



**Universidade de Brasília  
Departamento de Estatística**

**Um estudo sobre a performance de modelos preditivos**

**Thiago Patricio Soares de Oliveira**

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2022**

**Thiago Patricio Soares de Oliveira**

**Um estudo sobre a performance de modelos preditivos**

Orientador: Prof. Dr. José Augusto Fiorucci

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2022**

---

# Resumo

A previsão é um dos maiores objetivos quando buscamos modelos para séries temporais. Para tanto, existem famílias de modelos que são usadas na maioria dos estudos: ARIMA e ETS. Este estudo tem como objetivo apresentar uma alternativa bayesiana, conhecida como BSTS, para previsão de séries temporais, comparar a sua performance com os métodos clássicos citados anteriormente no banco de dados M3 e conferir a sua viabilidade. Ademais, será utilizada uma técnica nos modelos clássicos onde as previsões são calculadas a partir de variações da série original conhecida como Bagging. O estudo começa com uma série do banco de dados como exemplo para detalhar o modelo bayesiano. Nesse exemplo, percebe-se que o modelo com a melhor performance é o ETS utilizando a técnica de Bagging nas previsões e o bayesiano aparece logo em seguida com o segundo melhor desempenho. Ao expandir o estudo para o banco de dados completo, foi avaliado que os modelos ETS com Bagging, em geral, obtiveram os melhores resultados nas previsões. Essa alta performance é obtida em troca de um aumento considerável no tempo de computação. Os BSTS tiveram, em geral, um desempenho levemente superior aos ARIMA, que se deu basicamente pela superioridade nas previsões de séries mensais as quais são maioria no banco de dados. Sendo assim, tais modelos se mostraram capazes de ser uma alternativa viável na previsão das séries do banco de dados em questão.

Palavras-chaves: Modelos ARIMA, Modelos de Alisamento Exponencial, Bagging, Modelos BSTS, M3.

## **Lista de Tabelas**

1	Decomposição do banco de dados M3 . . . . .	19
2	Horizontes de previsão do banco de dados M3 . . . . .	19
3	MAPE calculado em cada modelo . . . . .	28
4	Resultados empíricos usando sMAPE em cada modelo (%) . . . . .	28

## Lista de Figuras

1	Decomposição da série de id 2107 do banco de dados M3 . . . . .	15
2	Exemplo de variações da série de id 2107 . . . . .	16
3	Exemplo de previsões pontuais das variações da série de id 2107 . . . . .	16
4	Intervalo de 95% de confiança para as previsões da série de id 2107 usando bagging . . . . .	17
5	Decomposição MSTL da série de id 780 . . . . .	22
6	Previsão da série usando o modelo SARIMA, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões.	23
7	Previsão da série usando o modelo ETS, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões. . . .	24
8	Previsão da série usando o modelo ETS com bagging, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões. . . . .	25
9	Previsão da série usando o modelo SARIMA com bagging, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões. . . . .	25
10	Gráfico Q-Q para normalidade e gráfico da função de autocorrelação (ACF) dos resíduos do modelo de tendência local com sazonalidade bayesiano . . .	26
11	Ajuste do modelo BSTS de tendência local com sazonalidade à série analisada	27
12	Previsão da série usando o modelo BSTS de tendência local com sazonalidade, num horizonte de 8 trimestres. A linha azul é a mediana das previsões, a parte sombreada é a densidade da distribuição à posteriori e o intervalo de 95% de confiança é representado pelas linhas verdes. . . . .	27
13	Intervalos MCB para todos os modelos em estudo. Considerando um nível de significância de 5%. . . . .	29

14	Intervalos MCB para os modelos selecionados. Considerando um nível de significância de 5%. . . . .	30
----	--	----

## **Lista de siglas**

AIC - Critério de informação de Akaike

AICc - Critério de informação de Akaike corrigido

AR - Autoregressivo

ARMA - *Autoregressive moving average*

ARIMA - *Autoregressive integrated moving average*

BSTS - *Bayesian Structured Time Series*

ETS - Modelos de alisamento exponencial

HWT - Holt-Winters

MA - Médias Móveis

MAPE - Erro percentual médio absoluto

MCB - *Multiple Comparisons with the best*

MCMC - *Markov Chain Monte Carlo*

MSTL - *Multiple Seasonal-Trend decomposition using LOESS*

SARIMA - *Seasonal Autoregressive integrated moving average*

SEH - Suavização exponencial de Holt

SES - Suavização exponencial simples

# Sumário

<b>1 Introdução</b> . . . . .	8
<b>2 Referencial Teórico</b> . . . . .	10
2.1 Modelo ARIMA . . . . .	10
2.1.1 Modelo SARIMA . . . . .	10
2.2 Alisamento Exponencial . . . . .	11
2.2.1 Suavização exponencial simples . . . . .	11
2.2.2 Suavização exponencial de Holt . . . . .	11
2.2.3 Suavização exponencial sazonal de Holt-Winters . . . . .	12
2.3 Modelos Estruturais . . . . .	13
2.3.1 Modelo de nível local . . . . .	13
2.3.2 Modelo de tendência local . . . . .	13
2.3.3 Modelo com tendência local e sazonalidade . . . . .	14
2.4 Bagging . . . . .	14
2.5 <i>Multiple comparisons with the best</i> . . . . .	17
<b>3 Metodologia</b> . . . . .	19
3.1 Conjunto de dados . . . . .	19
3.2 Critério de informação de Akaike . . . . .	19
3.3 Modelos estruturais Bayesianos . . . . .	20
3.4 Markov Chain Monte Carlo . . . . .	20
3.5 Performance. . . . .	21
3.5.1 MAPE . . . . .	21
3.5.2 sMAPE . . . . .	21
<b>4 Resultados</b> . . . . .	22
4.1 Ilustração dos modelos para uma série temporal . . . . .	22
4.2 Avaliação dos modelos . . . . .	28
<b>5 Conclusão</b> . . . . .	31



**Referências** . . . . . 33

# 1 Introdução

Um dos principais objetivos da análise de séries temporais é prever como uma série temporal evoluirá observando o passado, ou seja, realizar previsões para a série em questão, sejam elas de curto ou longo prazo (MORETTIN; TOLOI, 2006). Para tanto, devemos selecionar um modelo que represente o comportamento da série de maneira mais condizente com a realidade e, então, as previsões são calculadas para um período especificado no futuro. O alisamento exponencial (ETS) e os modelos autoregressivos integrados de médias móveis (ARIMA) são as abordagens mais utilizadas para análises de séries temporais, consideradas abordagens clássicas pela literatura. O primeiro, garante um maior peso para observações mais recentes e se baseia na descrição da tendência e sazonalidade, enquanto o segundo descreve a autocorrelação dos dados (HYNDMAN; ATHANASOPOULOS, 2021).

Contudo, esses modelos clássicos possuem limitações. As previsões dos modelos ARIMA, por exemplo, dependem de padrões prévios da série e dos erros das previsões anteriores (FEROZE, 2020). Somado a isso, esses modelos sofrem com sobreajuste, principalmente na presença de covariáveis (BROCKWELL; DAVIS, 2001), e com problemas na previsão quando temos poucas observações. Já os modelos ETS, não lidam bem com valores faltantes, nos restando algumas opções como, por exemplo, substituir o valor faltante por uma estimativa (interpolação) ou usar os dados seguintes, desde que faça sentido (HYNDMAN; ATHANASOPOULOS, 2021).

Segundo Hyndman e Athanasopoulos (2021), é possível usar uma técnica de agregação de bootstrap, conhecida como *bootstrap aggregation* (bagging), para melhorar a precisão das previsões ao reduzir a variância dos resultados sem aumentar o viés. Contudo, essa técnica se baseia em criar uma quantidade arbitrária de variações de uma série, encontrar um modelo e realizar previsões para cada uma das variações da série original, ou seja, é uma técnica computacionalmente intensiva. Em geral, o bagging resulta em previsões melhores do que aplicar o modelo diretamente (BERGMEIR et al., 2016).

Neste trabalho, vamos analisar uma abordagem alternativa para a modelagem e previsão de séries temporais. Para contornar as limitações dos modelos clássicos e encontrar uma opção diferente do bagging, analisaremos o problema de um ponto de vista bayesiano, através dos modelos estruturais bayesianos (BSTS), propostos por Scott e Varian (2013). Estes modelos possuem algumas vantagens: os parâmetros são atualizados ao passar do tempo, revelam o comportamento estocástico da série (HARVEY, 1990) e realizam previsões razoáveis com poucas observações (MCQUIRE et al., 2019). A própria

abordagem bayesiana traz benefícios em algumas situações, como aproveitar informações prévias das variáveis, quando estas estão presentes na série temporal de interesse (FERROZE, 2020). Porém, a distribuição a posteriori não é tão simples de ser calculada analiticamente. Então, uma aproximação numérica razoável é obtida através de um algoritmo *Markov Chain Monte Carlo* (MCMC).

Este trabalho tem como objetivo apresentar e comparar o desempenho dos modelos BSTS com os modelos clássicos existentes na literatura. Para tanto, utilizaremos métricas livres de escala para comparar o desempenho desses diferentes modelos na previsão das séries temporais contidas no banco de dados da terceira competição de Makridakis (MAKRIDAKIS; HIBON, 2000). Além disso, serão realizados testes de hipóteses para comparar estatisticamente tais modelos, e assim verificar a viabilidade do uso desses modelos bayesianos para a previsão dessas séries.

O restante deste relatório está organizado no seguinte formato: na Seção 2 são apresentadas as bases teóricas utilizadas na elaboração do trabalho; a Seção 3 apresenta o banco de dados e os métodos utilizados para a seleção e comparação dos modelos em análise; a Seção 4 mostra um exemplo da metodologia do estudo e uma discussão dos resultados obtidos; e por fim, a conclusão da comparação de desempenho e suas implicações são apresentadas na Seção 5.

## 2 Referencial Teórico

### 2.1 Modelo ARIMA

Os modelos ARMA, que unem modelos autoregressivos (AR) e de médias móveis (MA), só podem ser aplicados em séries estacionárias. No entanto, grande parte das séries são não estacionárias. Para contornar este problema, o modelo ARMA( $p, q$ ) é aplicado na  $d$ -ésima diferença da série, que resulta em um modelo que pode ser escrito como

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (2.1.1)$$

em que  $Y_t$  é a série diferenciada (HYNDMAN; ATHANASOPOULOS, 2021). Tal abordagem é chamada de modelo ARIMA( $p, d, q$ ), em que  $p$  é a ordem da parte autoregressiva,  $d$  o grau da diferenciação e  $q$  a ordem da parte de médias móveis. Esses modelos podem ser generalizados incluindo um operador sazonal (MORETTIN; TOLOI, 2006).

#### 2.1.1 Modelo SARIMA

Modelos ARIMA são usados na previsão para séries temporais univariadas. Tais modelos dão suporte tanto para elementos de modelos autoregressivos quanto médias móveis. A diferenciação permite que modelos ARIMA possam ser utilizados na análise de séries com tendência. Contudo, esses modelos não consideram sazonalidade nas séries. Para solucionar este problema, temos a generalização do ARIMA em ARIMA Sazonal ou, apenas, SARIMA.

Segundo Brockwell e Davis (2001), diferenciar a série  $X_t$  no lag  $m$  é uma maneira conveniente de eliminar a raiz sazonal do período  $m$ . Sendo assim, se  $d$  e  $D$  são inteiros não-negativos, então  $X_t$  é SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_m$  com período  $m$  se a série diferenciada  $Y_t = (1 - B)^d (1 - B^m)^D X_t$  é um processo ARMA causal

$$\phi(B)\Phi(B^m)Y_t = \theta(B)\Theta(B^m)\varepsilon_t \quad (2.1.2)$$

em que,  $\phi(B)$  é a parcela AR,  $\Phi(B)$  a AR sazonal,  $\theta(B)$  a parte de MA,  $\Theta(B)$  a MA sazonal e

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \Phi(B) &= 1 - \Phi_1 B^m - \dots - \Phi_P B^{mP}, \end{aligned}$$

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q,$$

$$\Theta(B) = 1 + \Theta_1 B^m + \dots + \Theta_q B^{mQ}.$$

## 2.2 Alisamento Exponencial

De acordo com Morettin e Toloí (2006), grande parte dos métodos de previsão presentes na literatura são baseados no conceito de que observações passadas possuem informações do comportamento da série. As técnicas de suavização exponencial assumem que valores extremos representam a aleatoriedade, e usando a suavização desses valores podemos extrair o padrão de comportamento da série analisada.

### 2.2.1 Suavização exponencial simples

Considere a seguinte série temporal  $Y_1, \dots, Y_n$ :

$$Y_t = \mu_t + \epsilon_t, \quad t = 1, \dots, N. \quad (2.2.1)$$

Em que  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \sigma_\epsilon^2$  e  $\mu_t$  um parâmetro desconhecido, ou seja, uma série localmente composta por seu nível e um ruído aleatório.

A suavização exponencial simples (SES) é descrita por

$$\bar{Y}_t = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k Y_{t-k} + (1 - \alpha)^t \bar{Y}_0, \quad t = 1, \dots, N, \quad (2.2.2)$$

em que  $\bar{Y}_t$  é o valor suavizado e  $\alpha$  a constante de suavização, entre 0 e 1. Pela expressão, nota-se que SES é uma média ponderada com pesos maiores em observações mais recentes.

### 2.2.2 Suavização exponencial de Holt

Caso seja necessário implementar o método SES em uma série com tendência linear, as previsões subestimarão ou superestimarão constantemente os valores reais (MORETTIN; TOLOI, 2006). Para contornar este problema, usa-se o método de suavização exponencial de Holt (SEH). O método é semelhante aos SES, porém, além de suavizar o nível, também há uma constante para suavizar a tendência. E assim, resulta nas seguintes expressões:

$$\bar{Y}_t = AY_t + (1 - A)(\bar{Y}_{t-1} + \hat{T}_{t-1}) \quad t = 2, \dots, N, \quad (2.2.3)$$

$$\hat{T}_t = C(\bar{Y}_t - \bar{Y}_{t-1}) + (1 - C)\hat{T}_{t-1} \quad t = 2, \dots, N, \quad (2.2.4)$$

$A$  e  $C$  são as constantes de suavização,  $0 < A, C < 1$ ,  $\bar{Y}_t$  e  $\hat{T}_t$  são os estimadores para o nível e tendência da série, respectivamente.

### 2.2.3 Suavização exponencial sazonal de Holt-Winters

Para séries mais complexas, existe o método de suavização exponencial sazonal de Holt-Winters (HW). De acordo com Morettin e Tolo (2006), há dois procedimentos que levam em consideração algumas características da série analisada.

#### (a) Série com Sazonalidade Multiplicativa

Considere um período sazonal  $m$ , o método HW considera o fator sazonal  $F_t$  sendo multiplicativa e a tendência aditiva, ou seja,

$$Y_t = \mu_t F_t + T_t + a_t \quad t = 1, \dots, N. \quad (2.2.5)$$

As equações de suavização são descritas por

$$\hat{F}_t = D \left( \frac{Y_t}{\bar{Y}_t} \right) + (1 - D)\hat{F}_{t-s}, \quad t = m + 1, \dots, N, \quad (2.2.6)$$

$$\bar{Y}_t = A \left( \frac{Y_t}{\hat{F}_{t-s}} \right) + (1 - A)(\hat{Y}_{t-1} + \hat{T}_{t-1}), \quad t = m + 1, \dots, N, \quad (2.2.7)$$

$$\hat{T}_t = C(\bar{Y}_t - \bar{Y}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad t = m + 1, \dots, N, \quad (2.2.8)$$

que dão estimativas do fator sazonal, nível e tendência, respectivamente.  $A$ ,  $C$  e  $D$  são constantes de suavização e possuem valores entre 0 e 1.

#### (b) Série com Sazonalidade Aditiva

Quando o fator de sazonalidade é aditivo,

$$Y_t = \mu_t + T_t + F_t + a_t. \quad (2.2.9)$$

E as novas estimativas são dadas por

$$\hat{F}_t = D(Y_t - \bar{Y}_t) + (1 - D)\hat{F}_{t-s}, \quad (2.2.10)$$

$$\bar{Y}_t = A(Y_t - \hat{F}_{t-s}) + (1 - A)(\bar{Y}_{t-1} + \hat{T}_{t-1}), \quad (2.2.11)$$

$$\hat{T}_t = C(\bar{Y}_t - \bar{Y}_{t-1}) + (1 - C)\hat{T}_{t-1}. \quad (2.2.12)$$

Com todos os elementos anteriormente definidos.

## 2.3 Modelos Estruturais

Segundo Morettin e Tolo (2006), o principal diferencial dos modelos estruturais é considerar as observações como combinações lineares de um nível e uma componente irregular, tal nível podendo representar tendências fixas ou aleatórias. Basicamente, ao abordar uma série temporal com esses modelos, ela será representada como uma combinação de componentes. É dada ênfase na formulação das componentes pelo conhecimento da série e a sugestão de tendências nas observações é obtida por análise gráfica. A seguir, temos os principais modelos estruturais.

### 2.3.1 Modelo de nível local

Considerado o modelo mais básico, é usado quando o nível local da série ( $\mu_t$ ) muda de acordo com um passeio aleatório, logo não é capaz de discernir tendências ou sazonalidades. Sendo assim, a previsão é constante, ou seja, a função previsão é simplesmente uma reta horizontal.

$$Z_t = \mu_t + \varepsilon_t \quad (2.3.1)$$

$$\mu_t = \mu_{t-1} + \eta_t \quad (2.3.2)$$

onde  $t = 1, \dots, N$  e  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ ,  $\eta_t \sim N(0, \sigma_n^2)$  são independentes e não correlacionadas (MORETTIN; TOLOI, 2006).

### 2.3.2 Modelo de tendência local

Para que seja possível avaliar tendências na série, deve-se adicionar uma componente de tendência ( $\beta_t$ ) ao nível local. Com isso, o modelo leva em consideração a

tendência da série em seu ajuste, e é descrito por

$$Z_t = \mu_t + \varepsilon_t, \quad (2.3.3)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad (2.3.4)$$

$$\beta_t = \beta_{t-1} + \xi_t, \quad (2.3.5)$$

onde  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ ,  $\eta_t \sim N(0, \sigma_\eta^2)$  e  $\xi_t \sim N(0, \sigma_\xi^2)$  com  $\eta_t$  e  $\xi_t$  mutuamente não correlacionados e não correlacionados com  $\varepsilon_t$ .  $\mu_t$  é o nível local e  $\beta_t$  a inclinação local (MORETTIN; TOLOI, 2006).

### 2.3.3 Modelo com tendência local e sazonalidade

Se for necessário, é possível incluir uma componente de sazonalidade no modelo. Sendo escrito na forma

$$Z_t = \mu_t + S_t + \varepsilon_t, \quad (2.3.6)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad (2.3.7)$$

$$\beta_t = \beta_{t-1} + \xi_t, \quad (2.3.8)$$

em que  $S_t$  é a componente sazonal,  $\mu_t$  é o nível local com tendência,  $\beta_t$  a inclinação local (ou componente de tendência) e  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  não correlacionado com os outros ruídos do modelo (MORETTIN; TOLOI, 2006).

## 2.4 Bagging

A técnica de agregação de bootstrap, conhecida como bagging, é bastante utilizada para reduzir a variância sem aumentar o viés das previsões, e com isso alcançar previsões pontuais mais precisas (HASTIE et al., 2009). Em um contexto de séries temporais, a principal dificuldade está em considerar a não estacionaridade e autocorrelação no procedimento bootstrap (PETROPOULOS et al., 2018). Considere a série de id 2107 do banco de dados da competição M3, segundo Hyndman e Athanasopoulos (2021), primeiramente é preciso decompor a série em tendência e sazonalidade, o que restar é considerado como resíduo. A decomposição dessa série é apresentada na Figura 1.



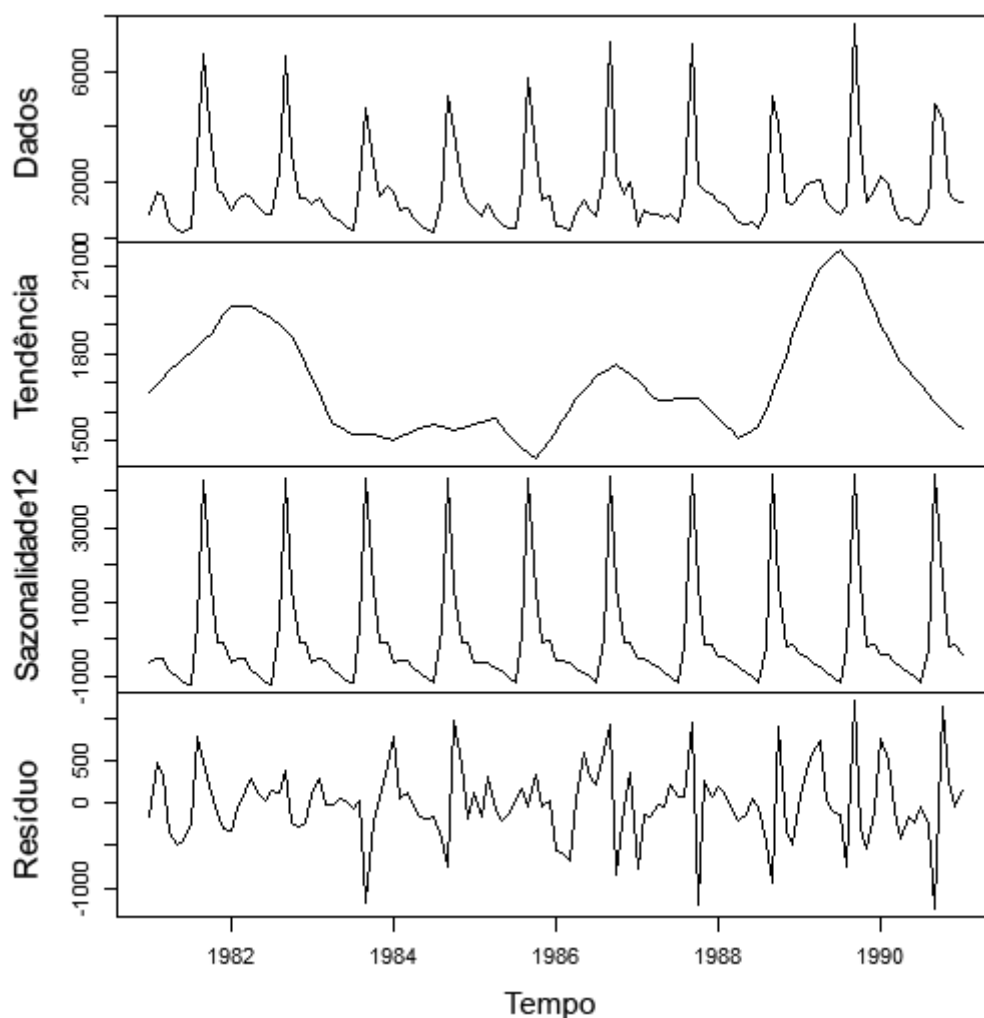


Figura 1: Decomposição da série de id 2107 do banco de dados M3

Ao obter os resíduos, realiza-se bootstrap em blocos para obter versões misturadas dos resíduos. Resumidamente, os resíduos são separados em "setores" ou "blocos" contíguos e os reorganizamos de forma aleatória. Um exemplo seria reorganizar, de forma aleatória, os resíduos relacionados ao mês de Janeiro de todos os anos envolvidos em dada série temporal. Esses resíduos são somados às demais componentes, resultando em variações da série original, como pode ser visto na Figura 2.

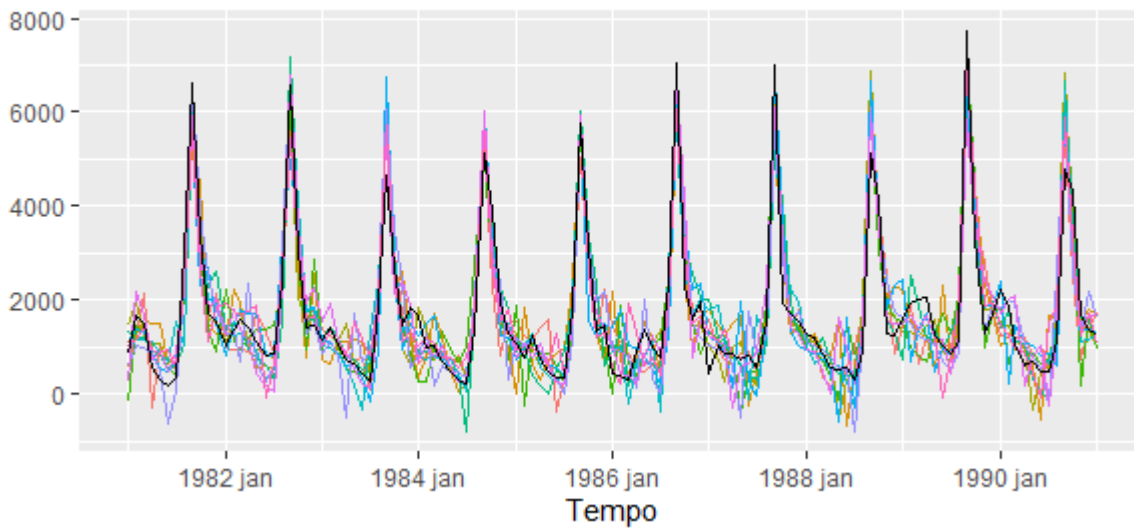


Figura 2: Exemplo de variações da série de id 2107

Para cada uma dessas séries, é realizado um ajuste de um modelo. Um modelo diferente pode ser selecionado para alguma variação, porém, é mais provável que todas as variações sejam ajustadas pelo mesmo modelo, dado que são séries semelhantes. Contudo, os parâmetros estimados não serão os mesmos, logo, as previsões serão diferentes mesmo que essas variações sejam ajustadas pelo mesmo modelo, como é possível observar na Figura 3.

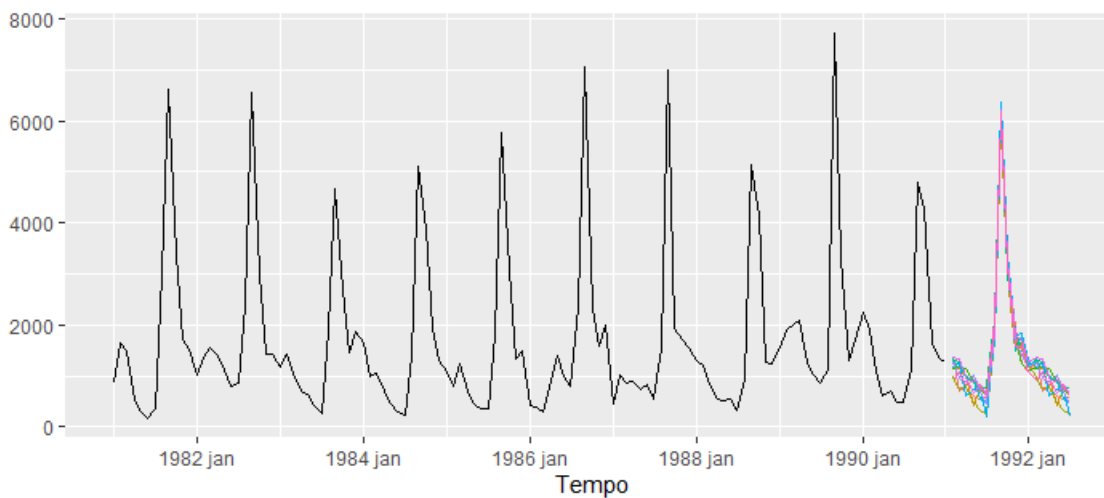


Figura 3: Exemplo de previsões pontuais das variações da série de id 2107

Por fim, levando em consideração todas as previsões, obtém-se um resultado médio para a previsão da série original (Figura 4), que é melhor (mais preciso) do que prever a série original diretamente. (HYNDMAN; ATHANASOPOULOS, 2021).



Figura 4: Intervalo de 95% de confiança para as previsões da série de id 2107 usando bagging

## 2.5 *Multiple comparisons with the best*

Quando a hipótese nula não é rejeitada em um teste estatístico, é aceito que todos os componentes dessa hipótese também sejam considerados como não rejeitados. Contudo, quando a hipótese nula é rejeitada surge a questão de quais componentes dessa hipótese também rejeitaram. Procedimentos estatísticos conhecidos como comparações múltiplas foram criadas com o objetivo de resolver esse problema. Um desses procedimentos, conhecido como *Multiple comparisons with the best* (MCB), será usado na comparação dos modelos.

Seja  $A_{nk}$  a acurácia do método  $k$  para a série temporal  $n$ , calculada usando SMAPE (que será apresentada posteriormente) ou qualquer outra métrica. Suponha que, para cada método  $k$ , o ranque seja a posição de  $A_{nk}$  entre  $A_{n1}, A_{n2}, A_{nK}$ , com  $K$  sendo o número total de métodos. Ao fazer esse ranqueamento para as  $n$  séries, pode-se obter o ranque médio  $\bar{R}_k$  para cada método. Considere a seguinte hipótese nula:

$$H_{0,k_1k_2} : \tau_{k_1} = \tau_{k_2}$$

onde  $k_1 = 1, 2, \dots, k_2 - 1$  e  $k_2 = 1, 2, \dots, K$  e  $\tau_j$  é o efeito do  $j$ -ésimo método. Os componentes de  $H_{0,k_1k_2}$  serão rejeitadas se, e somente se

$$|\bar{R}_{k_1} - \bar{R}_{k_2}| \geq r_{\alpha,K,N}$$

onde o valor crítico  $r_{\alpha,K,N}$  é determinado para manter o erro ao longo das comparações igual a  $\alpha$  (HSU, 1996).

De acordo com Hollander e Wolfe (1999), para um  $N$  grande:

$$r_{\alpha,K,N} \approx q_{\alpha,K} \sqrt{\frac{K(K+1)}{12N}}$$

onde  $q_{\alpha,K}$  é o percentil superior  $\alpha$  do intervalo da variável independente normal padrão  $K$ . Os valores de  $q_{\alpha,K}$  foram consultados na Tabela A.17 em Hollander e Wolfe (1999).

Como descrito em Koning et al. (2005), podemos visualizar os resultados do teste de uma maneira mais direta através de uma análise gráfica. Para cada método  $k$ , um intervalo é construído com comprimento  $r_{\alpha,K,N}$  e centrado em  $\bar{R}_k$ . Se os intervalos dos métodos  $k_1$  e  $k_2$  não se sobrepuserem, então  $H_{0,k_1k_2}$  é rejeitada. Para facilitar a visualização dos resultados é possível traçar uma linha na altura do limite superior do intervalo do melhor método, que naturalmente terá o menor limite superior entre todos os métodos. Assim, se o limite inferior do intervalo de outro método estiver acima dessa linha, rejeitará a hipótese nula e será considerado significativamente pior do que o melhor.

### 3 Metodologia

#### 3.1 Conjunto de dados

Em 1982, Spyros Makridakis liderou vários pesquisadores para realizarem uma competição com o intuito de comparar a acurácia dos mais diversos métodos de previsão de séries temporais. Essas competições ficaram conhecidas como Competições M.

As séries temporais da terceira competição de Makridakis, conhecida como M3, serão usadas para as comparações de performance dos modelos. Os resultados da competição foram publicados em 2000 (MAKRIDAKIS; HIBON, 2000). O banco de dados é composto por 3003 séries, distribuídas da seguinte forma: 1428 mensais, 756 trimestrais, 645 anuais e 174 em nenhuma das categorias anteriores.

Tabela 1: Decomposição do banco de dados M3

Intervalo	Micro	Indústria	Macro	Finanças	Demog	Outros	Total
Anual	146	102	83	58	245	11	645
Trimestral	204	83	336	76	57	0	756
Mensal	474	334	312	145	111	52	1428
Outro	4	0	0	29	0	141	174
Total	828	519	731	308	413	204	3003

Fonte: International Institute of Forecasters.

A Tabela 2 apresenta os horizontes de previsão utilizados em cada frequência das séries temporais presentes no banco de dados.

Tabela 2: Horizontes de previsão do banco de dados M3

Frequência	Horizonte de previsão	Quantidade de séries temporais
Anual	6	645
Trimestral	8	756
Mensal	18	1428
Outro	8	174
Total		3003

#### 3.2 Critério de informação de Akaike

Segundo Richard (2016), o critério de informação de Akaike (AIC) mensura a qualidade de um modelo levando em consideração a sua simplicidade. Sendo assim, é

uma métrica que permite comparar e selecionar modelos, onde os menores valores de AIC significam uma maior simplicidade e qualidade do modelo. O AIC é dado por:

$$AIC = 2k - 2\ln(\hat{L})$$

onde  $k$  é o número de parâmetros do modelo e  $\hat{L}$  o valor máximo da função de verossimilhança.

Entretanto, os modelos deste estudo foram selecionados por algoritmos de seleção automática no software *R* que se baseiam no critério de informação de Akaike corrigido (AICc). Quando uma amostra for pequena, ao usar o AIC é possível que modelos menos parcimoniosos sejam selecionado (BURNHAM; ANDERSON, 2002). Então, para evitar um possível sobreajuste utiliza-se uma correção no AIC:

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1}$$

em que  $n$  é o tamanho amostral. Perceba que quando  $n$  é grande,  $AICc$  se aproxima de  $AIC$ .

### 3.3 Modelos estruturais Bayesianos

Os modelos da família BSTS (Bayesian Structural Time Series) são, resumidamente, uma união entre os modelos estruturais de séries temporais, apresentados na seção anterior, e estatística bayesiana (ALMARASHI; KHAN, 2020). Primeiramente, as distribuições a priori dos parâmetros são definidas e a posteriori é obtida numericamente usando o algoritmo MCMC. Calcular essas distribuições pode ser desafiador, contudo, o pacote *bsts* no software *R* facilitará a obtenção dessas distribuições de forma numérica.

### 3.4 Markov Chain Monte Carlo

A integração Monte Carlo calcula  $E[f(X)]$  retirando amostras  $X_t$  ( $t = 1, \dots, n$ ) da distribuição posteriori  $\pi(\cdot)$  e então aproximando pela média amostral (CHAPMAN et al., 1996), ou seja,

$$E[f(X)] \approx \frac{1}{n} \sum_{t=1}^n f(X_t). \quad (3.4.1)$$

Logo, a média populacional é estimada pela média amostral. Pela lei dos grandes números, o tamanho da amostra  $n$  pode ser aumentado para aumentar a precisão da aproximação,

desde que as amostras  $X_t$  sejam independentes.

Retirar amostras independentes de  $\pi(\cdot)$  pode se revelar um verdadeiro desafio, já que a distribuição pode ser um tanto complicada. Contudo, as amostras não precisam ser independentes. De acordo com Chapman et al (1996), o conjunto das amostras  $\{X_t\}$  pode ser gerado por qualquer processo, desde que se retire amostras no suporte da distribuição posteriori nas proporções corretas. E uma forma de se fazer isso é usando uma cadeia de Markov com  $\pi(\cdot)$  sendo sua distribuição estacionária, dando origem aos métodos de Markov Chain Monte Carlo.

## 3.5 Performance

### 3.5.1 MAPE

As performances dos diferentes modelos na previsão da série exemplo foram comparadas com base no erro percentual médio absoluto (MAPE) por conta da sua fácil interpretação. É uma métrica de escala livre dada por:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (3.5.1)$$

onde  $Y_i$  e  $\hat{Y}_i$  são a  $i$ -ésima observação e previsão, respectivamente.

### 3.5.2 sMAPE

Ao realizar o estudo nas demais séries do banco de dados, o erro percentual médio absoluto simétrico (sMAPE) foi usado nas comparação visto que é uma das métricas usadas na publicação dos resultados da competição M3 (MAKRIDAKIS; HIBON, 2000) e, assim, futuras comparações podem ser realizadas. O sMAPE é dado por:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{(|Y_i| + |\hat{Y}_i|)/2} \quad (3.5.2)$$

com  $Y_i$  e  $\hat{Y}_i$  definidos anteriormente.

## 4 Resultados

### 4.1 Ilustração dos modelos para uma série temporal

Usando uma série do banco de dados M3 de exemplo, disponível no pacote *Mcomp* no *R*, será apresentado como o trabalho foi desenvolvido. O exemplo se iniciará com uma breve análise exploratória e será realizada modelagens com os métodos clássicos, juntamente da técnica de bagging. Concluindo com a modelagem da série utilizando um modelo estrutural bayesiano e a comparação dos desempenhos.

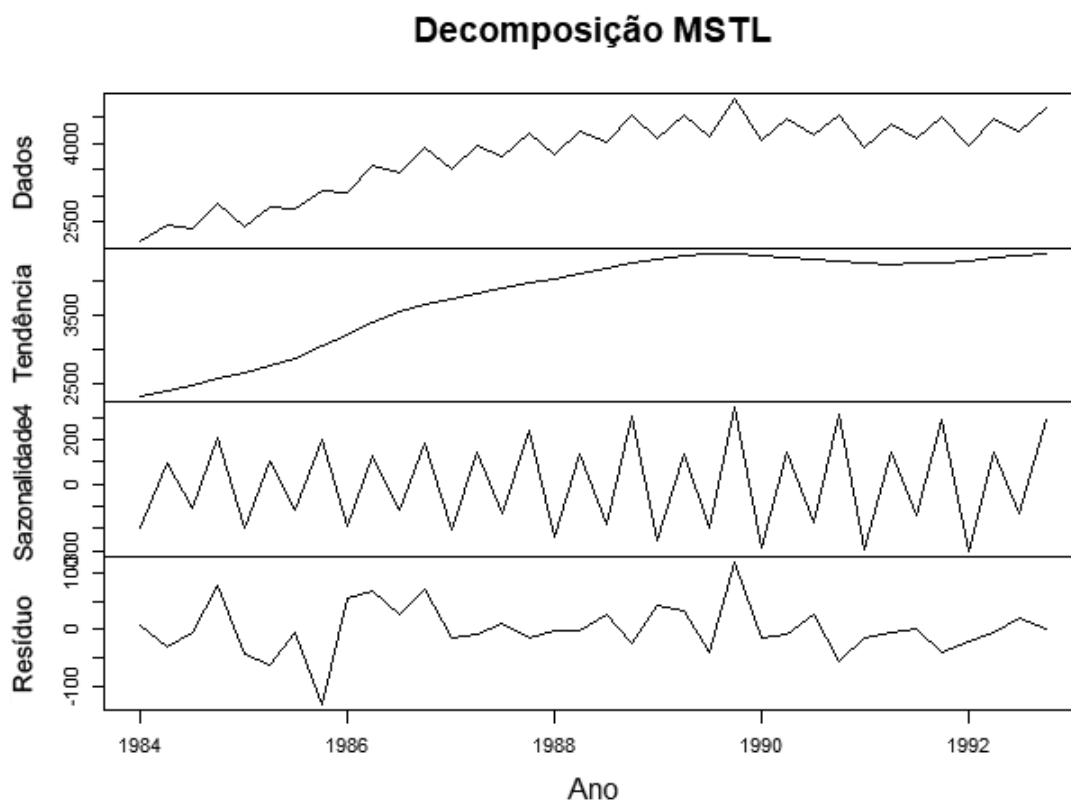


Figura 5: Decomposição MSTL da série de id 780

A série de número de vendas, com id 780, possui 44 observações no total e possui sazonalidade trimestral, ou seja, 4 observações por ano. Na Figura 5, a decomposição MSTL foi utilizada para separar as componentes que formam a série, e assim observa-se uma tendência positiva, onde as primeiras observações foram menores que 2500 unidades



e nos trimestres finais acima de 3500 unidades. Além disso, nota-se um comportamento sazonal de ciclo anual, o qual aparenta não sofrer grandes alterações ao longo do tempo.

De início, a série será ajustada por modelos ARIMA. Para tanto, a função de seleção automática de modelos do pacote *forecast* é aplicada. Por conta do grande número de séries presentes no banco de dados, a seleção dos modelos seria inviável manualmente. Para esta série, foi selecionado o modelo SARIMA  $(0, 1, 0) \times (0, 1, 1)_4$ . Então, o estudo segue adiante com esse modelo para prever o comportamento da série.

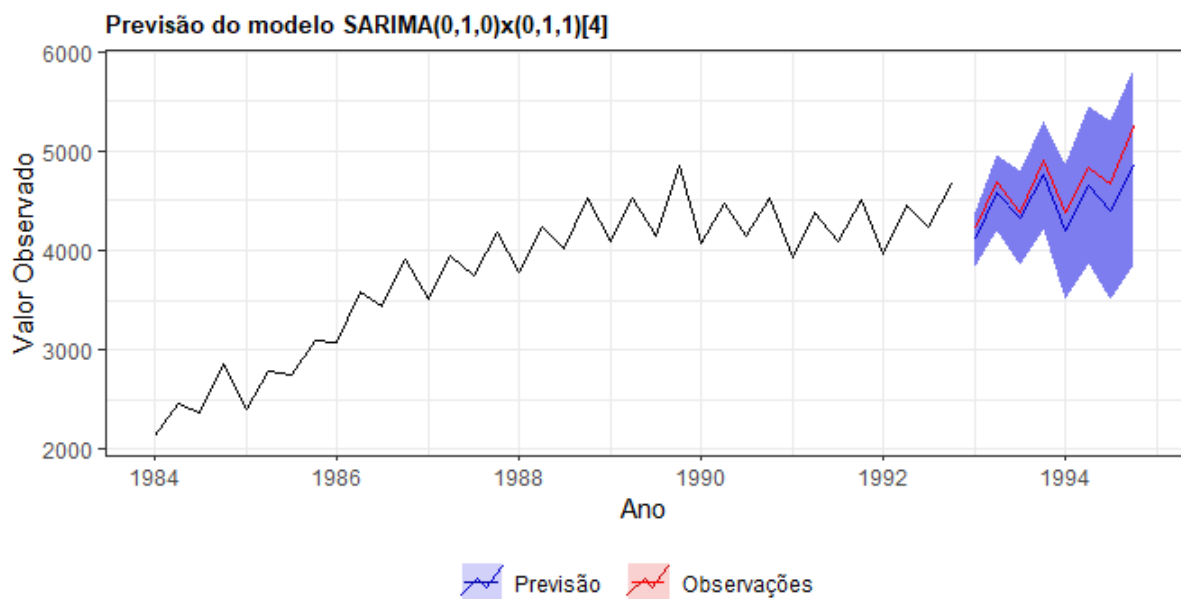


Figura 6: Previsão da série usando o modelo SARIMA, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% de confiança das previsões.

Na Figura 6, nota-se que todas as observações para validação do modelo estão contidas no intervalo de previsão. O que resulta em um MAPE de aproximadamente 3,8%, isso indica que a diferença entre a previsão do modelo e o valor observado é, em média, 3,8%, o que é considerado um resultado satisfatório. Seguindo adiante, a abordagem será feita com os modelos de alisamento exponencial de maneira similar: usando uma função de seleção automática de modelos do pacote *forecast*. O modelo selecionado foi ETS(M,A,M), o que resulta nas previsões abaixo.

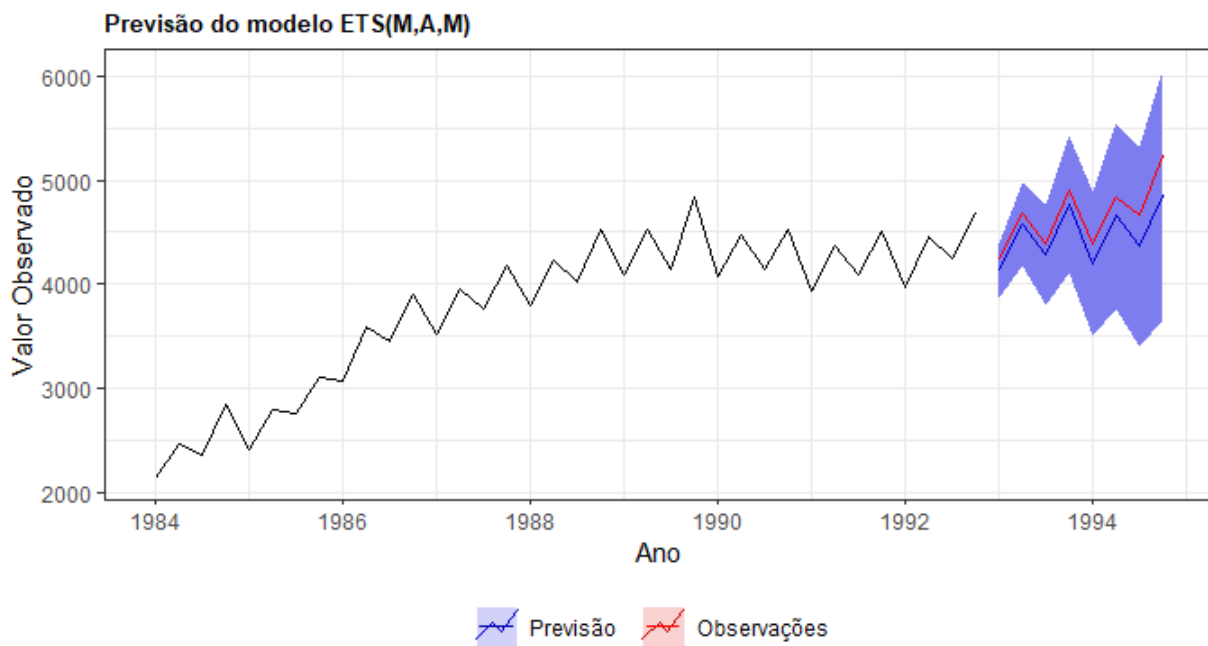


Figura 7: Previsão da série usando o modelo ETS, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões.

Na Figura 7, é possível observar que todas as observações de validação do modelo estão contidas nos intervalos de previsão do modelo. O MAPE resultante foi de 4%, ou seja, um desempenho inferior, porém próximo do modelo SARIMA. Agora será utilizado o bagging para calcular as previsões desses modelos.

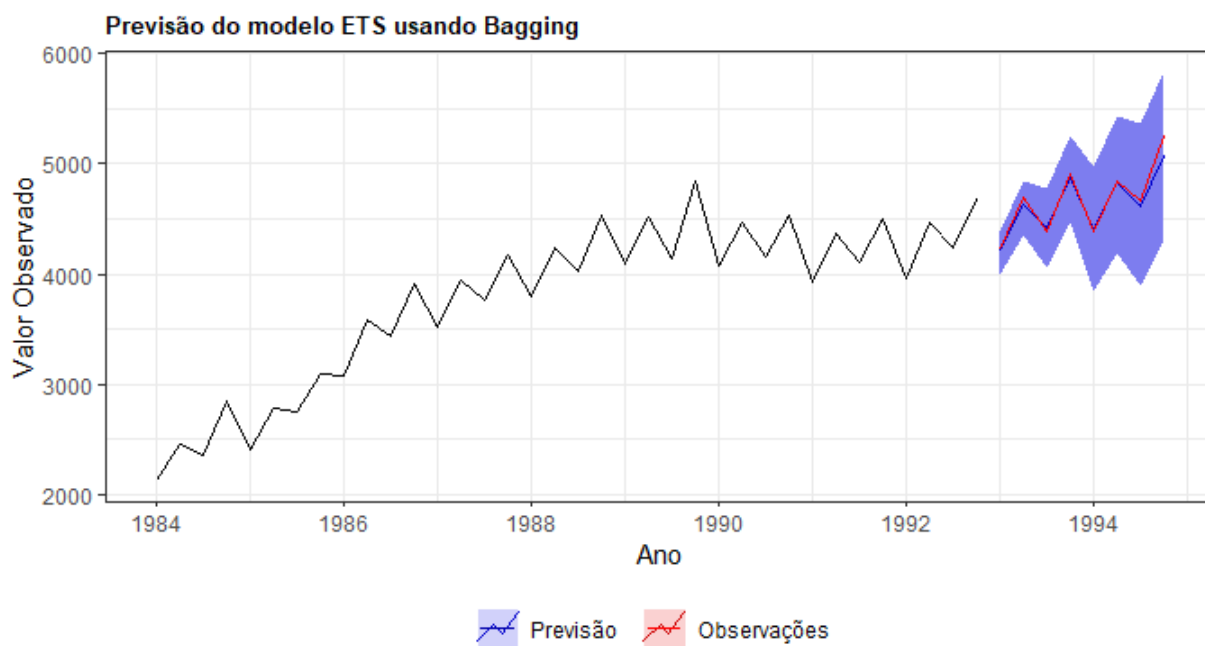


Figura 8: Previsão da série usando o modelo ETS com bagging, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões.

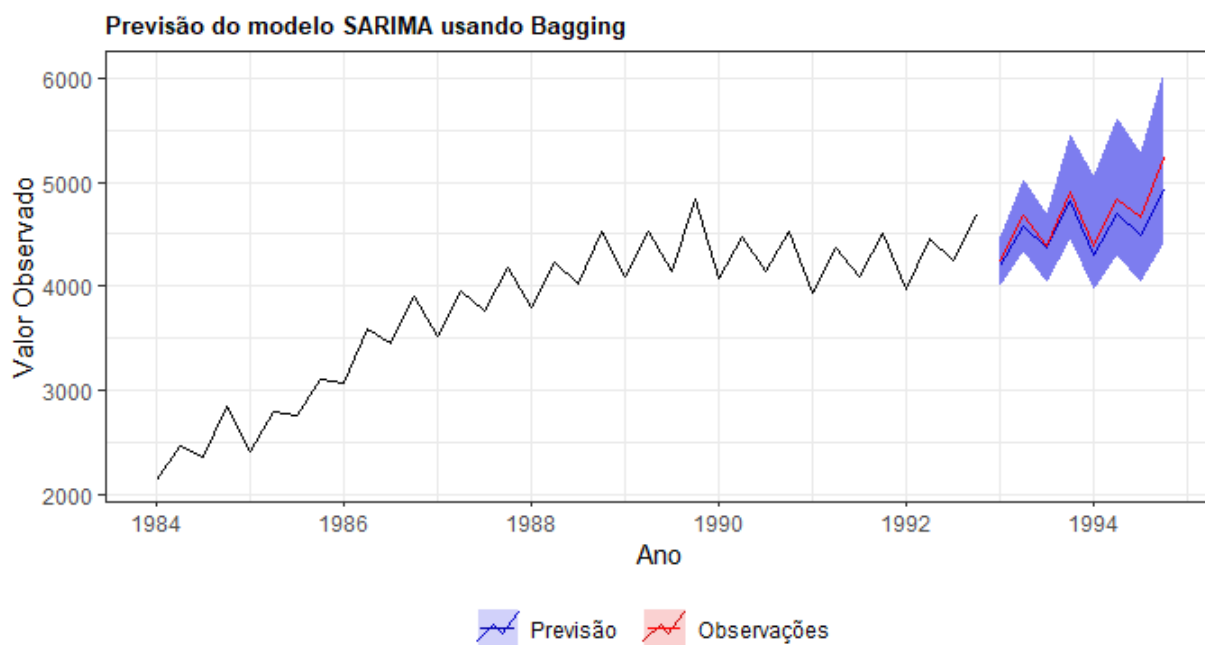


Figura 9: Previsão da série usando o modelo SARIMA com bagging, com horizonte de 8 trimestres. As previsões e os dados observados estão em azul e vermelho, respectivamente. A área azulada representa o intervalo de 95% confiança das previsões.

É notório o melhor desempenho das previsões usando bagging nas Figuras acima, o erro percentual médio absoluto calculado foi de 2,5% e 1% para o modelo SARIMA e ETS, respectivamente.

Por fim, um modelo BSTS será ajustado para os dados em análise. Neste caso, um modelo com componentes de tendência local e sazonalidade deve ser o mais adequado. Na Figura 10, a suposição de normalidade pode estar sendo violada devido à valores extremos nas caudas, já na autocorrelação, o 0 está contido na distribuição dos lags e isso indica que a suposição de autocorrelação nula está sendo obedecida.

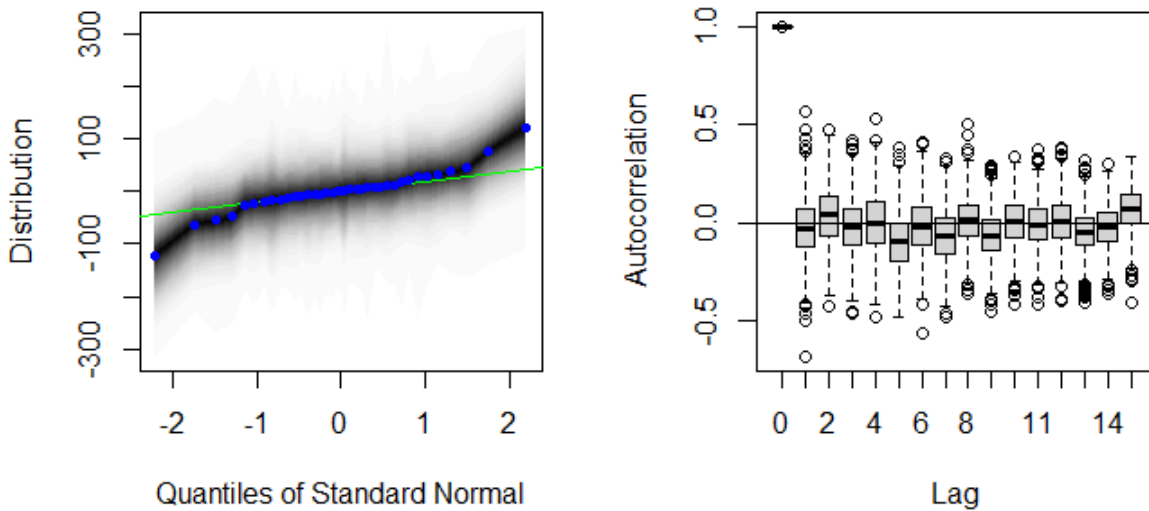


Figura 10: Gráfico Q-Q para normalidade e gráfico da função de autocorrelação (ACF) dos resíduos do modelo de tendência local com sazonalidade bayesiano

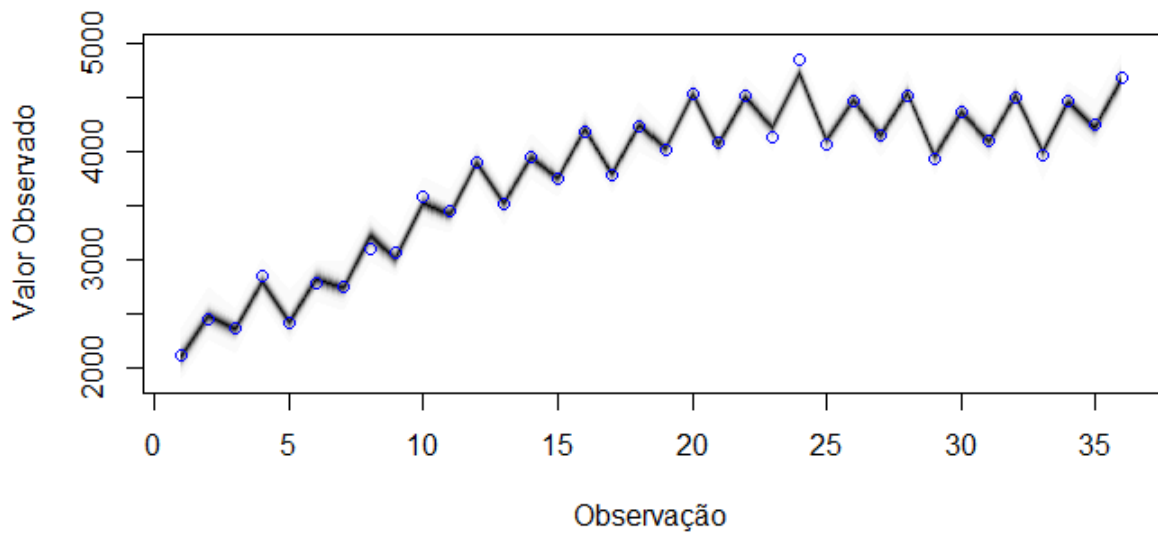


Figura 11: Ajuste do modelo BSTS de tendência local com sazonalidade à série analisada

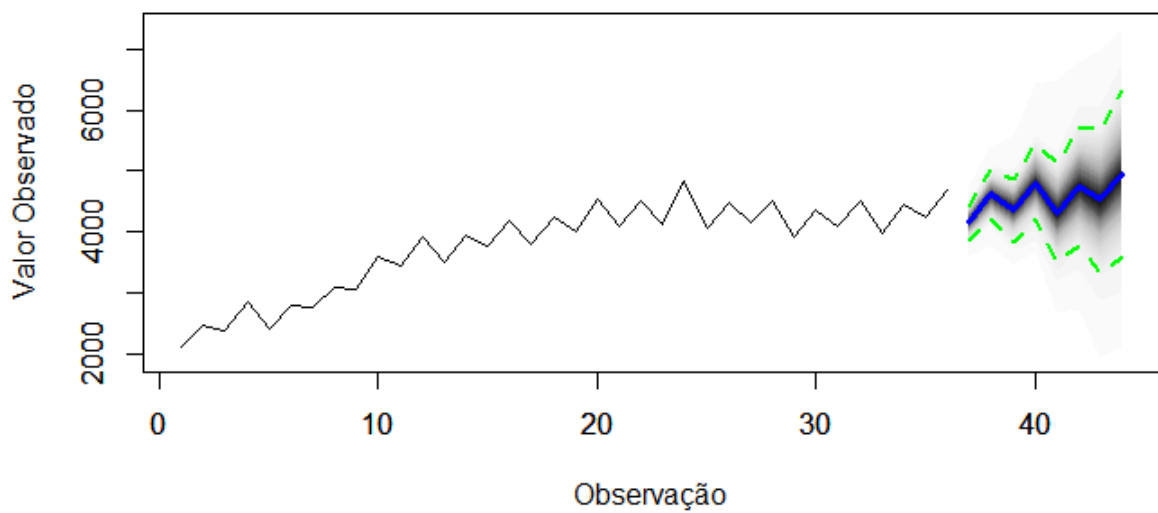


Figura 12: Previsão da série usando o modelo BSTS de tendência local com sazonalidade, num horizonte de 8 trimestres. A linha azul é a mediana das previsões, a parte sombreada é a densidade da distribuição à posteriori e o intervalo de 95% de confiança é representado pelas linhas verdes.

Na Figura 11, é apresentado o ajuste do modelo bayesiano à série fornecida. As linhas representam o ajuste do modelo aos dados enquanto os círculos são as observações da série original. O modelo se ajustou bem aos dados. Na Figura 12, tem-se as previsões de 8 observações, ou trimestres, e foi obtido um MAPE de, aproximadamente, 2,1%. Na Tabela 3, é realizada a comparação do MAPE entre todos os modelos para esta série. O modelo BSTS se saiu melhor que os modelos clássicos e o ARIMA com bagging, o modelo de alisamento exponencial somado à técnica de bagging teve o melhor desempenho, com um MAPE de 1%. É interessante perceber como o bagging melhorou o desempenho das previsões nos modelos ARIMA e ETS em troca de uma exigência computacional maior já que consideramos 100 variações da série original, assim como Hyndman e Athanasopoulos (2021) comentaram em seu livro.

Tabela 3: MAPE calculado em cada modelo

Modelo	MAPE (%)
<b>ETS Bagging</b>	1,0
<b>BSTS</b>	2,1
<b>ARIMA Bagging</b>	2,5
<b>ARIMA</b>	3,8
<b>ETS</b>	4,0

## 4.2 Avaliação dos modelos

Agora, o estudo será expandido para todas as 3003 séries do banco de dados. Para realizar essa tarefa, foi usado um computador equipado com um processador Intel i5-8400, 24GB de RAM e sistema operacional Windows 11. Para resumir todas as categorias de séries temporais, foi utilizado uma média ponderada pelos horizontes de previsão mencionados na Tabela 2. Os resultados para as 3003 séries são apresentados na Tabela 4.

Tabela 4: Resultados empíricos usando sMAPE em cada modelo (%)

Métodos	Anual	Trimestral	Mensal	Outro	Todos	Tempo (min)
<b>ARIMA</b>	17,10	10,00	14,96	4,51	13,98	43,94
<b>ETS</b>	17,00	9,68	14,13	4,37	13,34	16,78
<b>BSTS</b>	17,23	10,18	14,67	4,71	13,83	26,34
<b>ARIMA Bagging</b>	17,42	9,85	14,04	4,64	13,36	410*
<b>ETS Bagging</b>	17,23	9,83	13,69	4,70	13,12	330*

\* Tempo de computação aproximado.

Considerando os métodos clássicos, percebe-se que em todas as categorias os modelos ETS obtiveram menores valores de sMAPE. Ao considerar os modelos bayesianos, nota-se um desempenho superior aos modelos ARIMA nas séries mensais, e também quando todas as séries são analisadas ao mesmo tempo, o que nos fornece bons indicativos da viabilidade desses modelos como uma alternativa aos modelos clássicos. Os que obtiveram melhores resultados foram os modelos clássicos usando bagging para construir os intervalos de previsão, um resultado esperado. Entre eles, o ETS com bagging se mostrou o melhor, com um sMAPE de 13,09%.

Ainda na Tabela 4, já eram esperados longos tempos de computação nos modelos em que as previsões foram calculadas usando a técnica de bagging. Os modelos ETS e ARIMA, aplicando bagging, realizaram a tarefa em, aproximadamente, 5,5 e 6,8 horas, respectivamente. Estes tempos foram obtidos usando o pacote *foreach* no R que permite o uso de mais de um núcleo do processador na computação de um *for loop*, diminuindo o tempo necessário para a conclusão de um trabalho computacionalmente intensivo. Os modelos BSTS finalizaram a previsão de todas as séries em 26,34 minutos, tempo inferior aos modelos ARIMA, 43,94 minutos, e superior aos ETS, que concluíram a computação em 16,78 minutos. Como o pacote *bsts* é recente, ainda não existe um algoritmo para seleção de modelos automática, ou seja, o algoritmo utilizado neste trabalho pode não ser o mais eficiente e os modelos BSTS conseguirem a realizar a tarefa em um tempo menor ao relatado acima.

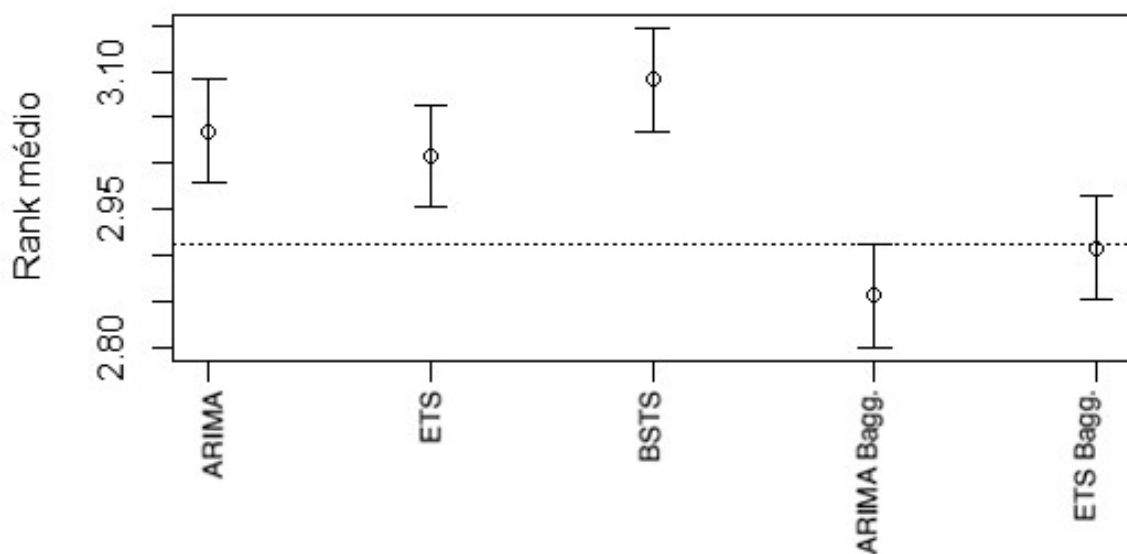


Figura 13: Intervalos MCB para todos os modelos em estudo. Considerando um nível de significância de 5%.

Ademais, o teste MCB foi realizado para as 3003 séries com o objetivo de comparar os modelos estatisticamente. Neste teste, o intervalo do ranque foi construído para cada modelo usando o erro percentual médio absoluto, que resultará em ranques equivalentes aos do sMAPE. Os métodos que possuem os intervalos do ranque médio totalmente acima da linha de referência obtiveram desempenhos significativamente piores do que o melhor modelo. Foi considerado um nível de significância de 5%.

Na Figura 13, a linha de referência é o limite superior do intervalo do ARIMA bagging, ou seja, é considerado o melhor modelo no teste. E os modelos ARIMA, ETS e BSTS estão todos acima dessa linha, sendo assim, desempenharam significativamente pior do que o melhor modelo. Ambos modelos com bagging tiveram os menores ranques médios. É discutível se o aumento significativo no tempo de computação é vantajoso dado os resultados, porém, vemos que esses modelos obtiveram previsões mais robustas do que os demais.

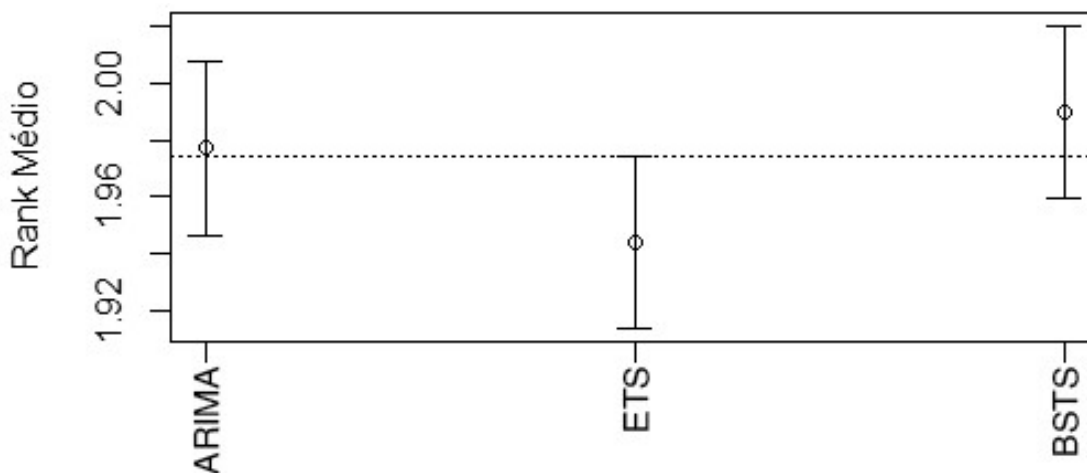


Figura 14: Intervalos MCB para os modelos selecionados. Considerando um nível de significância de 5%.

Considerando apenas os modelos sem bagging, vemos na Figura 14 que não houveram evidências suficientes de que os modelos ARIMA e BSTS obtiveram desempenhos significativamente piores do que os modelos de alisamento exponencial nesse estudo. O que condiz com os resultados apresentados na Tabela 4, em que os modelos BSTS e ARIMA tiveram performances semelhantes e os modelos ETS tiveram melhores resultados. Tais resultados fortalecem os bons indícios da viabilidade na previsão de séries temporais dos modelos bayesianos apresentados neste trabalho.



## 5 Conclusão

Neste estudo, foram apresentados modelos e métodos que, geralmente, não são lecionados durante a graduação. No caso exemplo, o procedimento bayesiano foi esmiuçado para demonstrar como esses modelos funcionam e, em seguida, as métricas de performance calculadas. Com isso, conclui-se que, na série de id 780, o modelo BSTS se saiu melhor que os modelos clássicos ARIMA, ETS e ARIMA com bagging com um MAPE de 2,1%. O modelo que teve os melhores resultados foi o ETS com bagging que obteve um MAPE de 1%.

Ao expandir o estudo às demais séries, o modelo BSTS teve um desempenho superior aos ARIMA em séries mensais, e como estas são a maior parte do banco de dados, o desempenho geral também foi superior. Entre os modelos sem bagging, os modelos de alisamento exponencial foram os melhores posicionados com um sMAPE geral de 13,34%, além disso, foram superiores em todas as outras categorias. Com esses resultados, o modelo BSTS pode ser uma alternativa viável aos modelos clássicos em situações que seja mais interessante uma abordagem bayesiana.

A revisão de literatura apontou que utilizar o bagging sempre resultaria em previsões mais precisas. Os resultados observados neste relatório também apontam nesse sentido, observou-se um aumento considerável na performance dos modelos ARIMA e ETS ao realizarem as previsões usando essa técnica. Calculando as previsões desta maneira, obteve-se sMAPE geral de 13,36% e 13,12% para os modelos ARIMA e ETS, respectivamente. Sendo assim, considerando os resultados usando a previsão tradicional, realmente houve uma melhora na precisão das previsões. Tal melhora veio a custo de um grande aumento no tempo de computação.

Ao realizar comparações múltiplas com o melhor, o modelo ARIMA com bagging obteve o menor ranque médio entre os demais métodos. E assim, com base no teste estatístico, não houveram evidências suficientes de que o modelo ETS com bagging seja significativamente pior. Em contraste, os modelos ARIMA, ETS e BSTS todos rejeitaram a hipótese nula, ou seja, houveram evidências de que esses modelos sejam significativamente piores do que o melhor modelo. É questionável se o aumento considerável no tempo de computação vale a pena em relação ao desempenho apresentado, porém, é inegável que houve uma melhora.

Em estudos futuros, seria interessante verificar o comportamento desses modelos bayesianos em outros bancos de dados, por exemplo, o banco da quarta competição (M4)

que conta com 100 000 séries temporais. Além disso, verificar o desempenho em previsões intervalares - já que, usualmente, são usadas previsões pontuais - e o comportamento em casos reais e nas mais diversas situações: séries com a presença de componentes cíclicos, covariáveis, entre outros. E assim, aprofundar em quais situações os modelos bayesianos são viáveis como alternativa aos clássicos.

## Referências

- ALMARASHI, A.; KHAN, K. Bayesian structural time series. *Nanoscience and Nanotechnology Letters*, v. 12, p. 54–61, Jan. 2020. Disponível em: <https://doi.org/10.1166/nnl.2020.3083>.
- BERGMEIR, C. et al. Bagging exponential smoothing methods using stl decomposition and box-cox transformation. *International Journal of Forecasting*, s.l, v. 32, n.2, p. 303–312, Jun. 2016. DOI 10.1016/j.ijforecast. Acesso em: 8 de jul. 2022. Disponível em: <https://doi.org/10.1016/j.ijforecast.2015.07.002>.
- BROCKWELL, P. J.; DAVIS, R. *Introduction to Time Series and Forecasting*. 2. ed. Colorado: Springer, 2001. 83-283 p.
- BURNHAM, K. P.; ANDERSON, D. R. *Model selection and multimodel inference : a practical information-theoretic approach*. 2. ed. Nova Iorque: Springer, 2002. 488 p.
- CHAPMAN et al. *Markov Chain Monte Carlo in practice*. 1. ed. Estados Unidos da América: Taylor & Francis, 1996. 1-12 p.
- FEROZE, N. Forecasting the patterns of covid-19 and causal impacts of lockdown in top five affected countries using bayesian structural time series models. *Chaos, Solitons Fractals*, Elsevier, Paquistão, v. 140, Dez. 2020. DOI 10.1016/j.chaos.2020.110196. Acesso em: 9 de jul. 2022. Disponível em: <https://doi.org/10.1016/j.chaos.2020.110196>.
- HARVEY, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. 1. ed. Nova Iorque: Cambridge University Press, 1990. 572 p.
- HASTIE et al. *The elements of statistical learning*. 2. ed. Nova Iorque: Springer, 2009. 745 p. DOI 10.1007/978-0-387-84858-7. Acesso em: 14 de jul. 2022. Disponível em: <https://doi.org/10.1007/978-0-387-84858-7>.
- HOLLANDER, M.; WOLFE, D. A. *Nonparametric Statistical Methods*. 2. ed. Nova Iorque: Wiley-Interscience, 1999. 806 p.
- HSU, J. C. *Multiple comparisons. Theory and methods*. 1. ed. Londres: Chapman and Hall, 1996. 27 p.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 3. ed. Melbourne, Australia: OTexts, 2021. Acesso em: 8 de jul. 2022. Disponível em: <https://otexts.com/fpp3/>.
- KONING, A. J. et al. The m3 competition: Statistical tests of the results. *International Journal of Forecasting*, v. 21 n.3, p. 397–409, Jul. - Set. 2005. DOI 10.1016/j.ijforecast.2004.10.003. Acesso em: 28 de Dez. 2022. Disponível em: <https://doi.org/10.1016/j.ijforecast.2004.10.003>.
- MAKRIDAKIS, S.; HIBON, M. The m3-competition: results, conclusions and implications. *International Journal of Forecasting*, v. 16 n.4, p. 451–476, Out. - Dez. 2000. Acesso em: 10 de Set. 2022. Disponível em: [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1).

- MCQUIRE, C. et al. Forecasting the 2021 local burden of population alcoholrelated harms using bayesian structural time-series. *Addiction*, s.l., v. 114, p. 994–1003, Jan. 2019. DOI 10.1111/add.14568. Acesso em: 10 de jul. 2022. Disponível em: [⟨https://doi.org/10.1111/add.14568⟩](https://doi.org/10.1111/add.14568).
- MORETTIN; TOLOI. *Análise de séries temporais*. 2. ed. São Paulo, Brasil: Egard Blucher, 2006. 538 p.
- PETROPOULOS, F. et al. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, v. 268 n.2, p. 545–554, Jan. 2018. Acesso em: 14 de jul. 2022. Disponível em: [⟨https://doi.org/10.1016/j.ejor.2018.01.045⟩](https://doi.org/10.1016/j.ejor.2018.01.045).
- RICHARD, M. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 1. ed. S.I: CRC Press., 2016. 189 p.
- SCOTT, S. L.; VARIAN, H. Predicting the present with bayesian structural time series. SSRN, s.l., Ago. 2013. DOI 10.2139/ssrn.2304426. Acesso em: 9 de jul. 2022. Disponível em: [⟨https://dx.doi.org/10.2139/ssrn.2304426⟩](https://dx.doi.org/10.2139/ssrn.2304426).