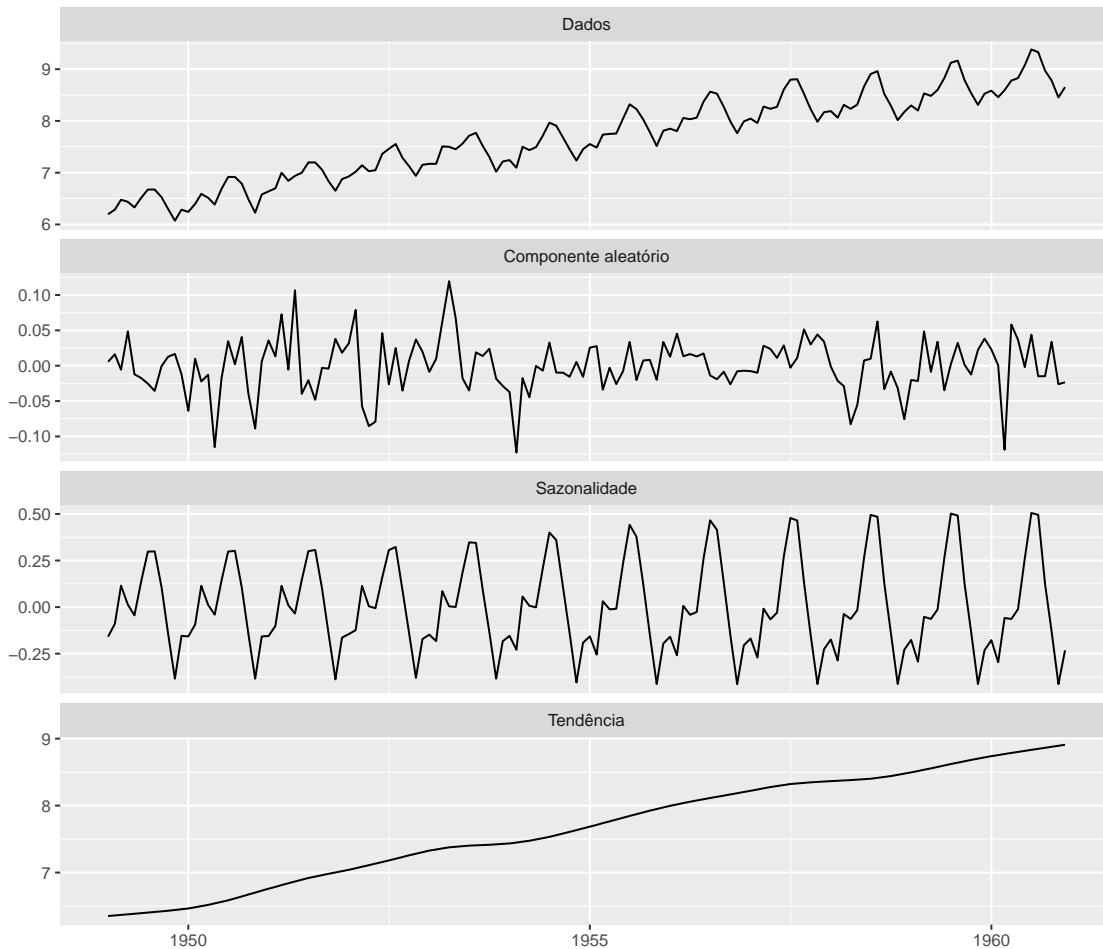


Figura 8 – Decomposição STL aplicada no conjunto de dados de números mensais de passageiros em voos internacionais, de 1949 a 1960, utilizando a série transformada (Box-Cox).

Fonte – Elaborada pela autora.

### 3.5.2 Modelo de médias móveis

O processo  $X_t$  é chamado de Modelo de médias móveis (MA) de ordem  $q$ , denotado por  $MA(q)$  (BOX *et al.*, 2015), se satisfizer a equação de diferenças

$$X_t = \theta_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (3.11)$$

em que  $\theta_0$  é constante,  $(\theta_1, \dots, \theta_q)$  são parâmetros reais,  $a_t$  é IID e  $a_t \sim N(0, \sigma_a^2)$ .

Também pode ser escrito em termos do operador de retardo como

$$X_t = \theta_0 + \theta(B)a_t, \quad (3.12)$$

em que  $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ .

O processo  $MA(q)$  é definido como sendo sempre estacionário, e é invertível se a equação característica  $\theta(B) = 0$  tiver todas as raízes fora do círculo unitário. As estimativas de  $(\theta_1, \dots, \theta_q)$  foram obtidas neste trabalho via máxima verossimilhança.

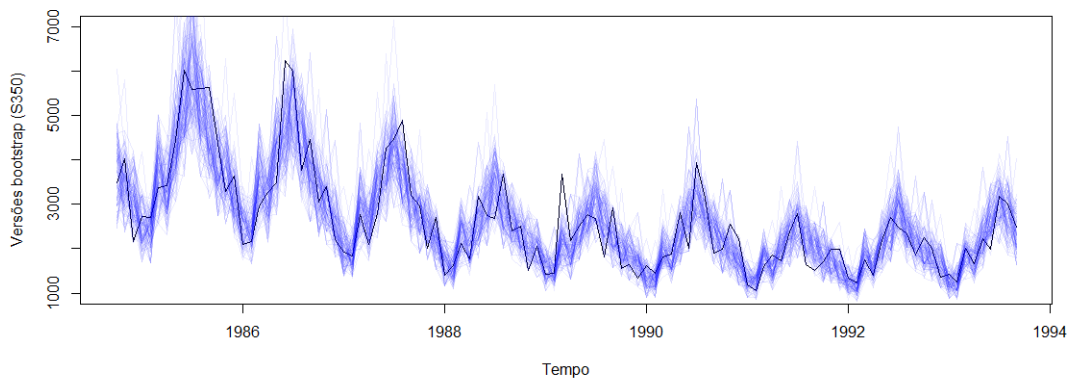


Figura 9 – Versões *bootstrap* (azul) da série temporal mensal número 350 da competição M3 (preto). São mostradas  $B = 100$  séries *bootstrap*, com tamanho de bloco  $l = 24$ .

Fonte – Elaborada pela autora.

### 3.5.3 Modelo autorregressivo e de médias móveis

Descrito pelo matemático neozelandês [Whittle \(1951\)](#) e popularizado pelos estatísticos britânicos [Box e Jenkins \(1970\)](#), o Modelo Autorregressivo e de Médias Móveis (ARMA) é a junção dos modelos AR e MA. A parte AR do processo envolve regressar a variável em seus próprios valores defasados, enquanto a parte MA envolve modelar o termo de erro como uma combinação linear de termos de erro que ocorreram no passado.

Dizemos que um processo  $X_t$  é um modelo autorregressivo e de médias móveis de ordem  $p$  e  $q$ , denotado por  $ARMA(p,q)$ , se satisfizer a seguinte forma

$$X_t = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (3.13)$$

em que  $\alpha$  é constante,  $(\phi_1, \dots, \phi_p)$  e  $(\theta_1, \dots, \theta_q)$  são parâmetros reais,  $a_t$  é IID e  $a_t \sim N(0, \sigma_a^2)$ .

Também pode ser escrito em termos de seu operador de retardo como

$$\phi(B)X_t = \alpha + \theta(B)a_t, \quad (3.14)$$

em que  $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  e  $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ .

A estacionariedade do processo exige que as raízes de  $\phi(B) = 0$  estejam todas fora do círculo unitário, e a invertibilidade requer a mesma condição para as raízes de  $\theta(B) = 0$ . As estimativas de  $p$  e  $q$  foram obtidas neste trabalho via máxima verossimilhança.

### 3.5.4 Modelo autorregressivo, integrado e de médias móveis

O Modelo Autorregressivo, Integrado e de Médias Móveis (ARIMA) foi introduzido por [Box e Jenkins \(1970\)](#), e é uma generalização do modelo ARMA. É aplicado em casos em que os dados possuem evidências de não-estacionariedade, ou seja, apresentam tendência e/ou sazonalidade, e que diferenciações da mesma (correspondente à parte integrada do modelo) pode eliminar essa não-estacionariedade.

O modelo ARIMA para dados não-sazonais é denotado por

$$ARIMA(p, d, q), \quad (3.15)$$

em que  $p$  representa a ordem do modelo autorregressivo não-sazonal,  $d$  representa a ordem de integração (ou diferenciação) não-sazonal e  $q$  representa a ordem do modelo de médias móveis não-sazonal.

O modelo ARIMA que também leva em conta a sazonalidade dos dados é chamado de SARIMA (modelo ARIMA sazonal), e é denotado por

$$ARIMA(p, d, q)(P, D, Q)s, \quad (3.16)$$

em que os parâmetros  $p$ ,  $d$  e  $q$  são definidos como na [Equação 3.15](#),  $P$  representa a ordem do modelo autorregressivo sazonal,  $D$  representa a ordem de integração (ou diferenciação) sazonal,  $Q$  representa a ordem do modelo de médias móveis sazonal e  $s$  é o número de observações por ano.

A série temporal agora pode ser diferenciada tanto em seu componente sazonal quanto em seu componente não-sazonal, a fim de se obter uma série estacionária, que pode ser denotada como um modelo ARMA ([3.13](#)).

A modelagem de um processo SARIMA pode ser escrita, de forma compacta e em termos dos operadores de retardos, como

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D X_t = \alpha + \Theta(B^s)\theta(B)a_t, \quad (3.17)$$

em que  $\alpha$  é constante,  $\phi(B)$  representa os coeficientes não-sazonais do processo autorregressivo,  $\theta(B)$  representa os coeficientes não-sazonais do processo de médias móveis,  $(1-B)^d$  é o operador diferença de ordem  $d$  para a diferenciação não-sazonal da série,  $\Phi(B^s)$  representa os coeficientes sazonais do processo autorregressivo,  $\Theta(B^s)$  representa os coeficientes sazonais do processo de médias móveis,  $(1-B^s)^D$  é o operador diferença de ordem  $D$  para a diferenciação sazonal da série e  $a_t \sim N(0, \sigma_a^2)$  é IID.

Vale ressaltar que o modelo ARIMA, com sazonalidade ou não, goza das mesmas propriedades do modelo ARMA( $p, q$ ). A estimação dos parâmetros do modelo foi obtida neste trabalho via máxima verossimilhança.

### 3.6 Previsão de modelos ARIMA sazonais

A previsão de modelos ARIMA sazonais é feita de maneira análoga à realizada em modelos ARIMA não-sazonais. Segundo [Brockwell \(1991\)](#), os operadores de diferenças podem ser aplicados para transformar a série temporal não-estacionária  $X_t$  no processo estacionário  $X_t^*$  utilizando a seguinte equação

$$X_t^* = (1-B)^d(1-B^s)^D X_t \quad (3.18)$$

Box, Jenkins e Reinsel (2008) propôs que o modelo SARIMA pode ser visto meramente como um caso especial de um modelo ARMA( $p + sP, q + sQ$ ), escrito como

$$\phi(B)^* X_t^* = \alpha + \theta(B)^* a_t, \quad (3.19)$$

em que a parte AR  $\phi(B)^*$  neste modelo é igual ao produto  $\Phi(B^s)\phi(B)$  do modelo SARIMA e a parte MA  $\theta(B)^*$  corresponde a  $\Theta(B^s)\theta(B)$ .

Sendo assim, a previsão de um modelo SARIMA pode ser realizada a partir da previsão de um modelo ARMA. A equação de previsão da observação  $h$  passos à frente de um modelo ARMA( $p, q$ ) é dada pela seguinte equação

$$\hat{X}_{t+h} = \hat{\alpha} + \hat{\phi}_1 X_{t+h-1} + \dots + \hat{\phi}_p X_{t+h-p} + a_{t+h} - \hat{\theta}_1 a_{t+h-1} - \dots - \hat{\theta}_q a_{t+h-q}, \quad (3.20)$$

em que os valores de  $X_t$  podem ser estimados, caso não estejam disponíveis. Os valores de  $a_t$  que não estão disponíveis são atribuídos como sendo iguais a zero.

### 3.7 Critério de informação de Akaike corrigido

O Critério de Informação de Akaike corrigido (AICc) é uma medida relativa da qualidade de ajuste de um modelo estocástico estimado, definido como

$$AICc = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{n - p - q - k - 2}, \quad (3.21)$$

em que  $p$  e  $q$  são os parâmetros do modelo ARIMA,  $k = 0$  se  $\alpha = 0$  e  $k = 1$  se  $\alpha \neq 0$ ,  $n$  é o número de observações da série, e o Critério de Informação de Akaike (AIC) é definido como

$$AIC = -2\log(L) + 2(p + q + k + 1), \quad (3.22)$$

em que  $L$  é o máximo da função de verossimilhança do modelo avaliada nas estimativas de máxima verossimilhança.

O AICc é indicado para lidar em situações em que a razão  $n/(p + q + k)$  é pequena ( $< 40$ ). Sakamoto, Ishiguro e Kitagawa (1986) e Sugiura (1978), afirmam que o AIC pode ter um desempenho ruim se o número de parâmetros for grande em relação ao tamanho da amostra, assim o AICc é apenas uma correção, de segunda ordem, do viés do AIC.

É importante notar que o AIC e o AICc tendem a não ser bons guias para selecionar a ordem apropriada de diferenciação de um modelo, mas apenas para selecionar os valores de  $p$  e  $q$ . Isso ocorre porque a diferenciação altera os dados nos quais a probabilidade é calculada, tornando os valores AIC entre os modelos com diferentes ordens de diferenciação não comparáveis.

### 3.8 Método *bagging* para previsão de séries temporais

O método *bootstrap aggregating*, mais conhecido como *bagging*, foi proposto por Breiman (1996) e está sendo bastante utilizado na área de Estatística Clássica e em Aprendizagem

Estatística. Este procedimento consiste em gerar conjuntos de dados por reamostragem *bootstrap*, que serão submetidos a algum procedimento de interesse e seus resultados combinados em um único resultado final. Os procedimentos mais comuns são técnicas de classificação, como por exemplo, árvores de decisões. O *bagging* é conhecido por reduzir a variância das previsões, ou seja, melhorar a precisão do que se deseja prever.

Este método também pode trazer benefícios na acurácia de previsões de séries temporais, embora ainda seja pouco explorado neste campo. A reamostragem é feita aplicando-se a técnica *bootstrap* em blocos no componente aleatório da série original de treino, e então as previsões de cada série reamostrada são combinadas de modo a se obter uma única previsão final, como ilustra a [Figura 10](#).

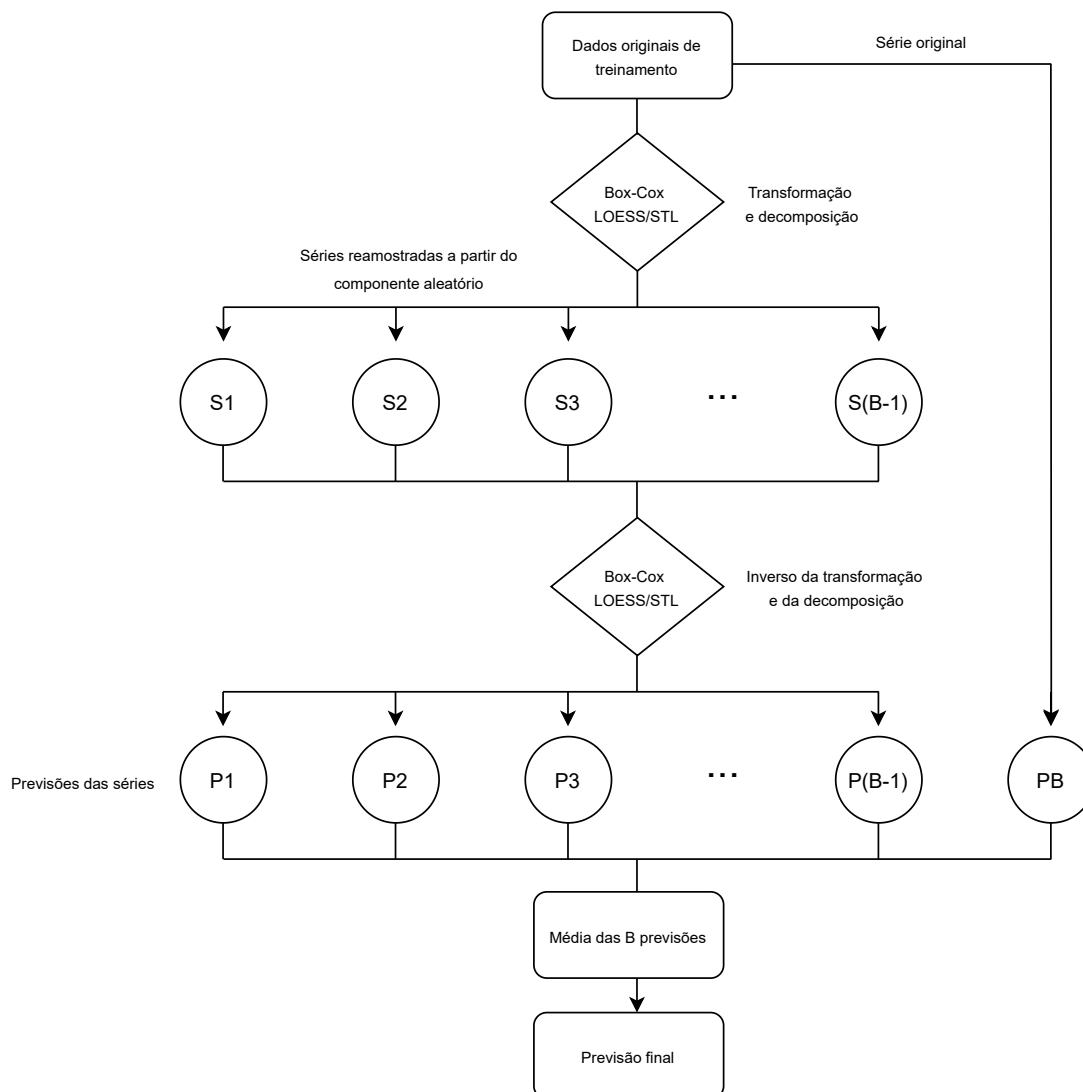


Figura 10 – Fluxograma *bagging*.

Fonte – Elaborada pela autora.

O procedimento mais detalhado e completo é descrito de acordo com o [Algoritmo 3](#).

**Algoritmo 3** – Método *bagging* para previsão de séries temporais

- 1: Divide-se a série temporal original em série de treinamento e série de teste. A série de treinamento será utilizada para obter as previsões, e a de teste para medir sua acurácia;
- 2: Aplica-se a transformação Box-Cox na série de treinamento com o intuito de estabilizar sua variância e torná-la mais simétrica. Séries com valores negativos ou iguais a zero não são transformadas;
- 3: A série de treinamento é decomposta em três componentes: tendência, sazonalidade e componente aleatório. Se a série não for sazonal, a decomposição utilizada é a LOESS, caso contrário utiliza-se a decomposição STL;
- 4: Aplica-se o método de reamostragem em blocos MBB no componente aleatório da série, gerando  $B - 1$  séries reamostradas;
- 5: Adiciona-se a tendência e a sazonalidade, e aplica-se o inverso da transformação Box-Cox em cada uma das  $B - 1$  séries, resultando em novas séries reamostradas a partir da série original de treinamento;
- 6: Cada uma das  $B$  séries, as  $B - 1$  séries reamostradas mais a série original de treinamento, são modeladas a partir de um modelo ARIMA. A função `auto.arima()` disponível no pacote *forecast* (HYNDMAN; KHANDAKAR, 2008) do *software* R foi utilizada para realizar a seleção do melhor modelo em cada caso, com o número máximo de diferenciações igual a 1;
- 7: Realiza-se então as previsões  $h$  passos à frente para cada uma das  $B$  séries e calcula-se as médias das  $B$  previsões para cada um dos  $h$  passos.

## 3.9 Medidas de acurácia

### 3.9.1 Erro quadrático médio

O MSE e o RMSE, são definidos na [Subseção 2.4.3](#).

### 3.9.2 Erro percentual absoluto médio arcotangente

O Erro Percentual Absoluto Médio Arcotangente (MAAPE) é uma extensão do Erro Percentual Absoluto Médio (MAPE), muito usado em estatística por ser de fácil interpretação. O MAPE expressa a porcentagem média dos erros, em valor absoluto, cometidos na previsão da série temporal; quanto menor esta medida, melhor será a previsão. A equação do MAPE é definida como a média do erro absoluto percentual, e é dada por

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|X_t - \hat{X}_t|}{|X_t|}, \quad (3.23)$$

em que  $|X_t|$  é o módulo do valor verdadeiro da série no instante  $t$ ,  $|\hat{X}_t|$  é o módulo do valor previsto da série no instante  $t$ , e  $n$  é a quantidade de observações previstas.

Este erro possui a desvantagem de que, quando a série possui valores próximos ou iguais a zero, o valor MAPE tenderá para infinito. Para resolver este problema o MAAPE foi proposto, e é definido como

$$MAAPE = \frac{100}{n} \sum_{t=1}^n (AAPE_t), \quad (3.24)$$

em que

$$AAPE_t = \arctan \left| \frac{X_t - \hat{X}_t}{X_t} \right|. \quad (3.25)$$

A função arcotangente é definida para todos os valores reais  $(-\infty, \infty)$  e  $\lim_{x \rightarrow \infty} \tan^{-1} x = \pi/2$ , e então  $AAPE_t$  está definido no intervalo  $[0, \pi/2]$ . O MAAPE apresenta uma penalidade mais equilibrada entre erros negativos e positivos e é mais robusto para lidar com *outliers* do que o MAPE.

### 3.9.3 Erro percentual absoluto médio simétrico

O Erro Percentual Absoluto Médio Simétrico (sMAPE) também é uma extensão do MAPE (3.23). A medida MAPE atribui uma maior penalidade em erros negativos ( $X_t < \hat{X}_t$ ) do que em erros positivos ( $\hat{X}_t < X_t$ ). Makridakis (1993) afirma que “erros iguais acima do valor real resultam em um valor de MAPE maior do que aqueles abaixo do valor real”, e utiliza um exemplo para mostrar este fato, em que  $X_t = 150$  e  $\hat{X}_t = 100$ , que resulta em um erro relativo igual a  $50/150 = 0,33$ , em contraste com a situação em que  $X_t = 100$  e  $\hat{X}_t = 150$ , que resulta em um erro relativo igual a  $50/100 = 0,50$ .

O sMAPE foi proposto para evitar o problema de assimetria dos erros positivos e negativos. Sua primeira versão foi escrita por Armstrong (1985), sofreu algumas alterações ao longo dos anos e finalmente chegou na forma mais utilizada nos dias atuais, que é dada por

$$sMAPE = \frac{100}{n} \sum_{t=1}^n \frac{|X_t - \hat{X}_t|}{(|X_t| + |\hat{X}_t|)/2}, \quad (3.26)$$

em que  $|X_t|$  é o módulo do valor verdadeiro da série no instante  $t$ ,  $|\hat{X}_t|$  é o módulo do valor previsto da série no instante  $t$ , e  $n$  é a quantidade de observações previstas.

A métrica sMAPE é definida no intervalo  $[0, 200]$ . Apesar de corrigir o problema de assimetria, o sMAPE ainda apresenta algumas desvantagens similares ao MAPE, como por exemplo, quando um dos valores, reais ou preditos, são iguais a zero, seu valor será 200, e quando ambos são iguais a zero, seu valor é indefinido. Apesar desta desvantagem, esta é uma métrica muito utilizada nos dias atuais para avaliar a acurácia de previsões de séries temporais e muito válida quando os valores da série não são iguais a zero.



---

## SIMULAÇÃO

---

Com o objetivo de avaliar o desempenho da acurácia do método *bagging* para o aprimoramento de previsões de séries temporais, foram realizados três estudos de simulação. No primeiro estudo (simulação 1), foram simuladas  $N = 100$  séries (fixando-se as sementes 1 a 100 no *software R*) provenientes de um modelo autorregressivo de ordem dois,  $\phi = (0,6; -0,2)$ , de tamanho  $n = 100$  cada. As séries simuladas foram divididas em treinamento, primeiras 80 observações, e teste, últimas 20 observações, todas tomadas em sequência no tempo. No segundo estudo (simulação 2), as séries foram simuladas a partir de um modelo de médias móveis de ordem três,  $\theta = (0,6; -0,5; 0,4)$ , e no terceiro estudo (simulação 3), as séries foram simuladas a partir da soma de um modelo autorregressivo de ordem dois,  $\phi = (0,6; 0,2)$ , com um modelo de médias móveis de ordem três,  $\theta = (0,6; -0,5; 0,4)$ , ambos sob as mesmas condições descritas anteriormente.

A [Figura 11](#) mostra as 3 primeiras séries simuladas a partir da simulação 1 e suas respectivas variâncias amostrais. A média das variâncias amostrais das  $N = 100$  séries é igual a 1,39. A [Figura 12](#) mostra as 3 primeiras séries simuladas a partir da simulação 2 e suas respectivas variâncias amostrais. A média das variâncias amostrais das  $N = 100$  séries é igual a 1,75. A [Figura 13](#) mostra as 3 primeiras séries simuladas a partir da simulação 3 e suas respectivas variâncias amostrais. A média das variâncias amostrais das  $N = 100$  séries é igual a 3,96. Nota-se que a média das variâncias amostrais das séries da simulação 1 é menor do que a média das variâncias amostrais das séries da simulação 2, que por sua vez é menor do que a média das variâncias amostrais das séries da simulação 3.

Para cada uma das séries, a partir da série de treinamento, foram geradas primeiramente  $B = 10$  e posteriormente  $B = 100$  séries *bootstrap* com tamanho de bloco  $l = (1, 5, 10, 15)$ , que foram combinadas a fim de se obter a previsão final. A partir da série de teste, calculou-se o número de séries em que cada modelo, individual e combinado, apresentou melhor previsão para os diferentes tamanhos de blocos, além da média e da mediana do MSE e do MAAPE. Os

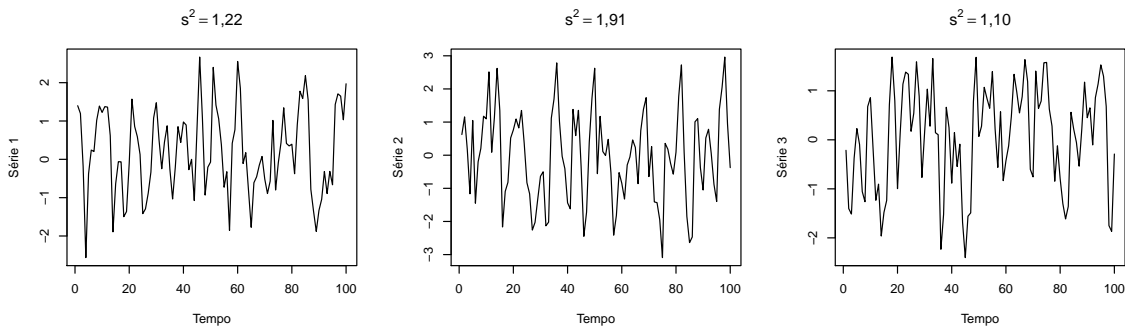


Figura 11 – Séries simuladas a partir de um modelo autorregressivo de ordem dois, com parâmetros fixados, e suas respectivas variâncias.

Fonte – Elaborada pela autora.

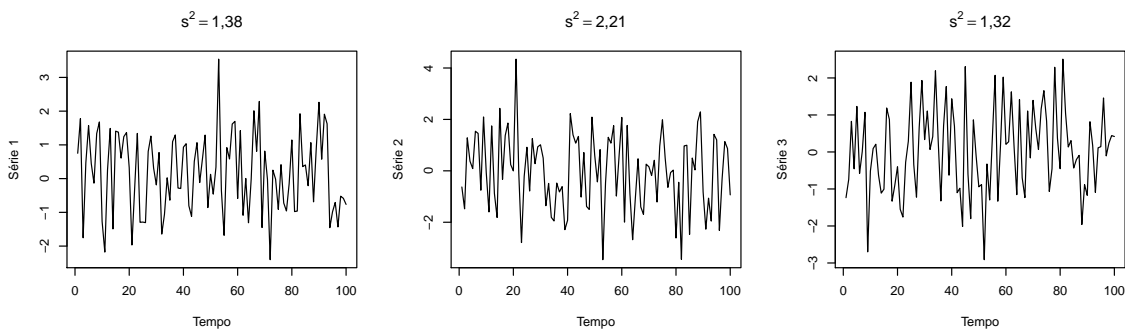


Figura 12 – Séries simuladas a partir de um modelo de médias móveis de ordem três, com parâmetros fixados, e suas respectivas variâncias.

Fonte – Elaborada pela autora.

resultados são apresentados na [Tabela 2](#), [Tabela 3](#), [Tabela 4](#) e [Tabela 5](#). Os modelos foram ajustados aos dados de treinamento de cada série segundo a função `auto.arima()` do *software* R, que combina testes de raiz unitária, minimização do AICc e estimação de máxima verossimilhança para obter o melhor modelo. Sob as condições descritas, o tempo computacional aproximado por simulação foi de 15 minutos para  $B = 10$  e 40 minutos para  $B = 100$ .

Tabela 2 – Número de séries simuladas de diferentes maneiras que obtiveram desempenho superior em cada modelo, assumindo  $B = 10$  séries *bootstrap*, diferentes medidas de acurácia e diferentes tamanhos de blocos utilizados no processo de reamostragem.

Medida de acurácia	Tamanho do bloco (Método individual não se aplica)	Simulação 1		Simulação 2		Simulação 3	
		Individual	Combinado	Individual	Combinado	Individual	Combinado
MSE	$l = 1$	54	46	48	52	44	56
	$l = 5$	56	44	52	48	48	52
	$l = 10$	57	43	49	51	44	56
	$l = 15$	52	48	50	50	42	58
MAAPE	$l = 1$	58	42	53	47	51	49
	$l = 5$	58	42	54	46	56	44
	$l = 10$	60	40	58	42	49	51
	$l = 15$	58	42	51	49	55	45

Fonte – Elaborada pela autora.

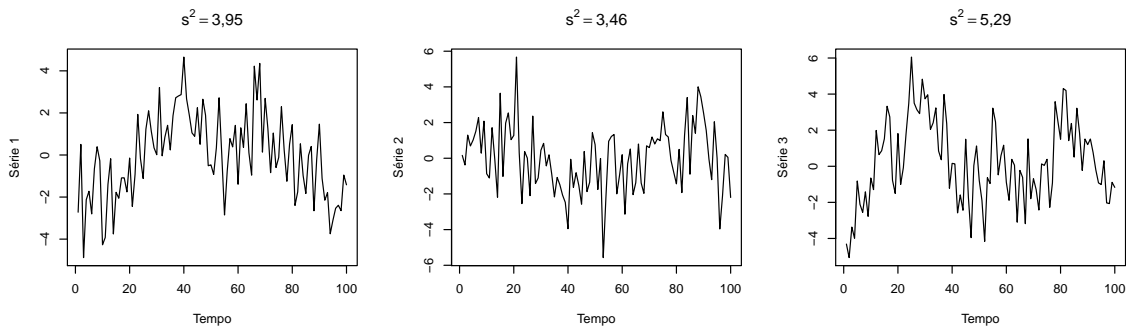


Figura 13 – Séries simuladas a partir da soma de uma série proveniente de um modelo autorregressivo de ordem dois com uma série proveniente de um modelo de médias móveis de ordem três, ambas com parâmetros fixados, e suas respectivas variâncias.

Fonte – Elaborada pela autora.

Tabela 3 – Número de séries simuladas de diferentes maneiras que obtiveram desempenho superior em cada modelo, assumindo  $B = 100$  séries *bootstrap*, diferentes medidas de acurácia e diferentes tamanhos de blocos utilizados no processo de reamostragem.

Medida de acurácia	Tamanho do bloco (Método individual não se aplica)	Simulação 1		Simulação 2		Simulação 3	
		Individual	Combinado	Individual	Combinado	Individual	Combinado
MSE	$l = 1$	60	40	46	54	40	60
	$l = 5$	59	41	49	50	46	54
	$l = 10$	58	42	44	56	44	56
	$l = 15$	58	42	48	52	46	54
MAAPE	$l = 1$	56	44	58	42	47	53
	$l = 5$	60	40	57	43	50	50
	$l = 10$	60	40	50	50	52	48
	$l = 15$	61	39	57	43	48	52

Fonte – Elaborada pela autora.

A [Tabela 2](#) e a [Tabela 3](#) apresentam comparações entre os desempenhos dos métodos individual e combinado ao realizar previsões. Na [Tabela 2](#), no caso em que as séries foram geradas a partir da simulação 1, assumindo-se  $B = 10$  séries *bootstrap* e tamanho de bloco  $l = 1$ , tem-se que, em todas as  $N = 100$  séries de treinamento, o MSE do método combinado calculado a partir das séries de teste obteve menor valor apenas em 46 séries. As 54 séries restantes, apresentaram menor valor de MSE ao utilizar o método individual. As interpretações são análogas para os demais valores da [Tabela 2](#) e da [Tabela 3](#).

Para  $B = 10$ , na simulação 1, para a maioria das séries, o método individual obteve melhor performance em relação ao método combinado. A média e a mediana do MSE e do MAAPE também obtiveram melhores resultados para o modelo individual. Na simulação 2, de acordo com o MSE, o método individual obteve desempenho equivalente ao método combinado para a maioria das séries, e de acordo com o MAAPE, o método individual obteve desempenho levemente superior ao combinado para a maioria das séries, porém, a média e a mediana do MSE e a média do MAAPE obtiveram melhores resultados para o modelo combinado, somente a mediana do MAAPE que favoreceu o método individual. Na simulação 3, de acordo com o

Tabela 4 – Média e mediana das medidas de acurácia das séries, simuladas de diferentes maneiras, assumindo  $B = 10$  séries *bootstrap* e diferentes tamanhos de blocos utilizados no processo de reamostragem.

Medida de acurácia	Tamanho do bloco	Simulação 1		Simulação 2		Simulação 3	
		Mediana	Média	Mediana	Média	Mediana	Média
MSE	Individual	1,36	1,42	1,69	1,79	4,15	4,97
	$l = 1$	1,33	1,46	1,62	1,77	4,14	4,81
	$l = 5$	1,48	1,48	1,63	1,78	4,19	4,87
	$l = 10$	1,41	1,47	1,67	1,77	4,03	4,89
	$l = 15$	1,40	1,47	1,67	1,77	4,06	4,88
MAAPE	Individual	78,20	78,53	78,54	79,15	79,40	82,02
	$l = 1$	78,71	79,07	78,77	78,65	79,82	81,48
	$l = 5$	78,47	79,64	78,64	78,92	80,58	81,92
	$l = 10$	78,75	79,93	78,72	79,11	80,08	81,96
	$l = 15$	78,38	79,62	78,77	78,92	80,23	81,98

Fonte – Elaborada pela autora.

Tabela 5 – Média e mediana das medidas de acurácia das séries, simuladas de diferentes maneiras, assumindo  $B = 100$  séries *bootstrap* e diferentes tamanhos de blocos utilizados no processo de reamostragem.

Medida de acurácia	Tamanho do bloco	Simulação 1		Simulação 2		Simulação 3	
		Mediana	Média	Mediana	Média	Mediana	Média
MSE	Individual	1,36	1,42	1,69	1,79	4,15	4,97
	$l = 1$	1,38	1,47	1,67	1,77	4,07	4,77
	$l = 5$	1,41	1,47	1,65	1,77	4,05	4,86
	$l = 10$	1,41	1,47	1,66	1,77	4,13	4,87
	$l = 15$	1,43	1,48	1,66	1,78	4,13	4,89
MAAPE	Individual	78,20	78,53	78,54	79,15	79,40	82,02
	$l = 1$	78,65	79,40	78,94	78,73	79,95	81,48
	$l = 5$	78,37	79,80	78,81	78,72	80,21	81,99
	$l = 10$	78,64	79,91	78,55	78,80	80,15	81,98
	$l = 15$	78,57	80,04	78,77	79,05	79,98	82,09

Fonte – Elaborada pela autora.

MSE, o método combinado apresentou melhor resultado para a maioria das séries, enquanto que olhando para o MAAPE o método individual permaneceu com desempenho levemente superior para a maioria das séries. A média e a mediana do MSE e do MAAPE apresentaram melhores resultados para o método combinado enquanto que a mediana do MAAPE favoreceu o método individual.

Para  $B = 100$ , nota-se que, na simulação 1, o método individual permaneceu superior ao método combinado para a maioria das séries, de acordo com o MSE e com o MAAPE. A média e a mediana do MSE e do MAAPE também obtiveram melhores resultados para o modelo individual. Na simulação 2, de acordo com o MSE, o método combinado apresentou vantagem

em relação ao método individual. Ao analisar o valor do MAAPE não foi observado o mesmo resultado, e o método individual se destaca. Como observado anteriormente, a média e a mediana do MSE e a média do MAAPE são menores para o modelo combinado, somente a mediana do MAAPE que favoreceu o método individual. Na simulação 3, no geral, de acordo com ambas as medidas, MSE e MAAPE, o desempenho do modelo combinado foi superior ao desempenho do modelo individual para a maioria das séries. Como observado anteriormente, a média e a mediana do MSE e do MAAPE apresentaram melhores resultados para o método combinado enquanto que a mediana do MAAPE favoreceu o método individual.

Aparentemente, quanto maior a variância amostral da série, melhor é a performance do método combinado em relação ao método individual. O método combinado também apresentou melhor desempenho quando aumentou-se o número de amostras *bootstrap* utilizadas no processo de reamostragem da série. Não é observada nenhuma melhora ou piora do método combinado quando o tamanho do bloco escolhido no processo de reamostragem é modificado. Aparentemente, para as séries simuladas neste estudo, o tamanho do bloco  $l = 1$ , ou seja, uma reamostragem considerando os dados IID, não torna a acurácia das previsões pior. Deve-se destacar também que a estimação do modelo realizada neste estudo busca minimizar o MSE, e por essa razão o MSE é uma medida mais adequada para avaliar os resultados do que o MAAPE, que foi calculado por ser uma medida para o erro de previsão que apresenta fácil interpretação.

## 4.1 Estudo do comportamento dos parâmetros do processo de reamostragem

Com o objetivo de analisar o comportamento da acurácia das previsões combinadas conforme o número de séries *bootstrap* e o tamanho de cada bloco gerados no processo de reamostragem aumentam ou diminuem, foi calculada a acurácia de duas séries temporais, simuladas de acordo com a simulação 3 descrita anteriormente, para diferentes tamanhos de blocos  $l = (1, 2, \dots, 50)$  e diferentes números de séries reamostradas  $B = (10, 100, 1000)$ .

A primeira série temporal foi a simulada a partir da semente de número 100, que obteve um desempenho melhor na acurácia das previsões utilizando as séries combinadas. Já a segunda série temporal foi simulada a partir da semente de número 95, que obteve um desempenho melhor na acurácia das previsões utilizando a série individual. Os resultados são apresentados na [Figura 14](#) e na [Figura 15](#).

Observa-se que, quanto maior o número de amostras *bootstrap* geradas, menor é a variação da acurácia da previsão quando alterado o tamanho do bloco. Aparentemente, quando é escolhido um número de amostras *bootstrap* geradas suficientemente grande e um tamanho do bloco grande (neste estudo foi observado que  $B \geq 100$  e  $l \geq 10$  aparentam ser valores razoáveis), a acurácia das previsões parece não sofrer muita alteração, apresentando apenas uma leve tendência conforme aumenta-se o tamanho do bloco escolhido. Para tamanhos de blocos

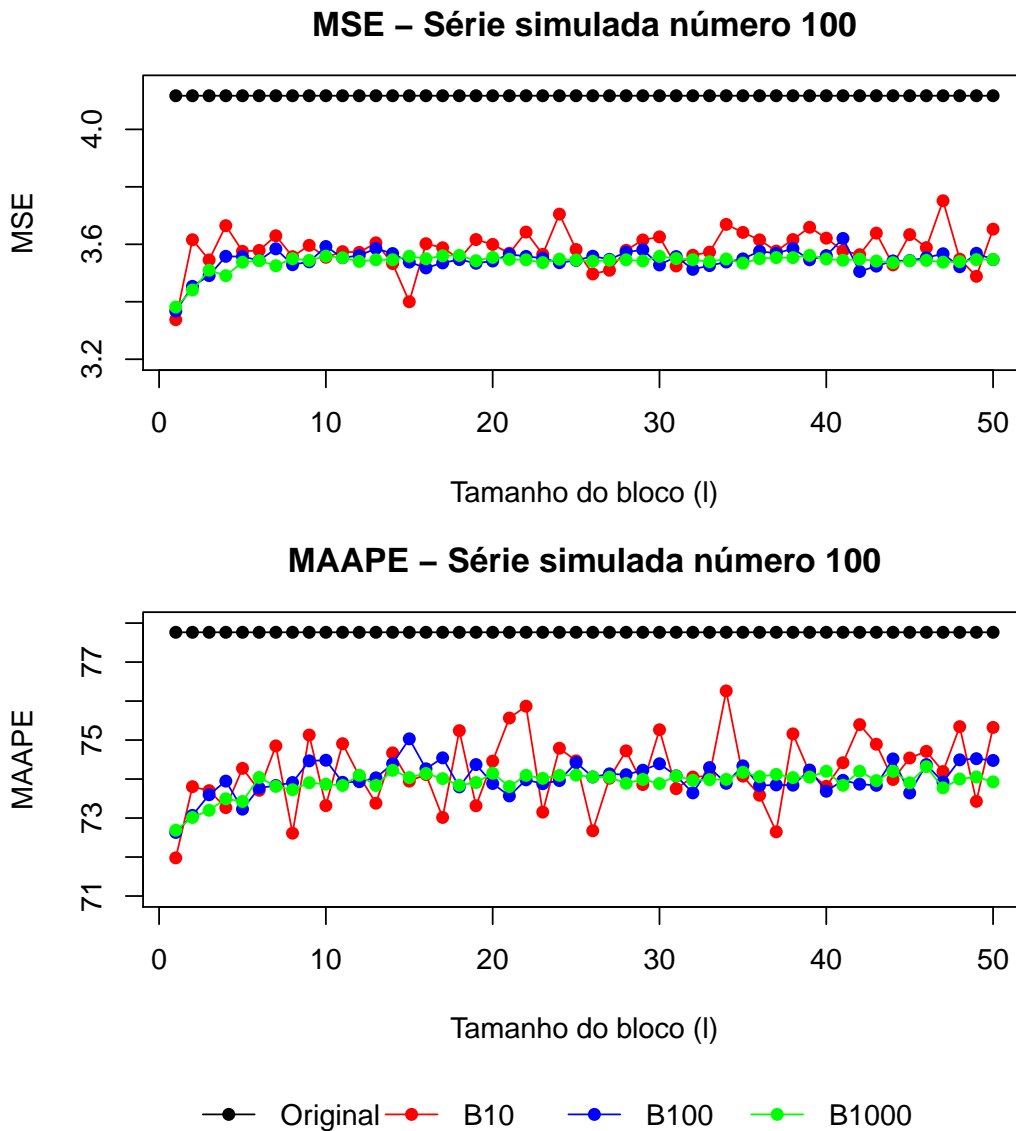


Figura 14 – Gráfico do comportamento do MSE e do MAAPE, utilizando a série gerada a partir da semente de número 100, quando o número de amostras *bootstrap* e o tamanho do bloco escolhido variam.

Fonte – Elaborada pela autora.

pequenos, a acurácia das previsões apresenta uma tendência mais evidente. Observa-se nos gráficos que a acurácia da previsão combinada na série gerada a partir da semente de número 100 é superior à previsão do modelo individual, enquanto que a acurácia da previsão combinada na série gerada a partir da semente de número 95 não é superior à previsão do modelo individual.

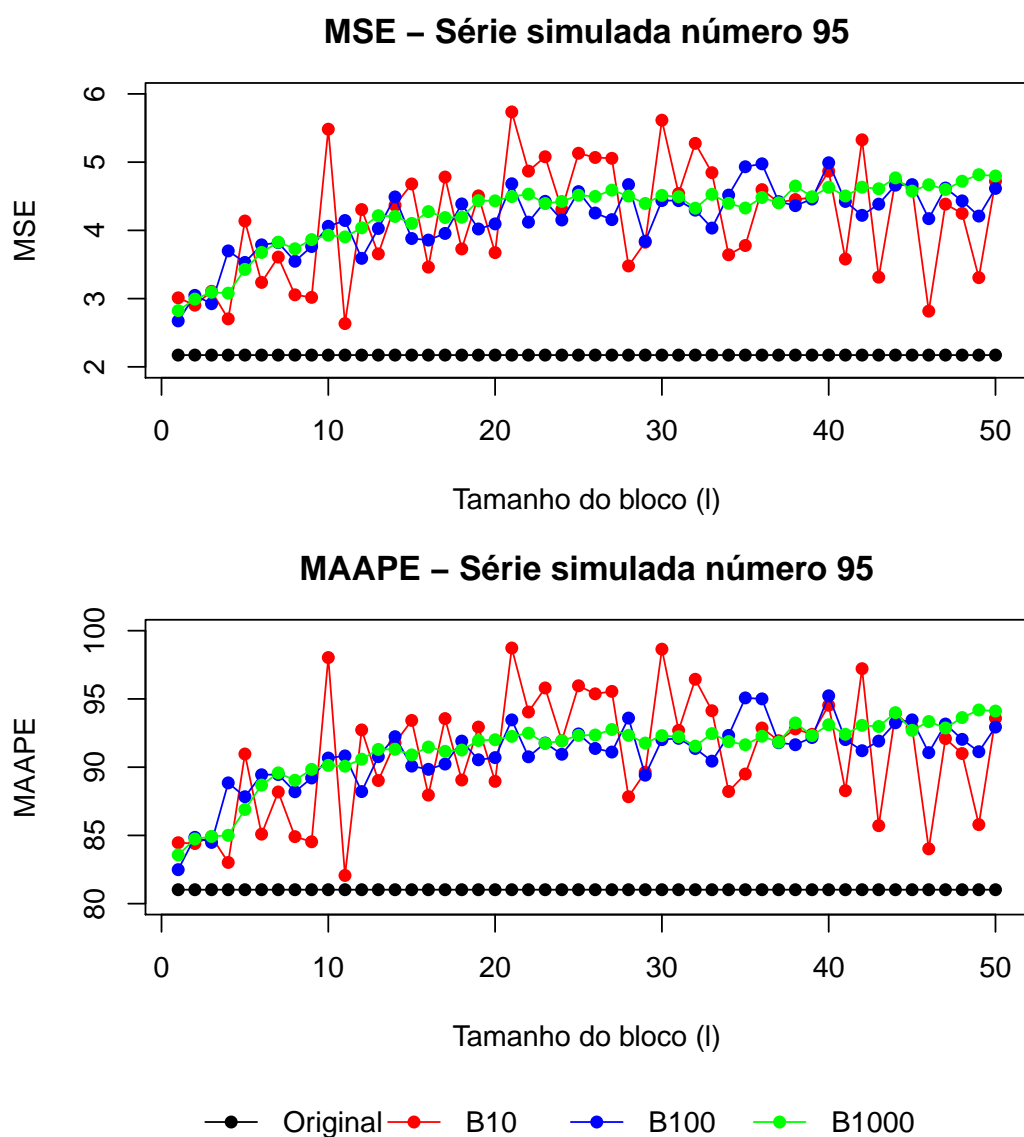


Figura 15 – Gráfico do comportamento do MSE e do MAAPE, utilizando a série gerada a partir da semente de número 95, quando o número de amostras *bootstrap* e o tamanho do bloco escolhido variam.

Fonte – Elaborada pela autora.





---

## APLICAÇÃO

---

Neste estudo foram analisadas 1.428 séries temporais reais mensais, provenientes de diferentes segmentos (séries financeiras, demográficas, industriais, etc.), disponibilizadas pela competição M3 de previsão. Assim como no estudo de simulação, tem-se por objetivo obter uma comparação entre os métodos de previsão individual e combinado, utilizando a classe de modelos ARIMA, e estudar o comportamento e o impacto dos parâmetros  $B$  (número de séries *bootstrap* geradas) e  $l$  (número de observações por bloco utilizado na reamostragem) na previsão final obtida, quando modificados.

Todas as 1.428 séries possuem seus valores estritamente positivos. O modelo utilizado na previsão do modelo individual e combinado foi, assim como nos estudos de simulação descritos anteriormente, o modelo ARIMA com o melhor ajuste fornecido pela função `auto.arima()` do *software* R, que combina testes de raiz unitária, minimização do AICc e estimação de máxima verossimilhança para obter o melhor modelo.

O tamanho de cada série de treinamento varia, porém a série de teste apresenta sempre tamanho 18, ou seja, é formada pelas últimas 18 observações da série temporal original. O número de séries *bootstrap* geradas via MBB foi  $B = 100$ , e o tamanho de bloco escolhido foi  $l = (1, 6, 12, 24)$ , ou seja: um mês, meio ano, um ano e dois anos. As medidas de acurácia utilizadas foram o RMSE, o MAAPE e o sMAPE. Sob as condições descritas, o tempo computacional aproximado para obter as previsões combinadas das 1.428 séries foi de aproximadamente sete dias.

Observa-se na [Tabela 6](#) e na [Tabela 7](#) que, independente da medida de acurácia utilizada, o desempenho do modelo combinado para gerar as previsões foi superior ao modelo individual na maioria das séries reais, em média e na mediana. Os valores da média e da mediana das medidas de acurácia adotadas neste estudo não parecem sofrer muitas alterações à medida em que variamos o valor de  $l$ . Então, considerando séries temporais reais mensais de diferentes segmentos, o método *bagging* obteve, em média e em mediana, melhores previsões do que o

Tabela 6 – Porcentagem e número de séries que obtiveram desempenho superior em cada modelo, para diferentes medidas de acurácia e diferentes tamanhos de blocos utilizados na reamostragem.

Medida de acurácia	Modelo	l = 1	l = 6	l = 12	l = 24
RMSE	Original	42% (594)	41% (588)	41% (586)	42% (596)
	Combinado	58% (834)	59% (840)	59% (842)	58% (832)
MAAPE	Original	42% (599)	41% (580)	41% (593)	42% (602)
	Combinado	58% (829)	59% (848)	59% (835)	58% (826)
sMAPE	Original	42% (602)	41% (581)	42% (594)	42% (605)
	Combinado	58% (826)	59% (847)	58% (834)	58% (823)

Fonte – Elaborada pela autora.

Tabela 7 – Média e mediana das medidas de acurácia de cada modelo, separadas por diferentes tamanhos de blocos utilizados na reamostragem.

Medida de acurácia	Medida de tendência	Individual/Combinado	l = 1	l = 6	l = 12	l = 24
RMSE	Média	762,38	734,24	733,11	734,15	736,37
	Mediana	509,77	481,93	480,61	486,39	489,05
MAAPE	Média	14,41	13,65	13,60	13,64	13,74
	Mediana	8,83	8,40	8,46	8,46	8,53
sMAPE	Média	14,71	13,83	13,77	13,82	13,93
	Mediana	8,93	8,46	8,57	8,63	8,51

Fonte – Elaborada pela autora.

método individual.

Com o objetivo de analisar os números obtidos para as medidas de acurácia do modelo original e combinado, gerou-se gráficos e intervalos de confiança para a diferença das médias. Levou-se em consideração essa diferença como sendo a média da medida de acurácia da previsão do modelo original menos a média da medida de acurácia da previsão do modelo combinado, utilizando blocos de tamanho  $l = 12$  e  $B = 100$  séries reamostradas. Os gráficos das análises descritivas são mostrados na [Figura 16](#), [Figura 17](#) e [Figura 18](#). Nota-se que a métrica RMSE não é livre de escala, por esse motivo, o interesse está apenas na comparação de valores maiores ou menores do que zero. Os intervalos de confiança calculados são observados na [Tabela 8](#).

Em todos os gráficos, percebe-se que grande parte dos valores se concentram como sendo maiores do que zero, ou seja, a média da acurácia do modelo original é maior do que a média da acurácia do modelo combinado, independentemente da medida de acurácia escolhida. De acordo com os intervalos de confiança, percebe-se que em todos eles o zero não está incluso, e seus valores são todos positivos, então conclui-se que, a média da diferença do erro do modelo original e do modelo combinado é maior do que zero, ou seja, em média a acurácia do modelo combinado apresenta melhor performance.

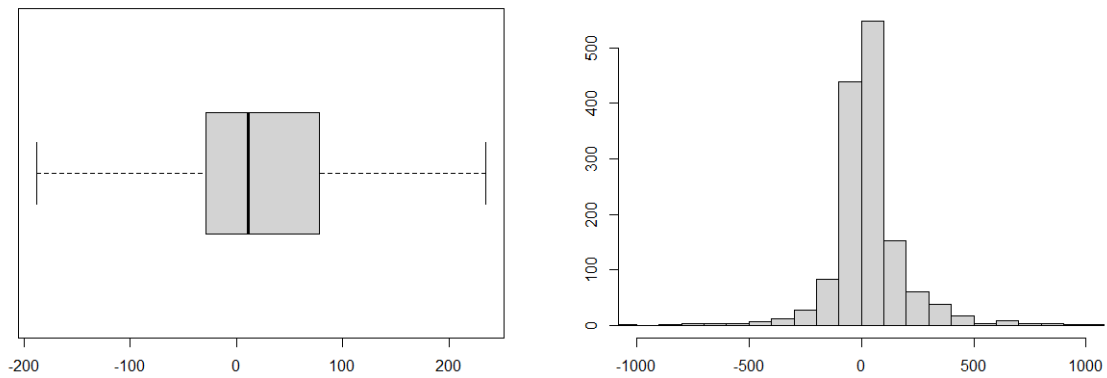


Figura 16 – Gráfico da diferença de médias do RMSE da previsão da série original e o RMSE da previsão da série combinada utilizando tamanho de bloco  $l = 12$  e amostras *bootstrap*  $B = 100$ .

Fonte – Elaborada pela autora.

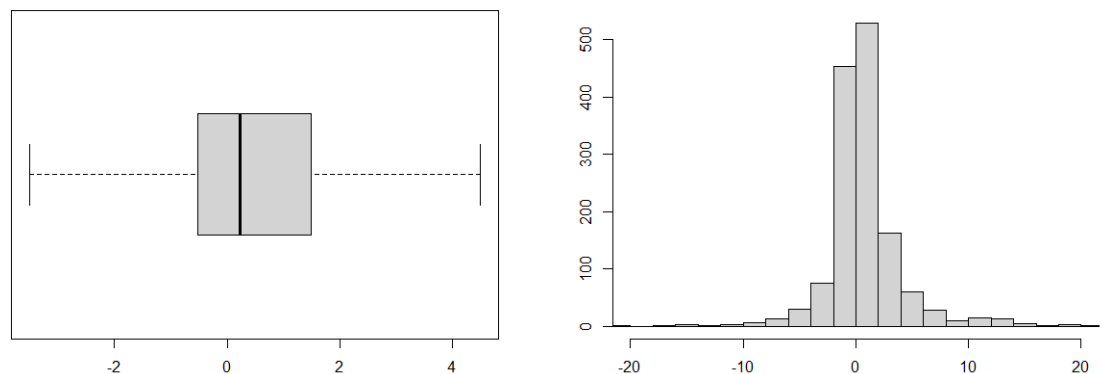


Figura 17 – Gráfico da diferença de médias do MAAPE da previsão da série original e o MAAPE da previsão da série combinada utilizando tamanho de bloco  $l = 12$  e amostras *bootstrap*  $B = 100$ .

Fonte – Elaborada pela autora.

Com o objetivo de estudar o que acontece com a acurácia das previsões combinadas em séries reais mensais quando aumentamos ou diminuímos o número de séries *bootstrap* geradas e o tamanho de bloco escolhido, foi calculada a acurácia de duas séries temporais para diferentes tamanhos de blocos  $l = (1, 2, \dots, 24)$  e diferentes números de séries reamostradas  $B = (10, 100, 1.000)$ .

A primeira série temporal foi a número 350, que obteve um desempenho superior na acurácia das previsões utilizando as séries combinadas. Já a segunda série temporal foi a número 360, que obteve um desempenho superior na acurácia das previsões utilizando o método individual. Os resultados são mostrados na [Figura 19](#) e na [Figura 20](#).

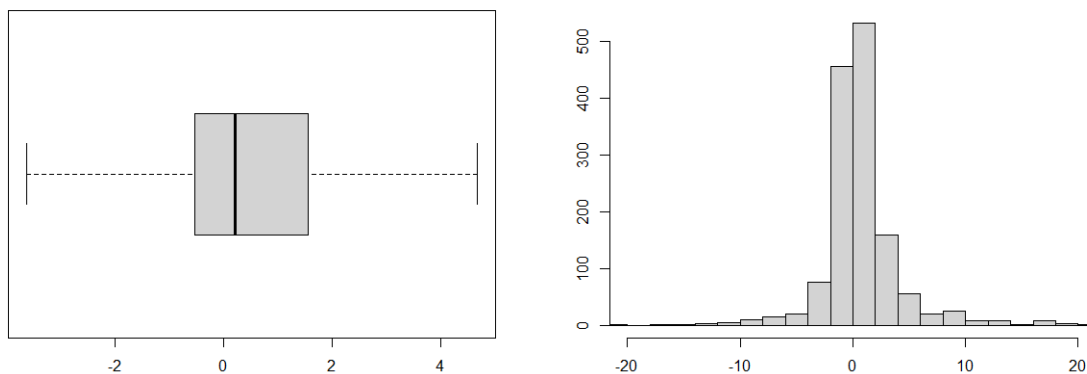


Figura 18 – Gráfico da diferença de médias do sMAPE da previsão da série original e o sMAPE da previsão da série combinada utilizando tamanho de bloco  $l = 12$  e amostras *bootstrap*  $B = 100$ .

Fonte – Elaborada pela autora.

Tabela 8 – Intervalos de confiança para a média da diferença das medidas de acurácia calculadas, fixando tamanho de bloco como  $l = 12$  e número de amostras *bootstrap*  $B = 100$  para o modelo combinado.

Medidas	Erro Padrão	Intervalo de Confiança
RMSE	6,18	[16,1153; 40,3487]
MAAPE	0,11	[0,5487; 0,9905]
sMAPE	0,14	[0,6104; 1,1653]

Fonte – Elaborada pela autora.

Como observado no estudo de simulação, percebe-se que, quanto maior o número de amostras *bootstrap* geradas menor é a variação da acurácia da previsão quando alterado o tamanho do bloco. Aparentemente, quando é escolhido um número de amostras *bootstrap* geradas grande e o tamanho do bloco grande a acurácia das previsões parece não sofrer muita alteração, apresentando apenas uma leve tendência conforme o tamanho do bloco escolhido aumenta, e para tamanhos de blocos pequenos a acurácia das previsões apresenta uma tendência mais evidente.

A acurácia da previsão combinada na série temporal mensal número 350 é superior à previsão do modelo individual, enquanto que a acurácia da previsão combinada na série temporal mensal número 360 não é superior à previsão do modelo individual. Como apresentado anteriormente, as previsões obtidas pelo modelo combinado são, em média e em mediana, superiores às obtidas no modelo individual.

A [Tabela 9](#) mostra o número de séries presentes em cada segmento do banco de dados mensais da competição M3 de previsão e a [Tabela 10](#) mostra a porcentagem em que cada modelo obteve melhor desempenho no conjunto de 1.428 séries mensais de acordo com o RMSE,

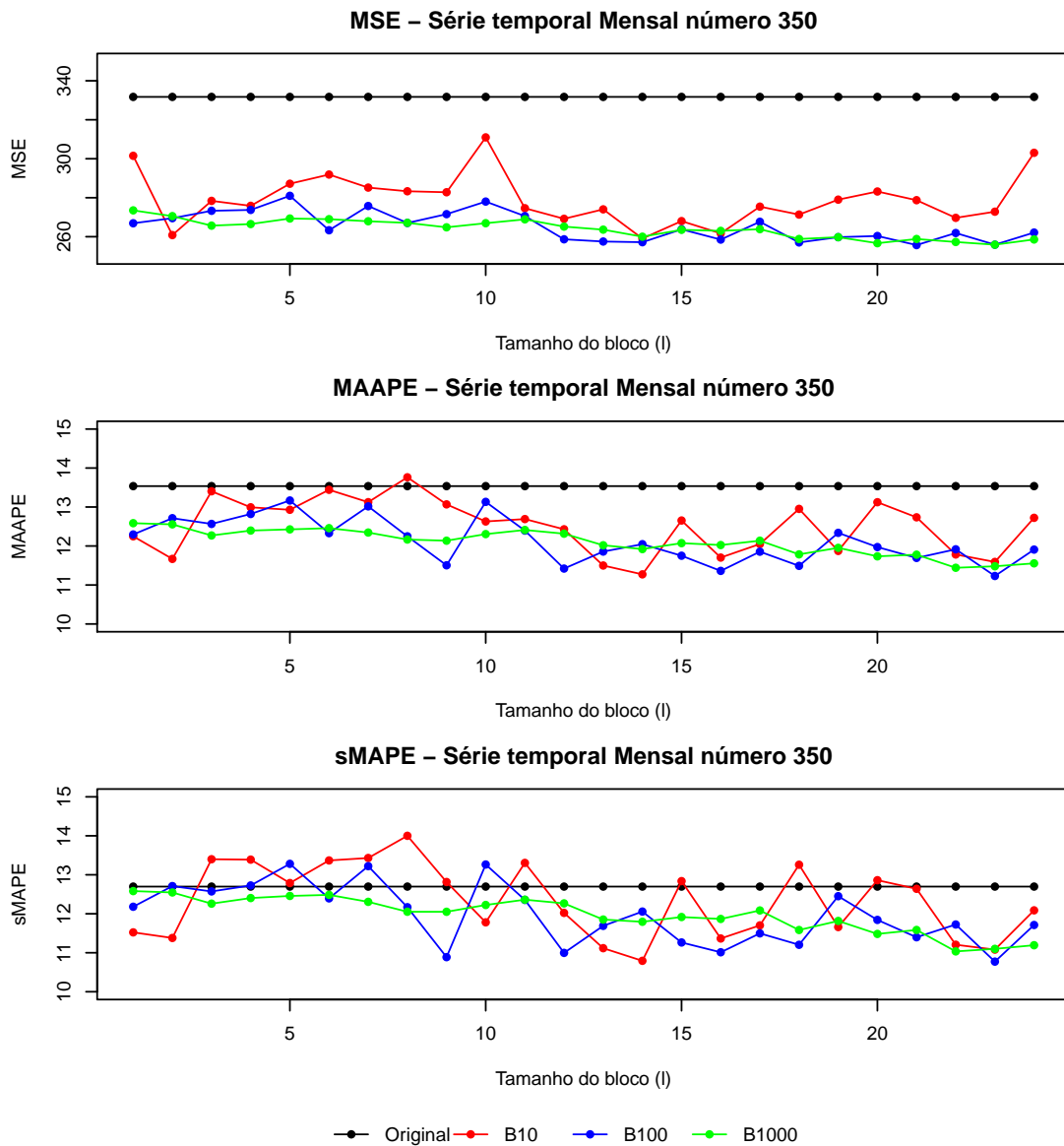


Figura 19 – Gráfico do comportamento do RMSE, do MAAPE e do sMAPE, utilizando a série temporal mensal número 350 da competição M3, quando o número de amostras *bootstrap* e o tamanho do bloco escolhido variam.

Fonte – Elaborada pela autora.

separadas pelo tamanho do bloco escolhido na reamostragem e pelos diferentes segmentos de cada série. Observa-se que em todos os segmentos, exceto para séries temporais financeiras, a previsão do modelo combinado foi, para a maioria das séries, superior à previsão do modelo individual.

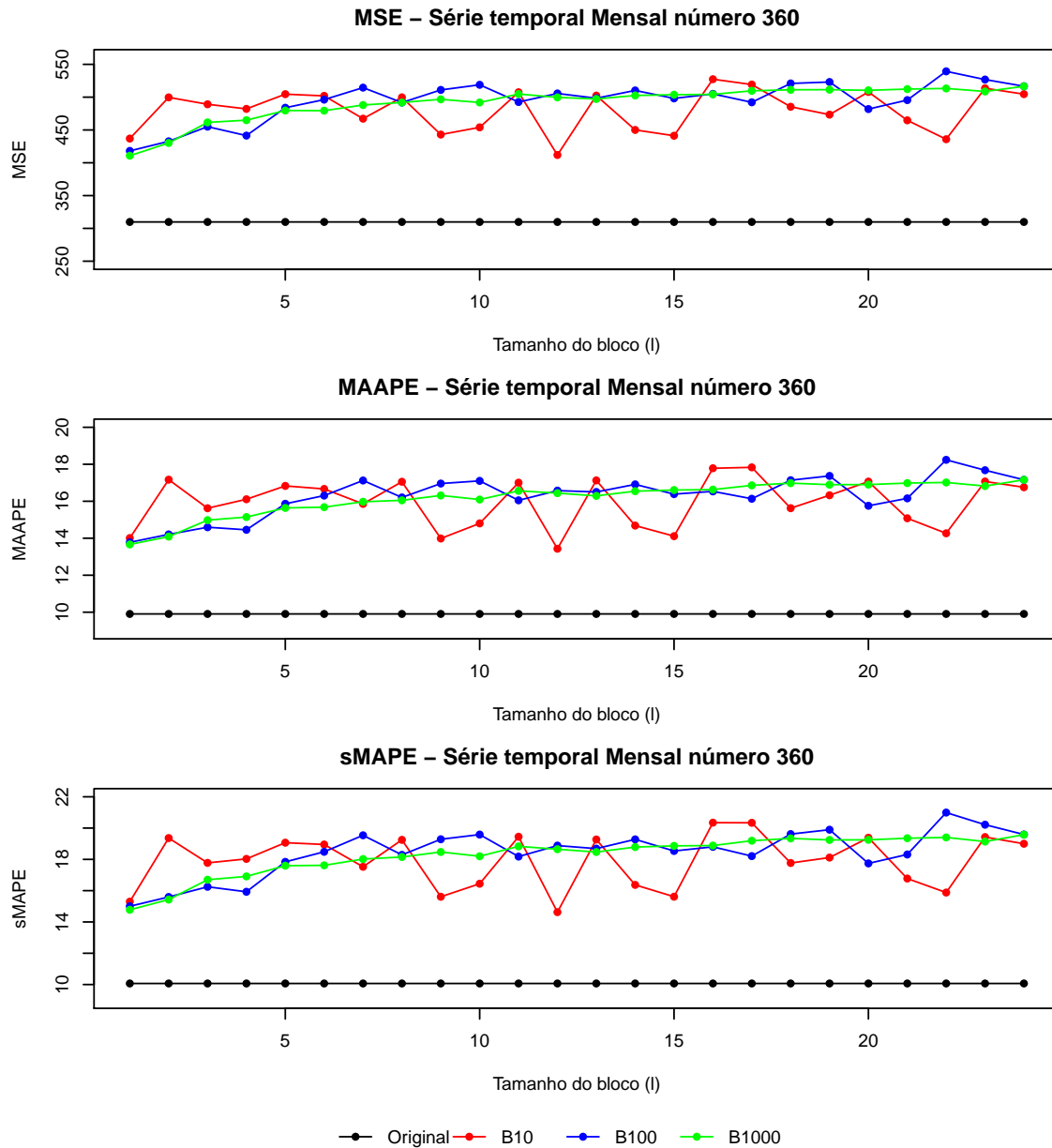


Figura 20 – Gráfico do comportamento do RMSE, do MAAPE e do sMAPE, utilizando a série temporal mensal número 360 da competição M3, quando o número de amostras *bootstrap* e o tamanho do bloco escolhido variam.

Fonte – Elaborada pela autora.

Tabela 9 – Número de séries presentes em cada segmento do banco de dados mensais da competição M3 de previsão.

Segmento	Número de séries
Demográfico	111
Financeiro	145
Industrial	334
Macro	312
Micro	474
Outros	52
Total	1.428

Fonte – Elaborada pela autora.

Tabela 10 – Porcentagem e número de séries que obtiveram desempenho superior em cada modelo de acordo com o RMSE, separadas por diferentes segmentos e tamanhos de blocos utilizados na reamostragem.

Tamanho do bloco	Modelo	Demogr.	Finan.	Industr.	Macro	Micro	Outras
$l = 1$	Individual	43% (48)	54% (78)	49% (163)	41% (127)	33% (156)	42% (22)
	Combinado	57% (63)	46% (67)	51% (171)	59% (185)	67% (318)	58% (30)
$l = 6$	Individual	46% (51)	56% (81)	49% (164)	40% (124)	31% (145)	44% (23)
	Combinado	54% (60)	44% (64)	51% (170)	60% (188)	69% (329)	56% (29)
$l = 12$	Individual	46% (51)	55% (80)	49% (163)	40% (126)	30% (144)	42% (22)
	Combinado	54% (60)	45% (65)	51% (171)	60% (186)	70% (330)	58% (30)
$l = 24$	Individual	47% (52)	57% (83)	50% (167)	41% (128)	30% (143)	44% (23)
	Combinado	53% (59)	43% (62)	50% (167)	59% (184)	70% (331)	56% (29)

Fonte – Elaborada pela autora.

## 5.1 Séries Temporais Financeiras

Tendo como motivação os resultados obtidos na seção anterior, com o objetivo de explorar e analisar o desempenho do método *bagging* em previsões de séries temporais financeiras sob diferentes perspectivas, realizou-se previsões de 145 séries temporais mensais e 58 séries temporais anuais. O modelo ajustado em cada série foi novamente o ARIMA com melhor ajuste fornecido pela função `auto.arima()` do *software* R. As séries estudadas foram as originais e as obtidas através de duas transformações, uma utilizando o log-retorno e outra utilizando o log-retorno ao quadrado das séries originais. Os resultados obtidos são mostrados na [Tabela 11](#), [Tabela 12](#), [Tabela 13](#) e [Tabela 14](#).

Tabela 11 – Porcentagem e número de séries que obtiveram desempenho superior em cada modelo sob diferentes métricas, utilizando 145 séries temporais financeiras mensais.

Séries Mensais	Medida de acurácia	Combinado ( $l = 12, B = 100$ )	Combinado ( $l = 12, B = 500$ )
Original	RMSE	45% (65)	43% (63)
	MAAPE	43% (63)	41% (59)
	sMAPE	43% (62)	42% (61)
Log Retorno	RMSE	59% (86)	59% (86)
	MAAPE	44% (64)	45% (66)
Log Retorno ao quadrado	RMSE	55% (79)	55% (79)
	MAAPE	54% (78)	55% (80)

Fonte – Elaborada pela autora.

Tabela 12 – Média e mediana das medidas de acurácia de cada modelo, sob diferentes métricas, utilizando 145 séries temporais financeiras mensais.

Séries Mensais	Medida de acurácia	Medida de tendência	Individual	Combinado ( $l = 12, B = 100$ )	Combinado ( $l = 12, B = 500$ )
Original	RMSE	Mediana	517,35	554,55	552,73
		Média	962,06	954,25	955,43
	MAAPE	Mediana	6,39	6,63	6,85
		Média	11,89	11,90	11,92
	sMAPE	Mediana	6,53	6,92	7,16
		Média	12,81	12,65	12,67
Log Retorno	RMSE	Mediana	0,0294	0,0271	0,0269
		Média	0,0971	0,0991	0,0991
	MAAPE	Mediana	76,79	76,24	76,01
		Média	73,10	73,23	73,46
Log Retorno ao quadrado	RMSE	Mediana	0,0016	0,0015	0,0014
		Média	0,1165	0,1129	0,1130
	MAAPE	Mediana	100,82	101,64	101,35
		Média	98,34	98,88	98,91

Fonte – Elaborada pela autora.



Tabela 13 – Porcentagem e número de séries que obtiveram desempenho superior em cada modelo sob diferentes métricas, utilizando 58 séries temporais financeiras anuais.

Séries Mensais	Medida de acurácia	Combinado (l=12, k=100)	Combinado (l=12, k=500)
Original	RMSE	52% (30)	55% (32)
	MAAPE	59% (34)	57% (33)
	sMAPE	57% (33)	57% (33)
Log Retorno	RMSE	62% (36)	59% (34)
	MAAPE	41% (28)	50% (29)
Log Retorno ao quadrado	RMSE	38% (22)	38% (22)
	MAAPE	43% (25)	40% (23)

Fonte – Elaborada pela autora.

Tabela 14 – Média e mediana das medidas de acurácia de cada modelo, sob diferentes métricas, utilizando 58 séries temporais financeiras anuais.

Séries Anuais	Medida de acurácia	Medida de tendência	Individual	Combinado (l=10, k=100)	Combinado (l=10, k=500)	
Original	RMSE	Mediana	1495,8	1106,9	1107,9	
		Média	2949,6	2430,4	2441,7	
	MAAPE	Mediana	20,65	22,20	21,13	
		Média	27,74	25,65	25,73	
	sMAPE	Mediana	21,12	21,90	20,98	
		Média	28,12	26,06	26,15	
Log Retorno	RMSE	Mediana	0,1386	0,1269	0,1366	
		Média	0,2475	0,2438	0,2441	
	MAAPE	Mediana	78,54	78,77	78,83	
		Média	79,90	78,99	79,25	
	Log Retorno ao quadrado	RMSE	Mediana	0,0205	0,0257	0,0256
			Média	0,2105	0,2075	0,2056
MAAPE		Mediana	99,70	101,80	102,34	
		Média	96,28	97,54	97,73	

Fonte – Elaborada pela autora.

Observa-se que, para séries temporais financeiras mensais, o modelo combinado se destaca obtendo melhor performance em séries transformadas utilizando o log-retorno e o log-retorno ao quadrado. Já para as séries originais, como mostrado anteriormente, o método individual apresentou melhores resultados. É interessante notar que mesmo o método individual obtendo melhor resultado para a maioria das séries originais, o método combinado apresentou um menor valor para a média do RMSE e do sMAPE.

As séries obtidas tomando o log-retorno apresentaram a mediana do RMSE e do MAAPE menores para o modelo combinado, mas a média de ambas as medidas foram menores para o modelo individual. Para a maioria das séries, olhando para o RMSE, o modelo combinado apresentou um melhor resultado, enquanto que olhando para o MAAPE, o modelo individual

foi o que obteve a melhor performance. Já as séries obtidas tomando o log-retorno ao quadrado apresentaram a média e a mediana do erro menor no modelo combinado utilizando como medida o RMSE. Olhando para o MAAPE, o modelo individual apresentou melhor desempenho. Para a maioria das séries, em ambas as medidas, RMSE e MAAPE, o modelo combinado apresentou um melhor resultado.

Para as séries temporais financeiras anuais, observa-se que, para as séries originais e transformadas pelo log-retorno, o modelo combinado apresentou melhor performance para a maioria das séries e também em média e mediana do RMSE e do MAAPE, exceto para o MAAPE em séries transformadas utilizando o log-retorno, em que o modelo original obteve melhor desempenho para a maioria das séries e sua mediana também foi menor. Já para as séries transformadas utilizando o log-retorno ao quadrado da série, foi observado que no geral, o método individual obteve um desempenho superior quando comparado ao método combinado. É interessante notar que mesmo o método original obtendo melhor resultado geral nas séries transformadas utilizando o log-retorno ao quadrado da série, o método combinado apresentou um erro menor para a média do RMSE.

Como discutido no estudo de simulação, deve-se destacar que a estimação do modelo realizada neste estudo busca minimizar o RMSE, e por essa razão o RMSE é uma medida mais adequada para avaliar os resultados do que as demais medidas de acurácia, que foram calculadas para fins de comparação e por serem medidas populares e de fácil interpretação.

De maneira resumida, os resultados mostraram que para séries mensais, em geral, o modelo combinado apresenta melhor performance em séries transformadas e pior performance para as séries originais, porém, mesmo nas séries originais nota-se que em média, o RMSE e o sMAPE são menores no modelo combinado. Já para séries anuais, o modelo combinado apresentou melhor performance nas séries originais e nas séries transformadas a partir do log-retorno, e apresentou pior performance nas séries transformadas a partir do log-retorno ao quadrado, porém, mesmo nas séries transformadas a partir do log-retorno ao quadrado, nota-se que em média, o RMSE é menor no modelo combinado. Não foi possível detectar um padrão ou uma explicação do comportamento do método *bagging* ao realizar previsões de séries temporais financeiras, e portanto, para trabalhos futuros, seria interessante testar novos métodos de previsão além do `auto.arima()` utilizado neste estudo.

---

## CONCLUSÃO

---

Este trabalho estudou e explorou algumas características do método *bagging* para o aprimoramento de previsões de modelos de séries temporais. O método de reamostragem utilizado no procedimento foi o MBB, escolhido dentre três métodos distintos de reamostragem em blocos por apresentar melhor desempenho. Todas as séries analisadas passaram pela transformação Box-Cox e foram submetidas a um método de decomposição (LOESS, caso a série fosse não-sazonal ou STL, caso a série fosse sazonal), pois o processo de reamostragem só pode ser aplicado em séries fracamente estacionárias, ou seja, sem tendência e sem sazonalidade. O modelo de série temporal ajustado aos dados nas simulações e nas aplicações foi o ARIMA, escolhido através da função `auto.arima()` do *software* R. As medidas de acurácia utilizadas foram o MSE, o RMSE, o MAAPE e o sMAPE.

O estudo de simulação foi dividido em três etapas. Na primeira, as séries foram simuladas de um modelo autorregressivo de ordem dois, a média das variâncias amostrais das séries foi igual a 1,39 e os parâmetros foram fixados. Na segunda, as séries foram simuladas de um modelo de médias móveis de ordem três, a média das variâncias amostrais das séries foi igual a 1,75 e os parâmetros foram fixados. Na terceira, as séries foram simuladas da soma de uma série simulada a partir de um modelo autorregressivo de ordem dois com uma série simulada a partir de um modelo de médias móveis de ordem três, a média das variâncias amostrais das séries foi igual a 3,96 e todos os parâmetros foram fixados. Em cada uma das três simulações o método foi aplicado em 100 séries de tamanho 100 cada e a previsão foi realizada para 20 passos à frente.

Os resultados do estudo de simulação indicaram que, em média, quanto maior a variância amostral da série, melhor o desempenho do modelo combinado em relação ao modelo individual para realizar as previsões. O estudo de simulação 1, que em média as séries apresentaram menores variâncias amostrais, obteve melhor desempenho no modelo individual em relação ao combinado para a maioria das séries e em média e mediana do MSE e do MAAPE, já o estudo de simulação 3, que em média as séries apresentaram maiores variâncias amostrais, obteve melhor

desempenho no modelo combinado em relação ao modelo individual para a maioria das séries e em média e mediana do MSE e em média do MAAPE. Já o estudo de simulação 2, que em média as séries apresentaram variâncias amostrais maiores do que a média das variâncias amostrais das séries geradas no estudo de simulação 1 e menores do que a média das variâncias amostrais das séries geradas no estudo de simulação 3, de acordo com o MSE, obteve desempenho superior no método combinado em relação ao método individual para a maioria das séries e em média e mediana, porém, em relação ao MAAPE, o desempenho do método individual aparenta ser melhor para maioria das séries e em sua mediana, porém, em média, o modelo combinado se destaca. Deve-se considerar que a estimação do modelo realizada neste estudo busca minimizar o MSE, e, portanto, o MSE é uma medida mais adequada para avaliar os resultados do que o MAAPE.

O tamanho do bloco escolhido no processo de reamostragem não apresentou ser relevante na acurácia das previsões. Mostrou-se que, quanto maior o número de amostras *bootstrap*  $B$  geradas, menor é a variação da acurácia da previsão quando alterado o tamanho do bloco. Aparentemente, quando escolhemos um valor de  $B$  grande e o tamanho do bloco grande (o estudo mostrou que  $B \geq 100$  e  $l \geq 10$  aparentam ser valores razoáveis), a acurácia das previsões parece não sofrer muita alteração, apresentando apenas uma leve tendência conforme aumentamos o tamanho do bloco escolhido, e para tamanhos de blocos pequenos a acurácia das previsões apresenta uma tendência mais evidente.

Posteriormente, o método foi aplicado em 1.428 séries temporais reais mensais, de diferentes segmentos, provenientes do banco de dados da competição M3 de previsão de séries temporais. Neste segundo momento, a previsão foi calculada, para todas as séries, 18 passos à frente. Os resultados mostraram que, em média e mediana, e para a maioria das séries, o modelo combinado obteve um desempenho melhor do que o modelo individual na acurácia das previsões. Em relação ao número de amostras *bootstrap*  $B$  e tamanho de bloco  $l$  definidos no processo de reamostragem, o que foi observado comprova as conclusões obtidas no estudo de séries simuladas. Foi observado que, em média, a previsão de séries temporais utilizando o método *bagging* apresentou melhor acurácia em relação à previsão utilizando apenas o modelo individual em séries reais mensais, exceto para séries temporais financeiras.

Realizou-se então, um pequeno estudo a fim de avaliar o comportamento da previsão combinada em relação à previsão individual utilizando séries temporais reais mensais e anuais, originais e transformadas (utilizando o log-retorno e o log-retorno ao quadrado das séries). De maneira resumida, os resultados mostraram que para séries mensais, em geral, o modelo combinado apresentou melhor performance nas séries transformadas e pior performance nas séries originais, porém, mesmo nas séries originais nota-se que em média, o RMSE e o sMAPE são menores no modelo combinado. Já para séries anuais, o modelo combinado apresentou melhor performance nas séries originais e nas séries transformadas a partir do log-retorno, e apresentou pior performance nas séries transformadas a partir do log-retorno ao quadrado, porém,

mesmo nas séries transformadas a partir do log-retorno ao quadrado, nota-se que em média, o RMSE é menor no modelo combinado. Não foi possível detectar um padrão ou uma explicação do comportamento do método *bagging* ao realizar previsões de séries temporais financeiras, e portanto, como sugestão de trabalhos futuros, seria interessante testar novos métodos de previsão além do `auto.arima()` utilizado neste estudo, além de explorar o método combinado em outros diferentes aspectos e abordagens com o objetivo de verificar seu desempenho na obtenção de melhores previsões para séries temporais financeiras.



## REFERÊNCIAS

---

---

- ARMSTRONG, J. **Long-range forecasting : from crystal ball to computer**. New York: Wiley, 1985. ISBN 0471823600. Citado na página 54.
- BATES, J. M.; GRANGER, C. W. J. The combination of forecasts. **Journal of the Operational Research Society**, Taylor & Francis, v. 20, n. 4, p. 451–468, 1969. Disponível em: <<https://doi.org/10.1057/jors.1969.103>>. Citado na página 23.
- BERGMEIR, C.; HYNDMAN, R. J.; BENÍTEZ, J. M. Bagging exponential smoothing methods using STL decomposition and box–cox transformation. **International Journal of Forecasting**, Elsevier BV, v. 32, n. 2, p. 303–312, apr 2016. Citado nas páginas 24 e 39.
- BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. Holden-Day, 1970. (Holden-Day series in time series analysis and digital processing). ISBN 9780816210947. Disponível em: <<https://books.google.com.br/books?id=5BVfnXaq03oC>>. Citado na página 49.
- BOX, G. E.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 26, n. 2, p. 211–243, 1964. Citado na página 40.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 4th. ed. Wiley, 2008. ISBN 9780470272848, 0470272848. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=9fa5414b790ea57bebc0125d13ba3c86>>. Citado na página 51.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015. Citado nas páginas 33 e 48.
- BREIMAN, L. Bagging predictors. **Machine Learning**, Springer Science and Business Media LLC, v. 24, n. 2, p. 123–140, aug 1996. Citado na página 51.
- BROCKWELL, P. **Time Series: Theory and Methods**. New York, NY: Springer New York, 1991. ISBN 9781441903204. Citado na página 50.
- BUNN, D. W. A bayesian approach to the linear combination of forecasts. **Operational Research Quarterly (1970-1977)**, JSTOR, v. 26, n. 2, p. 325, jun 1975. Citado na página 23.
- CARLSTEIN, E. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 14, n. 3, sep 1986. Citado nas páginas 24 e 30.
- CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. **International journal of forecasting**, Elsevier, v. 5, n. 4, p. 559–583, 1989. Citado na página 23.
- CLEVELAND, R. B.; CLEVELAND, W. S.; MCRAE, J. E.; TERPENNING, I. Stl: A seasonal-trend decomposition. **Journal of official statistics**, v. 6, n. 1, p. 3–73, 1990. Citado nas páginas 45 e 46.

- CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots. **Journal of the American Statistical Association**, Informa UK Limited, v. 74, n. 368, p. 829–836, dec 1979. Citado na página 41.
- CLEVELAND, W. S.; DEVLIN, S. J. Locally weighted regression: an approach to regression analysis by local fitting. **Journal of the American statistical association**, Taylor & Francis, v. 83, n. 403, p. 596–610, 1988. Citado na página 41.
- CLEVELAND, W. S.; LOADER, C. Smoothing by local regression: Principles and methods. In: **Statistical theory and computational aspects of smoothing**. [S.l.]: Springer, 1996. p. 10–49. Citado na página 43.
- CORDEIRO, C.; NEVES, M. Forecasting time series with boot. expos procedure. **REVSTAT-Statistical Journal**, v. 7, n. 2, p. 135–149, 2009. Citado na página 24.
- DANTAS, T. M.; OLIVEIRA, F. L. C.; REPOLHO, H. M. V. Air transportation demand forecast through bagging holt winters methods. **Journal of Air Transport Management**, Elsevier BV, v. 59, p. 116–123, mar 2017. Citado na página 24.
- DICKINSON, J. P. Some statistical results in the combination of forecasts. **Operational Research Quarterly (1970-1977)**, JSTOR, v. 24, n. 2, p. 253, jun 1973. Citado na página 23.
- DUDANI, S. The distance-weighted k-nearest-neighbor rule. **IEEE Transactions on Systems, Man, and Cybernetics**, n. 4, p. 325–327, 1976. Citado na página 42.
- EFRON, B. Bootstrap methods: Another look at the jackknife. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 7, n. 1, p. 1–26, jan. 1979. Disponível em: <<https://doi.org/10.1214%2Faos%2F1176344552>>. Citado nas páginas 24 e 27.
- GUERRERO, V. M. Time-series analysis supported by power transformations. **Journal of Forecasting**, Wiley, v. 12, n. 1, p. 37–48, jan 1993. Citado na página 40.
- GUTHRIE, W. F. **NIST/SEMATECH e-Handbook of Statistical Methods - Retrieved in November 17, 2020**. 2012. Disponível em: <<http://www.itl.nist.gov/div898/handbook/>>. Citado na página 44.
- HYNDMAN, R.; ATHANASOPOULOS, G.; BERGMEIR, C.; CACERES, G.; CHHAY, L.; O'HARA-WILD, M.; PETROPOULOS, F.; RAZBASH, S.; WANG, E.; YASMEEN, F. **forecast: Forecasting functions for time series and linear models**. [S.l.], 2020. R package version 8.13. Disponível em: <<https://pkg.robjhyndman.com/forecast/>>. Citado na página 43.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v. 26, n. 3, p. 1–22, 2008. Disponível em: <<https://www.jstatsoft.org/article/view/v027i03>>. Citado nas páginas 40, 43 e 53.
- IRIZARRY, R. A.; BRAVO, H. C. **Lecture 7: Smoothing**. 2010. Disponível em: <<http://users.umiacs.umd.edu/~hcorrada/PracticalML/pdf/lectures/smoothing.pdf>>. Citado na página 42.
- IZBICKI, R.; SANTOS, T. M. dos. **Aprendizado de máquina: uma abordagem estatística**. [S.l.: s.n.], 2020. ISBN 978-65-00-02410-4. Citado na página 43.
- KUNSCH, H. R. The jackknife and the bootstrap for general stationary observations. **The annals of Statistics**, JSTOR, p. 1217–1241, 1989. Citado nas páginas 24 e 29.



- LIU, R. Y.; SINGH, K. Moving blocks jackknife and bootstrap capture weak dependence. **Exploring the limits of bootstrap**, v. 225, p. 248, 1992. Citado na página 29.
- MAKRIDAKIS, S. Accuracy measures: theoretical and practical concerns. **International Journal of Forecasting**, Elsevier BV, v. 9, n. 4, p. 527–529, dec 1993. Citado na página 54.
- MANCUSO, A. C. B.; WERNER, L. Review of combining forecasts approaches. **Independent journal of management & production**, Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), v. 4, n. 1, p. 248–277, 2013. Citado na página 23.
- MOORE, D.; MCCABE, G.; DUCKWORTH, W.; ALWAN, L. The practice of business statistics. Wh Freeman, 2008. Citado na página 24.
- MORETTIN, P. A.; TOLOI, C. M. d. C. Análise de séries temporais. 2004. Citado na página 29.
- OLIVEIRA, E. M. de; OLIVEIRA, F. L. C. Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods. **Energy**, Elsevier, v. 144, p. 776–788, 2018. Citado na página 24.
- PETROPOULOS, F.; HYNDMAN, R. J.; BERGMEIR, C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? **European Journal of Operational Research**, Elsevier, v. 268, n. 2, p. 545–554, 2018. Citado na página 24.
- POLITIS, D. N.; ROMANO, J. P. A circular block-resampling procedure for stationary data. **Exploring the limits of bootstrap**, Wiley, New York, v. 2635270, 1992. Citado nas páginas 24, 31 e 32.
- PROENÇA, I. M. D. Estimativas jackknife e bootstrap para o enviesamento e desvio-padrão do índice de gini. **Doc.Trabalho Nº 67**, Lisboa, 1988. Citado na página 28.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 45 e 46.
- ROBERT, C. **Monte Carlo Statistical Methods**. New York, NY: Springer New York, 2004. ISBN 9781475741452. Citado na página 28.
- SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. Akaike information criterion statistics. **Dordrecht, The Netherlands: D. Reidel**, Taylor & Francis, v. 81, 1986. Citado na página 51.
- SUGIURA, N. Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 7, n. 1, p. 13–26, 1978. Citado na página 51.
- WHITTLE, P. **Hypothesis testing in time series analysis**. [S.l.]: Almqvist & Wiksells boktr., 1951. v. 4. Citado na página 49.

