



Universidade de Brasília
Departamento de Estatística

Bagging em modelos autoregressivos de duração condicional

Mateus Barros Kilson

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

**Brasília
2024**

Mateus Barros Kilson

Bagging em modelos autoregressivos de duração condicional

Orientador(a): Prof. Helton Saulo Bezerra dos Santos

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

**Brasília
2024**

Dedico este trabalho aos meus pais, Guilherme e Licia. Nada disso estaria acontecendo sem a presença deles na minha vida. Sem dúvidas, são meus maiores incentivadores para a conclusão do meu curso. Eles custearam todas as minhas necessidades, proporcionando um ambiente em que eu pudesse me dedicar exclusivamente aos meus estudos, com o intuito de obter um futuro melhor. Muito obrigado, mãe e pai.

Agradecimentos

- Gostaria de agradecer ao professor Helton, que sempre esteve disposto a me orientar, mesmo com outras demandas. Ele se dispôs a entrar nessa caminhada comigo, apresentando o tema, objetivos e referências fundamentais para o desenvolvimento deste trabalho. Sua orientação e apoio foram essenciais para a conclusão deste projeto. Muito obrigado, professor Helton, por todo o suporte e dedicação.

Resumo

O principal objetivo deste trabalho é possibilitar a previsão do tempo de mudança de preço dos ativos BASF SE, AAPL e BAYN, utilizando modelos autorregressivos de duração condicional (ACD) com e sem a presença do método bagging. Além disso, como iremos realizar diversas previsões, é de grande interesse comparar a capacidade e a qualidade de predição de cada modelo gerado. Para cada modelo ACD, são estimados parâmetros a partir de uma distribuição escolhida. Em seguida, realizamos predições para os dados e, por fim, as comparamos. Utilizaremos as distribuições Exponencial, Weibull, Burr e Gama-Generalizada sendo que para cada modelo ACD associado a uma distribuição, será gerado um modelo correspondente com a mesma distribuição, mas com a presença do método bagging.

Palavras-chaves: Big Data; Modelagem de Alta Frequência; Dados Financeiros; Weibull; Burr; Gama Generalizada; Generalizada F; Modelos de Previsão; Bagging; Duração entre Eventos Financeiros; Complexidade dos Dados Financeiros.

Lista de Tabelas

1	Distribuições do erro e resultados do teste Ljung-Box associados	19
2	Estatísticas das previsões, modelos EACD, Apple.	21
3	Estatísticas das previsões, modelos WACD, Apple.	22
4	Estatísticas das previsões, modelos BACD, Apple.	22
5	Estatísticas das previsões, modelos GGACD, Apple.	23
6	Distribuições do erro e resultados do teste Ljung-Box associados	25
7	Estatísticas das previsões, modelos EACD, Bayer.	27
8	Estatísticas das previsões, modelos WACD, Bayer.	28
9	Estatísticas das previsões, modelos BACD, Bayer.	28
10	Estatísticas das previsões, modelos GGACD, Bayer.	29
11	Distribuições do erro e resultados do teste Ljung-Box associados	31
12	Estatísticas das previsões, modelos EACD, BASF SE.	34
13	Estatísticas das previsões, modelos WACD, BASF SE.	34
14	Estatísticas das previsões, modelos BACD, BASF SE.	34
15	Estatísticas das previsões, modelos GGACD, BASF SE.	35

Lista de Figuras

1	Ilustração da técnica Bagging.Fonte:(CAMARGO, 2021)	15
2	Durações observadas, Apple	18
3	Durações sem sazonalidade, Apple	18
4	QQ-plot exponencial	20
5	QQ-plot weibull	20
6	QQ-plot burr	20
7	QQ-plot gamma-generalizada	20
8	Conjunto de QQ-plots para diferentes distribuições	20
9	Durações reamostradas com blocos de tamanho 50, Apple	21
10	Previsão do Modelo EACD comparada com dados reais, Apple.	21
11	Previsão do Modelo WACD comparada com dados reais, Apple.	22
12	Previsão do Modelo BACD comparada com dados reais, Apple.	22
13	Previsão do Modelo GGACD comparada com dados reais, Apple.	23
14	Durações observadas, Bayer	24
15	Durações sem sazonalidade, Bayer	24
16	QQ-plot exponencial	26
17	QQ-plot weibull	26
18	QQ-plot burr	26
19	QQ-plot gamma-generalizada	26
20	Conjunto de QQ-plots para diferentes distribuições	26
21	Durações reamostradas com blocos de tamanho 50, Bayer	27
22	Previsão do Modelo EACD comparada com dados reais, Bayer.	27
23	Previsão do Modelo WACD comparada com dados reais, Bayer.	28
24	Previsão do Modelo BACD comparada com dados reais, Bayer.	28
25	Previsão do Modelo GGACD comparada com dados reais, Bayer.	29
26	Durações observadas, BASF SE	30

27	Durações sem sazonalidade, BASF SE	30
28	QQ-plot exponencial	32
29	QQ-plot weibull	32
30	QQ-plot burr	32
31	QQ-plot gamma-generalizada	32
32	Conjunto de QQ-plots para diferentes distribuições	32
33	Durações reamostradas com blocos de tamanho 50, BASF SE	33
34	Previsão do Modelo EACD comparada com dados reais, BASF SE.	34
35	Previsão do Modelo WACD comparada com dados reais, BASF SE.	35
36	Previsão do Modelo BACD comparada com dados reais, BASF SE.	35
37	Previsão do Modelo GGACD comparada com dados reais, BASF SE.	36

Sumário

1 Introdução	8
2 Metodologia	9
2.1 Conjunto de dados	9
2.2 Modelos ACD.	9
2.3 Distribuição associada ao erro.	10
2.3.1 Erro com distribuição Exponencial	11
2.3.2 Erro com distribuição Weibull	12
2.3.3 Erro com distribuição Burr	13
2.3.4 Erro com distribuição Gamma-Generalizada	13
2.4 Adequação dos Modelos	14
2.5 Método Bagging	14
2.6 Previsão	15
2.7 Comparação dos Modelos	16
3 Resultados	18
3.1 Apple	18
3.1.1 Durações	18
3.1.2 Modelos e Adequações	19
3.1.3 Bagging	19
3.1.4 Previsões	20
3.2 Bayer	24
3.2.1 Durações	24
3.2.2 Modelos e Adequações	25
3.2.3 Bagging	25
3.2.4 Previsões	26
3.3 BASF SE	30
3.3.1 Durações	30

3.3.2	Modelos e Adequações	31
3.3.3	Bagging	33
3.3.4	Previsões	33
4	Conclusão	37
	Referências	38
	Apêndice	39

1 Introdução

Devido o aumento na quantidade de investidores no mercado financeiro, a demanda por informações tem sido cada vez mais frequente. Com a proposta dos modelos ACD (Autorregressive Conditional Duration) por Robert Engle e Russel em 1998 (ENGLER; RUSSELL, 1998), a modelagem e previsão da duração entre eventos financeiros tornaram-se uma das formas de fornecer informações relevantes nesse vasto mercado. Apesar de sua eficiência satisfatória, a busca por melhorias no meio acadêmico é constante. A aplicação de métodos que gerem resultados mais precisos nos leva ao método Bagging (Bootstrap Aggregating), que, quando associado aos modelos ACD, pode proporcionar resultados mais precisos.

Os modelos ACD são amplamente utilizados na previsão da duração dos intervalos entre eventos, como alterações no preço dos ativos. Essas mudanças de preço podem ser consequências da quantidade de informação disponível sobre o ativo e das movimentações financeiras significativas. Em contraste, a longa duração desses intervalos pode refletir a falta de informações e poucas movimentações relacionadas ao ativo.

A técnica de Bagging, introduzida por Leo Breiman em 1996 (BREIMAN, 1996), é um método de reamostragem que visa melhorar a previsão de dados. Essa técnica envolve o desenvolvimento de vários conjuntos de dados utilizando reamostragem com reposição (Bootstrap), onde cada conjunto é usado para treinar um modelo independente. Posteriormente, as previsões de cada modelo são agregadas para formar uma previsão final mais precisa.

Inspirado no trabalho realizado por Morais em 2023 (MORAIS, 2023), que comparou a modelagem e previsão de dados financeiros utilizando a distribuição Gamma-Generalizada com e sem a aplicação do método Bagging, este estudo pretende realizar uma comparação entre modelos ACD com e sem a utilização do método Bagging, variando as distribuições utilizadas nos modelos e o tamanho dos blocos na amostra.

Para a realização deste trabalho, utilizaremos as distribuições Exponencial, Weibull, Burr e Gamma-Generalizada, por já possuírem implementação no pacote R utilizado. Para cada distribuição, serão atribuídos parâmetros diferentes ao modelo, tanto com quanto sem a presença do método bagging, considerando que cada distribuição possui suas especificidades.

2 Metodologia

2.1 Conjunto de dados

Este trabalho será desenvolvido utilizando os dados de mercado referentes às empresas, BASF SE, uma multinacional cujo principal foco é a produção de produtos químicos, AAPL, referente a Apple uma empresa multinacional americana que projeta, fabrica e vende eletrônicos de consumo e BAYN, referente a Bayer que é uma multinacional alemã voltada principalmente para os setores de saúde e agricultura. Como não há necessidade de utilizar todas as variáveis presentes em nossos bancos de dados iremos selecionar as variáveis referentes ao preço, horário local e volume.

Realizaremos os ajustes necessários utilizando o R para a montagem e melhor manipulação da série temporal. A variável preço será calculada pela média dos valores originais ask e bid que representam os maiores de oferta e compra no momento observado. O horário, originalmente fornecido em GMT, também precisará ser ajustado para o horário local para o desenvolvimento da série.

Utilizaremos a função `computeDurations` para a formação da nossa série. Essa função utiliza os parâmetros `open` e `close`, que se referem ao horário de abertura e fechamento da janela de tempo que iremos utilizar, `type`, que se refere ao tipo de dado, `priceDiff` e `cumVol`, que se referem às diferenças de preço e volumes acumulados que consideraremos significativos para o desenvolvimento da série. Uma vez que a série estiver criada, utilizaremos a função `diurnalAdj` para a decomposição da série, por ser uma função especificamente projetada para lidar com dados financeiros diurnos.

2.2 Modelos ACD

Os modelos ACD são modelos autoregressivos de duração condicional, inicialmente propostos para lidar com dados financeiros e possuem diversas variações, para o desenvolvimento deste trabalho nos iremos utilizar o que é chamado de standard ACD como citado por Mayorov (MAYOROV, 2011) onde temos que t_i é o tempo em que ocorre o i -ésimo evento de interesse, o tempo é contado a partir de um momento considerado como o momento inicial t_0 . Dessa maneira definimos a variável x_i

$$x_i = t_i - t_{i-1}, i = 1, 2, 3, \dots \quad (2.2.1)$$

o modelo ACD propõe que:

$$x_i = \psi_i \epsilon_i \quad (2.2.2)$$

sendo ϵ_i uma variável independente indenticamente distribuída (i.i.d) com $E(\epsilon_i)=1$ e suporte positivo, e

$$\psi_i = \omega + \sum_{v=1}^q \beta_v x_{i-v} + \sum_{j=1}^p \alpha_j \psi_{i-j}, \quad (2.2.3)$$

sendo:

- ψ_i é a duração condicional esperada no tempo t_i ,
- ω é um termo constante positivo,
- x_i são as durações passadas observadas,
- β_v e α_j são parâmetros não negativos do modelo,
- q é a ordem do processo autorregressivo, sendo um inteiro positivo,
- p é a ordem do processo de médias móveis, sendo um inteiro positivo.

Uma vez que x_i é exclusivamente positivo podemos assumir que $\omega > 0$, $\beta_v \geq 0$ e $\alpha_j \geq 0$, para $j \in \{1, \dots, p\}$ e $v \in \{1, \dots, q\}$ o que é devidamente comprovado quando observamos que os zeros do polinômio $\omega(L) = 1 - \sum_{j=1}^g (\alpha_j + \beta_j)L^j$ estão forá do círculo unitário onde L representa o operador lag, sendo $g = \max(p, q)$, com $\beta_j = 0$ e $\alpha_j = 0$ para $j > q$ e $j > p$, respectivamente.

A utilização desse modelo está normalmente associada à suposição de dependência entre as durações ou seja durações passadas influenciam nos valores seguintes, estacionaridade fraca e média unitária para a componente ϵ_i , sendo a mesma aproximada por uma distribuição de probabilidade

Para o desenvolvimento dos modelos, utilizaremos o pacote Tools for Autoregressive Conditional Duration Models (ACDm) (BELFRAGE, 2022). Este pacote possibilita a escolha do modelo ACD, a determinação da ordem do modelo, a seleção da distribuição do erro, entre outras especificações.

2.3 Distribuição associada ao erro

O desenvolvimento de cada modelo depende da distribuição do erro escolhida, resultando em modelos distintos com parâmetros variados. Todos os modelos que iremos

observar nesse estudo, apesar de possuírem parâmetros diferentes, utilizam o método de máxima verossimilhança, exceto o modelo com erro exponencial, que utiliza quasi maximum likelihood(QML) para a estimação dos parâmetros, uma vez que QML não produz estimadores consistentes para as outras distribuições que iremos utilizar como comentado por Bhatti (BHATTI, 2010), o método QML produz estimadores consistentes para modelos cuja função condicional de log-verossimilhança é da família das distribuições exponenciais lineares. Contudo, não é o foco deste trabalho discutir os métodos de estimação de parâmetros; para mais informações, sugerimos a leitura de (GOURIEROUX; MONFORT; TROGNON, 1984).

Para cada distribuição do erro temos suas particularidades de utilização, funções de densidade da distribuição, métodos de estimação de parâmetros e funções de sobrevivência em seguida será brevemente comentado algumas dessas características para cada uma das distribuições associadas ao erro.

A função de risco definida por

$$h(z) = \frac{f(z)}{S(z)}, \quad (2.3.1)$$

e a função de sobrevivência $S(z)$, dada por:

$$S(z) = P(\epsilon > z) = 1 - P(\epsilon \leq z). \quad (2.3.2)$$

Apesar de citarmos nas seções seguintes, a realização da grande maioria dos cálculos não é necessária por já estarem devidamente implementados no pacote (BELFRAGE, 2022).

2.3.1 Erro com distribuição Exponencial

O modelo ACD associado à distribuição exponencial(EACD), utiliza o método QML para a estimação dos parâmetros e possui a seguinte função de densidade para o ϵ :

$$f(z) = \exp(-z), \quad (2.3.3)$$

tendo a sua $E(\epsilon) = 1$ e $Var(\epsilon) = 1$, logo para o modelo EACD(1,1), assumindo fraca estacionariedade, podemos concluir que

$$E(x_i) = \frac{\omega}{1 - \alpha_1 - \beta_1}, \quad (2.3.4)$$

$$Var(x_i) = \frac{\mu_x^2(1 - \beta_1^2 - 2\alpha_1\beta_1)}{1 - 2\alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1}, \quad (2.3.5)$$

e sua função de risco definida por:

$$h(x_i|\psi_i) = \frac{1}{\psi_i}. \quad (2.3.6)$$

É interessante mencionar que a função de risco associada ao modelo EACD é constante e, devido à natureza dos nossos dados, esse modelo acaba sendo menos eficiente quando comparado a outros modelos que possuem uma função de risco mais adequada aos dados

2.3.2 Erro com distribuição Weibull

O modelo ACD associado à distribuição Weibull(WACD), foi proposto por (EN-
GLE; RUSSELL, 1998), e possui a seguinte função de densidade para o erro:

$$f(z|\gamma) = \theta\gamma^{-1}e^{-\theta z^\gamma}, \quad (2.3.7)$$

onde:

$$\theta = [\Gamma(\gamma^{-1} + 1)]^\gamma, \quad (2.3.8)$$

Aqui γ é o parâmetro de forma e $\Gamma(\cdot)$ a função Gamma completa. A média e variância de ϵ são $E(\epsilon) = 1$ e $Var(\epsilon) = \Gamma(1 + \frac{2}{\gamma})/[\Gamma(1 + \frac{1}{\gamma})]^2 - 1$

Possuindo a seguinte função de risco:

$$h(z|\gamma) = \gamma\theta z^{\gamma-1}, \quad (2.3.9)$$

Logo podemos observar que se $\gamma > 1$ a função de risco é monotona crescente enquanto se $0 < \gamma < 1$ a função de risco é monotona decrescente.

Com base na equação(2.2.2) podemos realizar uma substituição na função de densidade de ϵ (2.3.7) com o intuito de obter a função de densidade condicional de x_i dado ψ_i :

$$f(x_i|\psi_i) = \theta\gamma \frac{x_i^{\gamma-1}}{\psi_i^\gamma} e^{-\theta(\frac{x_i}{\psi_i})^\gamma} \quad (2.3.10)$$

Essa função pode ser utilizada para a obtenção da função condicional de log-verossimilhança.

2.3.3 Erro com distribuição Burr

O modelo ACD associado à distribuição Burr(BACD), é proposto com o intuito de dar mais flexibilidade para o modelo e é descrito como uma mistura gamma de distribuições weibull, possuindo a seguinte função de densidade para o erro:

$$f(z) = \frac{\theta \kappa z^{\kappa-1}}{(1 + \sigma^2 \theta z^\kappa)^{\frac{1}{\sigma^2} + 1}}, \quad (2.3.11)$$

em que:

$$\theta = \sigma^{2(1+\frac{1}{\kappa})} \frac{\Gamma(\frac{1}{\sigma^2} + 1)}{\Gamma(\frac{1}{\kappa} + 1)\Gamma(\frac{1}{\sigma^2} - \frac{1}{\kappa})}. \quad (2.3.12)$$

Como os parâmetros da distribuição Burr são σ^2 , κ , para o desenvolvimento do modelo adotamos $0 < \sigma^2 < \kappa$ sendo $\kappa \geq 1$, $\sigma^2 > 0$ e $E(\epsilon) = 1$. Possuindo a seguinte função de risco

$$h(z|\kappa, \sigma) = \frac{\theta \kappa z_i^{\kappa-1}}{1 + \sigma^2 \theta z_i^\kappa}, \quad (2.3.13)$$

observando a função de risco, podemos notar que ela é não monotômica algo que é de nosso interesse pois pode se adequar melhor a natureza dos dados.

O modelo BACD possui casos especiais de modo que quando $\sigma^2 \rightarrow 0$ o modelo se torna um WACD, seguindo a mesma restrição com $\kappa = 1$ o modelo se torna um EACD como comentado por (GRAMMIG; MAURER, 2000).

2.3.4 Erro com distribuição Gamma-Generalizada

O modelo ACD associado à distribuição Gamma-Generalizada(GACD), foi proposto por (ZHANG; RUSSELL; TSAY, 2001), e possui a seguinte função de densidade para o erro tomando $E(\epsilon_i) = 1$:

$$f(z) = \frac{\gamma z^{\kappa\gamma-1}}{\lambda^{\kappa\gamma}\Gamma(\kappa)} \exp[-(\frac{z}{\lambda})^\gamma], z > 0 \quad (2.3.14)$$

em que:

$$\lambda = \frac{\Gamma(\kappa)}{\Gamma(\kappa + \frac{1}{\gamma})}. \quad (2.3.15)$$

Sendo os $\gamma > 0$ e $\kappa > 0$ parâmetros de forma da distribuição, possuindo mais flexibilidade em sua função de risco. A função de densidade condicional de x_i para o

modelo (GACD) segue:

$$f(x_i|\gamma, \kappa) = \frac{\gamma x_i^{\gamma\kappa-1}}{(\psi_i\lambda)^{\gamma\kappa}\Gamma(\kappa)} \exp[-(\frac{x_i}{\psi_i\lambda})]. \quad (2.3.16)$$

A partir dessa função podemos obter a função condicional de máxima verossimilhança para o modelo.

2.4 Adequação dos Modelos

Sendo $\hat{\epsilon}_i$, $\hat{\psi}_i$, o erro e a duração condicional obtidos na estimação do modelo. Utilizando $\hat{\epsilon}_i = x_i / \hat{\psi}_i$, podemos verificar a adequação dos modelos, uma vez que o $\hat{\epsilon}_i$ deve ser i.i.d, seguindo a distribuição assumida. As séries $\hat{\epsilon}_i$ e $\hat{\epsilon}_i^2$ não devem possuir correlação. Para a verificação dos critérios, podemos utilizar o teste Ljung-Box para a correlação. Além disso, podemos observar o ajuste a distribuição pelo QQ-plot, onde quanto mais semelhante à distribuição assumida, melhor. Como comentado por (TSAY, 2009), neste trabalho nos iremos utilizar o QQ-plot com distribuição exponencial, para os casos em que o erro não seja exponencial será utilizado o resíduo de Cox-Snell

$$-\ln(S(x_i)). \quad (2.4.1)$$

Aqui, $S(x_i)$ a função estimada de sobrevivência do modelo, nos levando para uma distribuição exponencial unitária, de maneira que possamos realizar uma melhor observação de adequação e posteriormente uma comparação entre os modelos

2.5 Método Bagging

Um dos objetivos deste trabalho é avaliar se o método Bootstrap Aggregating (bagging) proporciona resultados mais precisos. Este método envolve reamostrar a série temporal sem tendência e sazonalidade utilizando o bootstrap e, para cada uma das séries reamostradas, realizar previsões. Em seguida, calcula-se a média dessas previsões para obter uma previsão final. Para uma melhor visualização do processo podemos observar a Figura 1, é válido resaltar que para esse trabalho a decomposição da série será feita utilizando o função "diurnalAdj" divergindo do proposto na imagem.

Ressaltamos que existem diversos métodos de reamostragem bootstrap. Baseado no estudo realizado em (CAMARGO, 2021), utilizaremos o Moving Block Bootstrap (MBB) por ser o mais adequado para séries temporais. O MBB consiste na formação de

blocos consecutivos de tamanho "n", dos quais serão selecionados "k" blocos para cada nova amostra, que serão utilizadas para realizar as previsões. No estudo em questão iremos variar os valores dos blocos para observar em qual tamanho teremos as previsões mais precisas, os tamanhos utilizados para os blocos serão 5, 10, 20, 30, 40 e 50.

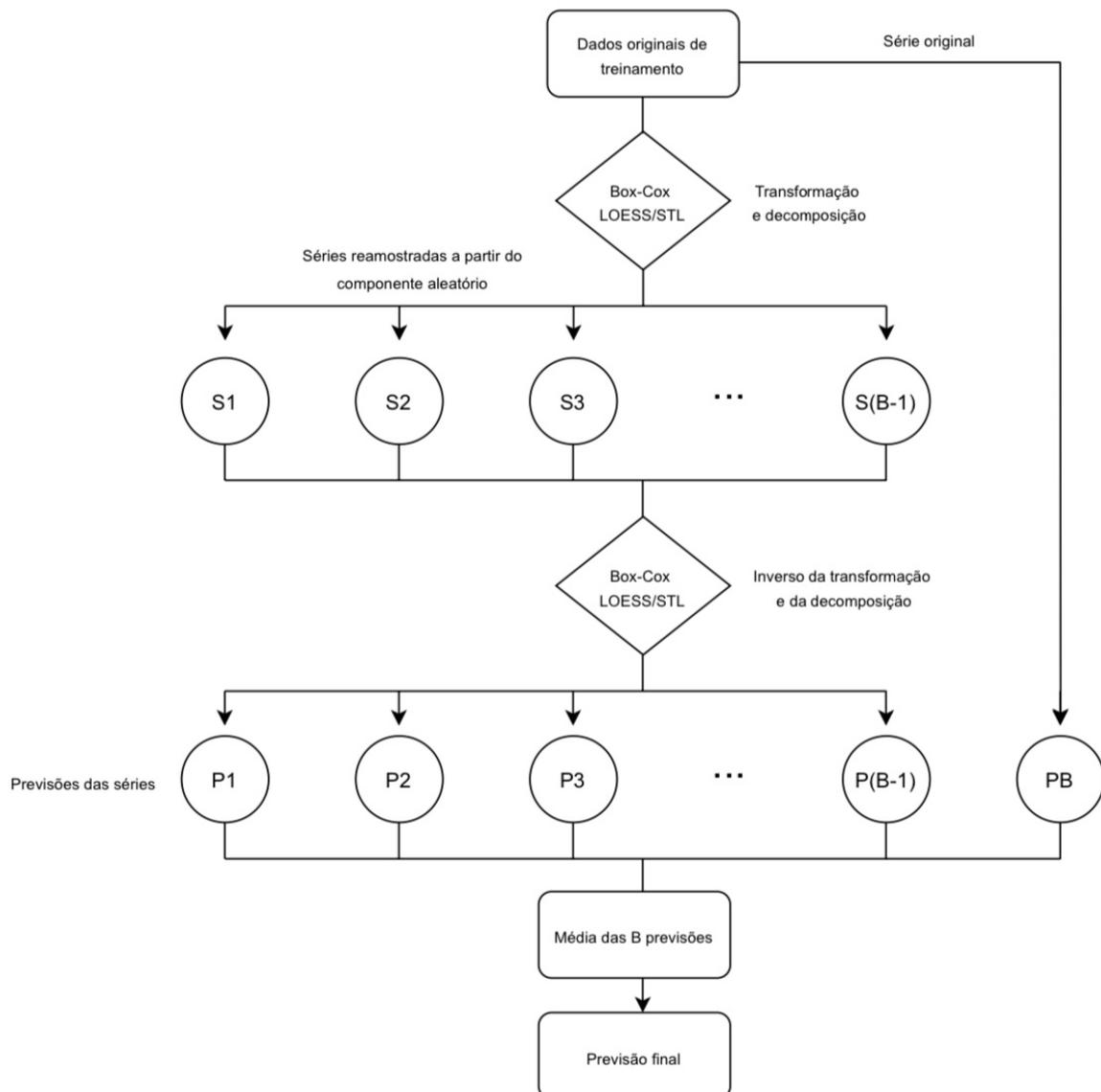


Figura 1: Ilustração da técnica Bagging. Fonte: (CAMARGO, 2021)

2.6 Previsão

As previsões serão feitas de maneira independente para cada banco de dados, sendo que para banco iremos observar as previsões para 50 passos a frente, para cada modelo elaborado.

Para a realização das previsões referentes a cada modelo, será definido como $i = h$ o momento de origem para início das previsões, seguindo a estrutura dos modelos ACD(1,1) utilizados teríamos $x_{h+1} = \psi_{h+1}\epsilon_{h+1}$, sendo $\psi_{h+1} = \omega + \alpha_1 x_h + \beta_1 \psi_h$, logo iremos definir $x_h(1)$ como sendo a primeira previsão ou seja a previsão de um passo

$$x_h(1) = E(x_{h+1}) = E(\psi_{h+1}\epsilon_{h+1}) = \psi_{h+1}. \quad (2.6.1)$$

Definindo o erro associado a está previsão como

$$e_h(1) = x_{h+1} - x_h(1) = \psi_{h+1}(\epsilon_{h+1} - 1). \quad (2.6.2)$$

De maneira geral podemos definir uma fórmula como sendo a fórmula geral para previsões

$$x_h(m) = \omega + (\alpha_1 + \beta_1)x_h(m-1). \quad (2.6.3)$$

2.7 Comparação dos Modelos

Para realizar a comparação entre os modelos iremos observar o modelo que melhor se ajusta aos dados e sua capacidade preditiva ou seja o modelo que nos fornece melhores previsões.

A comparação dos modelos quanto ao seu ajuste aos dados será feita pela observação do QQ-plot referente a cada modelo desenvolvido, também iremos produzir uma tabela com a estimação de cada parâmetro com seu erro-padrão, e p-valores referentes aos teste de correlação, também será observado o AIC e BIC de cada modelo sendo eles bons medidores de ajustes como feito em (SAULO; BALAKRISHNAN; VILA, 2023)

Sendo AIC e BIC definidos respectivamente por:

$$AIC = -2 \ln(L) + 2k, \quad (2.7.1)$$

$$BIC = -2 \ln(L) + k \ln(n). \quad (2.7.2)$$

Uma vez que L denota o máximo da função de verossimilhança do modelo em questão, k o número de parâmetros estimados do modelo e n o número de observações dos dados.

Para a comparação quanto a capacidade preditiva dos modelos nos iremos utilizar os dados separados para treinamento e comparalos aos obtidos nas previsões dos modelos utilizando o erro médio(ME), a raiz do erro quadrático médio(RMSE) e o erro percentual absoluto médio(MAPE) como comentado em (CAMARGO, 2021).

Sendo ME, RMSE e MAPE definidos respectivamente por:

$$ME = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i), \quad (2.7.3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)^2, \quad (2.7.4)$$

$$RMSE = \sqrt{MSE}, \quad (2.7.5)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left(\frac{|X_i - \hat{X}_i|}{X_i} \right), \quad (2.7.6)$$

sendo X_i o real valor observado e \hat{X}_i o valor previsto.

3 Resultados

3.1 Apple

3.1.1 Durações

Após a entrada dos dados e a aplicação da função, podemos visualizar as durações em segundos entre a mudança de preço acima de 0,01 dolares obtidas referentes a empresa Apple, figura 2, em seguida observa-se as mesmas durações retirando sua sazonalidade, figura 3.

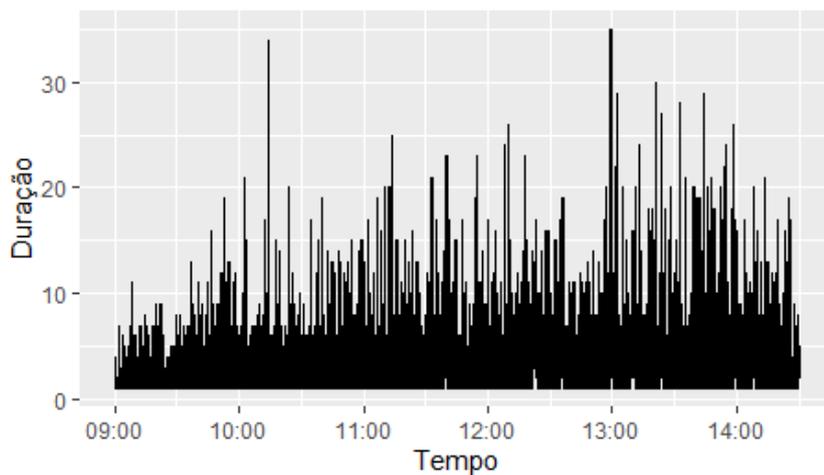


Figura 2: Durações observadas, Apple

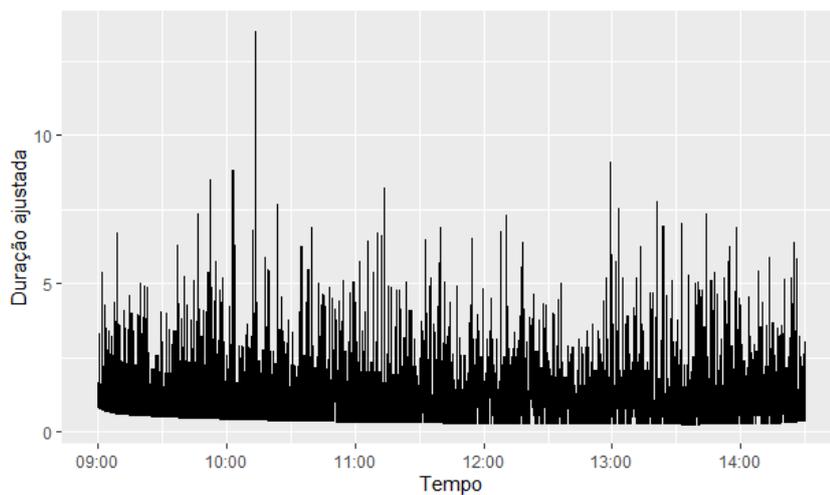


Figura 3: Durações sem sazonalidade, Apple

3.1.2 Modelos e Adequações

Utilizando as durações sem sazonalidade obtidas podemos partir para a estimação e verificação dos modelos. Os modelos para cada distribuição associada ao erro são estimados utilizando a função `acd.fit` e quando testados para correlação dos erros, gerão os seguintes resultados para os p-valores do teste Ljung-Box, bem como o AIC e BIC, na tabela 1 temos a distribuição atribuída ao erro e os resultados obtidos, para a interpretação dos resultados utilizaremos um nível de confiança de 0,95:

Tabela 1: Distribuições do erro e resultados do teste Ljung-Box associados

Erro	Lag=4	Lag=16	AIC	BIC
Exponencial	0.02567	0.01151	13494.62	13515.02
Weibull	0.05208	0.0267	12951.11	12978.31
Burr	0.001983	2.356e-07	10968.48	11002.48
Gamma-Gen	0.02445	0.003586	11418.31	11452.31

Os modelos gerados para o banco de dados da empresa Apple apresentaram grandes indícios de correlação no teste Ljung-Box e valores mais elevados de AIC e BIC. Apenas em um dos casos nos obtivemos um p-valor acima de 0,05 indicando que devemos rejeitar a hipótese nula de independência, algo que difere do desejado, contudo, a título de estudo, daremos continuidade.

Nas figuras 4, 5, 6 e 7, podemos ver os QQ-plots referentes aos erros de cada modelo sendo a linha de referência o equivalente ao QQ-plot de uma distribuição exponencial unitária e os modelos EACD e GGACD mostram um ajuste adequado.

3.1.3 Bagging

A aplicação do método bagging consiste em reamostrar a série de durações originais, estimar modelos e a partir destes modelos gerar previsões, tais serão combinadas por meio de uma média para cada passo das previsões. Neste trabalho a reamostragem será feita utilizando Moving Block Bootstrap com blocos de tamanho 5, 10, 20, 30, 40 e 50, sendo assim geradas 50 novas séries temporais para cada tamanho de bloco, que utilizaremos para gerar os modelos responsáveis pelas previsões. Com o intuito de facilitar a visualização do processo de reamostragem será gerado um gráfico, figura 9, com as 50 primeiras observações das 50 séries geradas pela reamostragem referentes aos blocos de tamanho 50, sendo a série original representada pela cor preta.

Ao observar as amostras geradas pelo bagging, podemos notar que sua sazonalidade

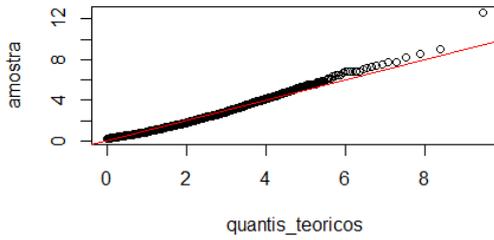


Figura 4: QQ-plot exponencial

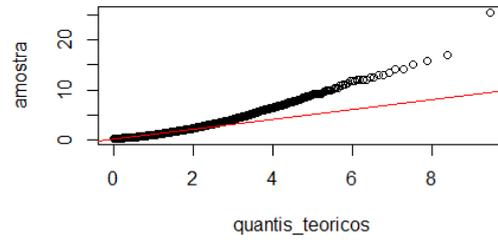


Figura 5: QQ-plot weibull

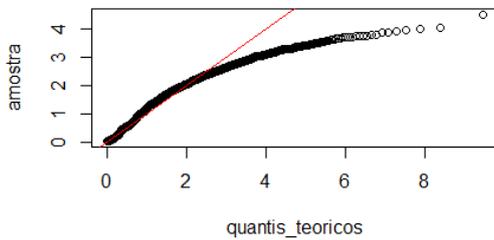


Figura 6: QQ-plot burr

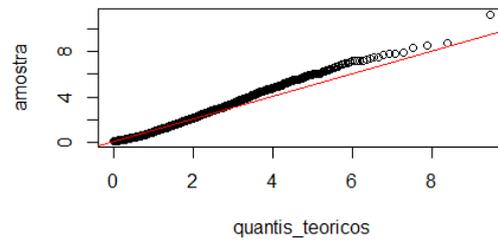


Figura 7: QQ-plot gamma-generalizada

Figura 8: Conjunto de QQ-plots para diferentes distribuições

dade não foi bem removida, produzindo séries que divergem bastante da original.

3.1.4 Previsões

Para concluir o método bagging, após a reamostragem das séries é elaborado um modelo para cada nova série gerada, utilizando os novos modelos realizaremos previsões 50 passos a frente, tais previsões serão combinadas utilizando uma média entre elas para cada tamanho de bloco, para cada distribuição associada ao erro. Os modelos gerados pela série original também irão produzir previsões com o objetivo de uma comparação entre a predição dos modelos com e sem a utilização do método bagging.

A comparação entre os modelos será realizada observando os valores de MSE, RMSE, MAAPE associados a cada previsão gerada, quando comparadas aos dados observados que foram deixados fora da amostra. Os valores comentados serão apresentados por meio de uma tabela para cada distribuição, contendo os resultados para cada tamanho de bloco e para a distribuição sem bagging, tabelas 2, 3, 4 e 5.

Para uma ilustração das previsões será apresentado um gráfico contendo os dados originais deixados fora da amostra para teste e as previsões realizadas utilizando o modelo

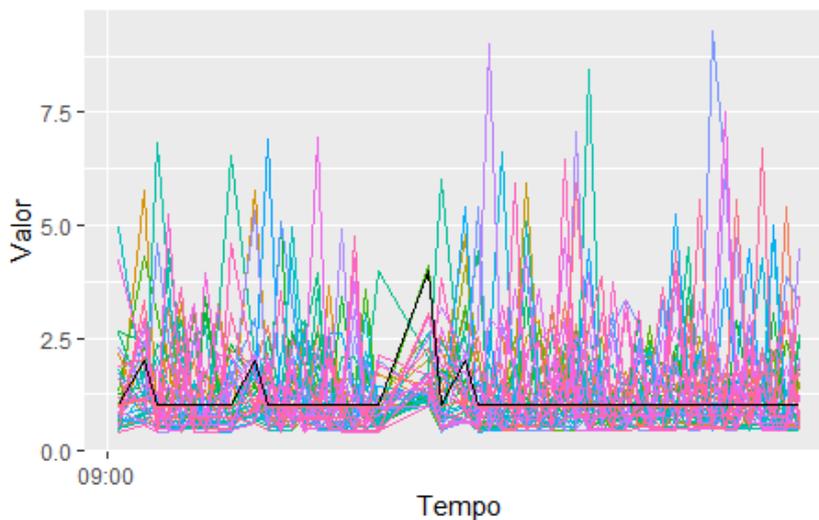


Figura 9: Durações reamostradas com blocos de tamanho 50, Apple

sem bagging referente a cada distribuição, figuras 10, 11, 12 e 13

Tabela 2: Estatísticas das previsões, modelos EACD, Apple.

Estatísticas	Tamanho dos blocos						
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	
ME	0.02494	0.02028	0.01940	0.02465	0.02218	0.02471	0.02004
RMSE	0.70279	0.70334	0.70285	0.70309	0.70268	0.70268	0.70220
MAPE	29.95765	30.15685	30.16924	29.97855	30.05603	29.95908	30.11716



Figura 10: Previsão do Modelo EACD comparada com dados reais, Apple.

Apesar de os dados anteriores divergirem bastante do esperado, seguimos para a avaliação de sua capacidade preditiva. No modelo exponencial, obtivemos valores baixos para ME e RMSE, indicando um bom ajuste, e valores aceitáveis para o MAPE. O modelo desenvolvido sem bagging obteve o maior ME, com uma diferença pequena em relação aos

Tabela 3: Estatísticas das previsões, modelos WACD, Apple.

Estatísticas	Tamanho dos blocos						
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	
ME	0.01495	0.00905	0.00826	0.01366	0.01139	0.01330	0.00856
RMSE	0.70123	0.70289	0.70208	0.70189	0.70143	0.70126	0.70086
MAPE	30.27562	30.56224	30.55709	30.34536	30.41262	30.33329	30.49377

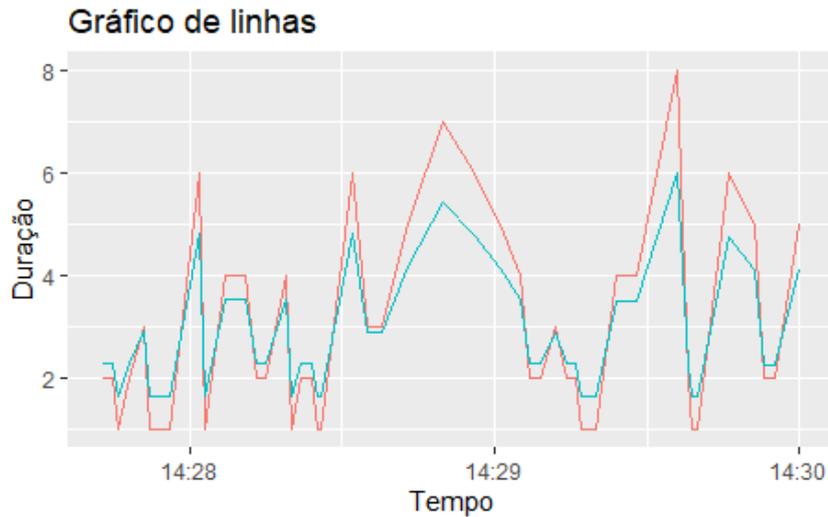


Figura 11: Previsão do Modelo WACD comparada com dados reais, Apple.

Tabela 4: Estatísticas das previsões, modelos BACD, Apple.

Estatísticas	Tamanho dos blocos						
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	
ME	-0.69878	-0.93738	-0.87700	-1.29804	-1.97484	-1.61334	-1.25915
RMSE	0.97208	1.16033	1.11214	1.46447	2.09268	1.75277	1.43132
MAPE	58.70199	72.5316	69.3872	91.10371	127.3676	107.1828	88.3614

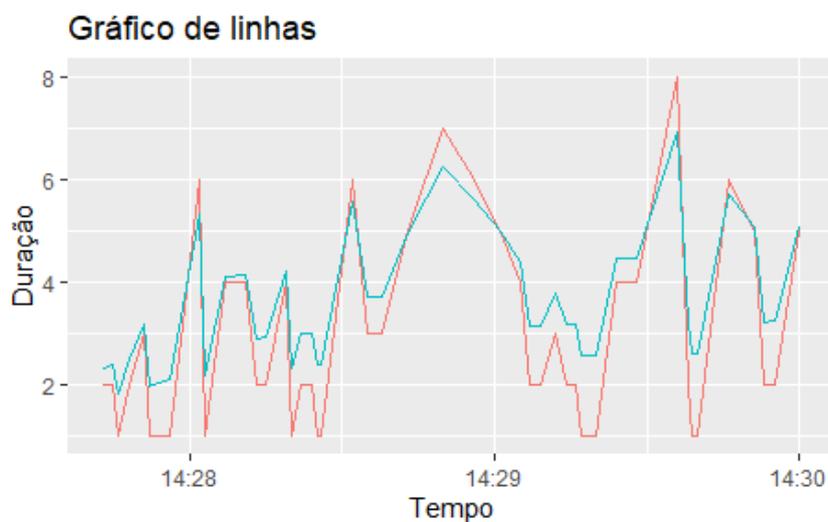


Figura 12: Previsão do Modelo BACD comparada com dados reais, Apple.

demais modelos, não indicando qualquer melhor desempenho dos modelos com bagging.

Tabela 5: Estatísticas das previsões, modelos GGACD, Apple.

Estatísticas	Tamanho dos blocos					
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
ME	0.04341	0.03451	0.04025	0.03921	0.04218	0.04481
RMSE	0.70616	0.70507	0.70560	0.70586	0.70586	0.70627
MAPE	29.44829	29.71016	29.53029	29.56494	29.4265	29.36128



Figura 13: Previsão do Modelo GGACD comparada com dados reais, Apple.

Nos modelos ajustados com erro Weibull, também são observados valores pequenos para o ME, sendo o modelo sem bagging o que apresenta o maior ME. Os valores de RMSE também indicam uma boa predição, e não se observa uma diferença significativa entre os valores obtidos para os modelos com e sem a presença do método bagging.

Os modelos ajustados com erro Burr apresentaram valores negativos para o ME, indicando que as previsões estão levemente inflacionadas, enquanto o MAPE apresenta valores elevados. Vale ressaltar que, nesses modelos, os menores valores para o ME e RMSE são obtidos pelo modelo sem bagging.

Nas previsões obtidas pelo modelo GGACD, observamos que tanto o ME quanto o RMSE possuem valores baixos, indicando uma boa predição, enquanto os valores para o MAPE se mantiveram no mesmo nível dos obtidos com os modelos EACD e WACD.

3.2 Bayer

3.2.1 Durações

Após a entrada dos dados e a aplicação da função, podemos visualizar as durações em segundos entre a mudança de preço acima de 0,01 dolares obtidas referentes a empresa Bayer, figura 14, em seguida observa-se as mesmas durações retirando sua sazonalidade, figura 15.

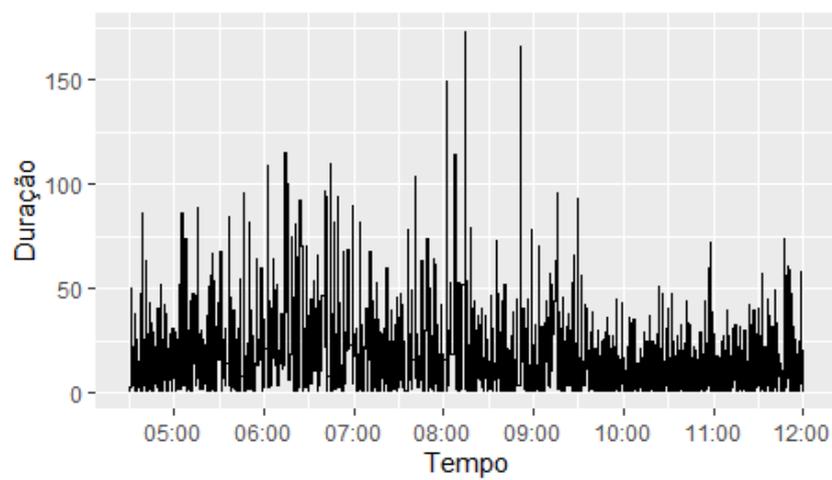


Figura 14: Durações observadas, Bayer

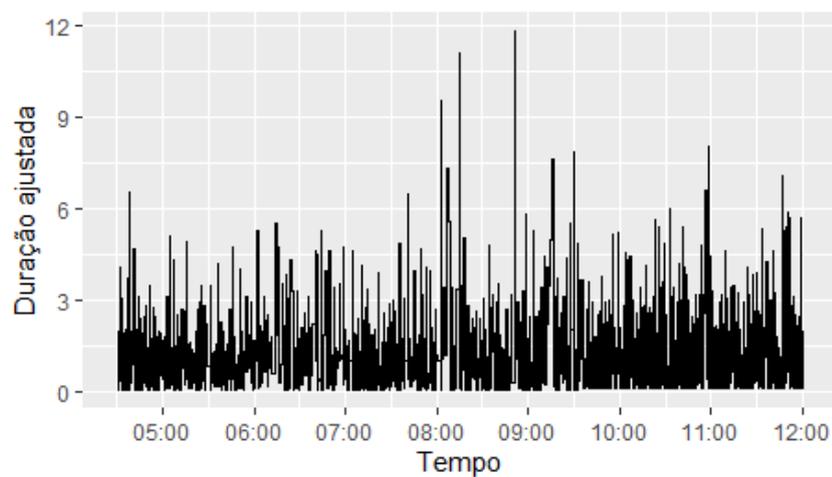


Figura 15: Durações sem sazonalidade, Bayer

3.2.2 Modelos e Adequações

Utilizando as durações sem sazonalidade obtidas podemos partir para a estimação e verificação dos modelos. Os modelos para cada distribuição associada ao erro são estimados utilizando a função `acd.fit` e quando testados para correlação dos erros, geram os seguintes resultados para os p-valores do teste Ljung-Box, bem como o AIC e BIC, na tabela 6 temos a distribuição atribuída ao erro e os resultados obtidos, para a interpretação dos resultados utilizaremos um nível de confiança de 0,95:

Tabela 6: Distribuições do erro e resultados do teste Ljung-Box associados

Erro	Lag=4	Lag=16	AIC	BIC
Exponencial	0.683	0.5897	3989.76	4006.48
Weibull	0.6498	0.5204	3967.22	3989.50
Burr	0.5651	0.3557	3942.81	3970.66
Gamma-Gen	0.1506	0.1831	4017.74	4045.72

Em todos os casos foram obtidos p-valores acima de 0,05 para o teste Ljung-Box logo não possuímos evidências para reitarmos a independência do erro, observando os valores do AIC e BIC podemos observar que o modelo apresenta resultados positivos.

Nas figuras 16, 17, 18 e 19, podemos ver os QQ-plots referentes aos erros de cada modelo sendo a linha de referência o equivalente ao QQ-plot de uma distribuição exponencial unitária, todos os modelos apresentam um ajuste adequado, principalmente o modelo WACD.

3.2.3 Bagging

A aplicação do método bagging consiste em reamostrar a série de durações originais, estimar modelos e apartir destes modelos gerar previsões, tais serão combinadas por meio de uma média para cada passo das previsões. Neste trabalho a reamostragem será feita utilizando Moving Block Bootstrap com blocos de tamanho 5, 10, 20, 30, 40 e 50, sendo assim geradas 50 novas séries temporais para cada tamanho de bloco, que utilizaremos para gerar os modelos responsáveis pelas previsões. Com o intuito de facilitar a visualização do processo de reamostragem será gerado um gráfico, figura 21, com as 50 primeiras observações das 50 séries geradas pela reamostragem referentes aos blocos de tamanho 50, sendo a série original representada pela cor preta.

A visualização do processo de reamostragem é bem interessante, pois, ao contrário do observado com os dados da Apple, obtivemos séries que se aproximam da série original,

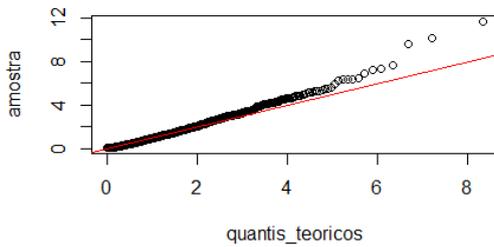


Figura 16: QQ-plot exponencial

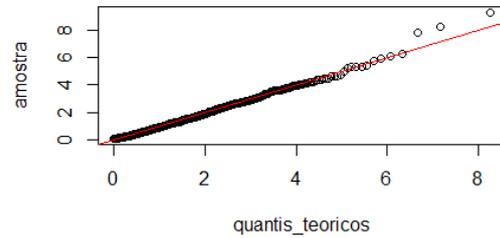


Figura 17: QQ-plot weibull

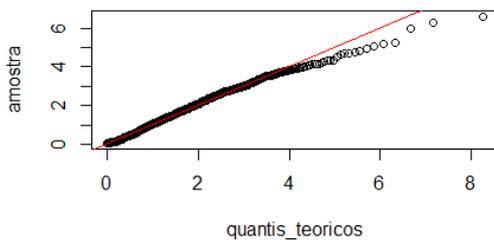


Figura 18: QQ-plot burr

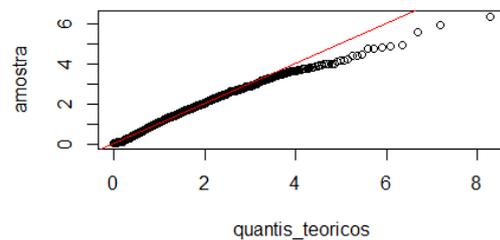


Figura 19: QQ-plot gamma-generalizada

Figura 20: Conjunto de QQ-plots para diferentes distribuições

o que nos leva a crer que a sazonalidade dos dados foi retirada com mais precisão, podendo resultar em previsões mais coerentes.

3.2.4 Previsões

Para concluir o método bagging, após a reamostragem das séries é elaborado um modelo para cada nova série gerada, utilizando os novos modelos realizaremos previsões 50 passos a frente, tais previsões serão combinadas utilizando uma média entre elas para cada tamanho de bloco, para cada distribuição associada ao erro. Os modelos gerados pela série original também irão produzir previsões com o objetivo de uma comparação entre a predição dos modelos com e sem a utilização do método bagging.

A comparação entre os modelos será realizada observando os valores de MSE, RMSE, MAAPE associados a cada previsão gerada, quando comparadas aos dados observados que foram deixados fora da amostra. Os valores comentados serão apresentados por meio de uma tabela para cada distribuição, contendo o resultados para cada tamanho de bloco e para a distribuição sem bagging, tabelas 7, 8, 9 e 10.

Para uma ilustração das previsões será apresentado um gráfico contendo os dados

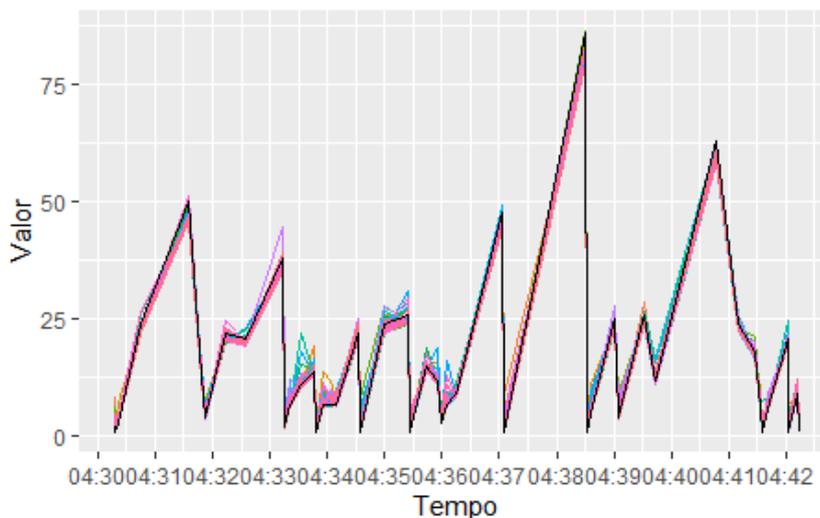


Figura 21: Durações reamostradas com blocos de tamanho 50, Bayer

originais deixados fora da amostra para teste e as previsões realizadas utilizando o modelo sem bagging referente a cada distribuição, figuras 22, 23, 24 e 25

Tabela 7: Estatísticas das previsões, modelos EACD, Bayer.

Estatísticas	Tamanho dos blocos						
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	
ME	0.451380	0.443800	0.447218	0.447643	0.445908	0.453202	0.444533
RMSE	1.722859	1.719792	1.721125	1.720351	1.721569	1.722721	1.720478
MAPE	27.36621	27.55805	27.47016	27.45563	27.50523	27.30930	27.54110

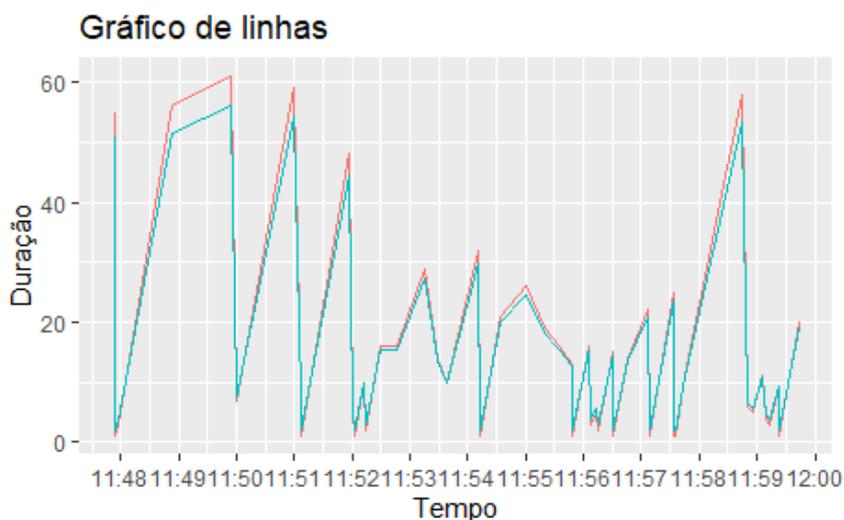


Figura 22: Previsão do Modelo EACD comparada com dados reais, Bayer.

Ao observar os resultados das predições e suas estatísticas, podemos perceber resultados extremamente homogêneos, de maneira que não notamos qualquer diferença entre os modelos, tamanho de bloco ou presença de bagging. Em todos os modelos

Tabela 8: Estatísticas das previsões, modelos WACD, Bayer.

Estatísticas	Tamanho dos blocos					
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
ME	0.45391	0.44701	0.45033	0.45064	0.44943	0.45627
RMSE	1.72317	1.72044	1.72168	1.72097	1.72217	1.72326
MAPE	27.29649	27.47274	27.38636	27.37571	27.41048	27.22667

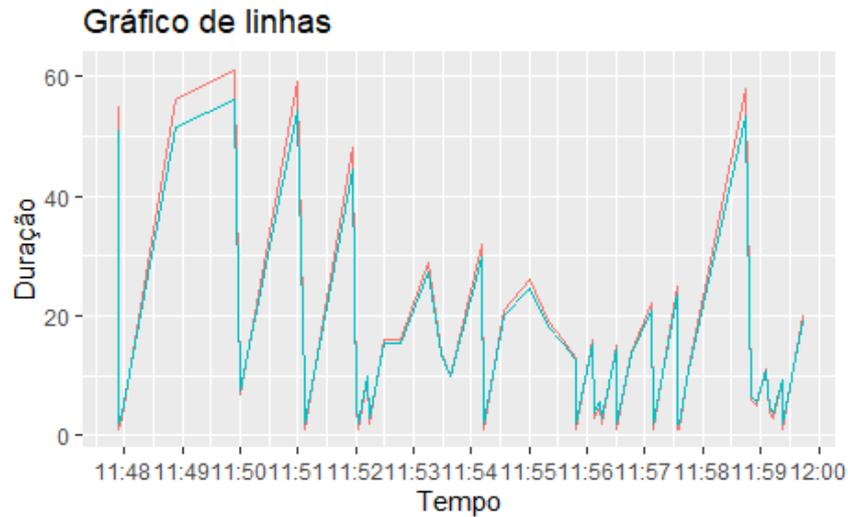


Figura 23: Previsão do Modelo WACD comparada com dados reais, Bayer.

Tabela 9: Estatísticas das previsões, modelos BACD, Bayer.

Estatísticas	Tamanho dos blocos					
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
ME	0.43774	0.43481	0.43597	0.43848	0.43269	0.44186
RMSE	1.72021	1.71860	1.71881	1.71856	1.71898	1.72030
MAPE	27.72889	27.80796	27.77392	27.70430	27.86056	27.93417



Figura 24: Previsão do Modelo BACD comparada com dados reais, Bayer.

elaborados, é observada uma boa capacidade preditiva, gerando baixos valores para ME,

Tabela 10: Estatísticas das previsões, modelos GGACD, Bayer.

Estatísticas	Tamanho dos blocos						
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	
ME	0.41667	0.41800	0.41795	0.42646	0.42622	0.43159	0.42272
RMSE	1.71140	1.71239	1.71272	1.71345	1.71588	1.71561	1.71405
MAPE	28.16487	28.23069	28.22529	28.00069	28.02077	27.86831	28.11525

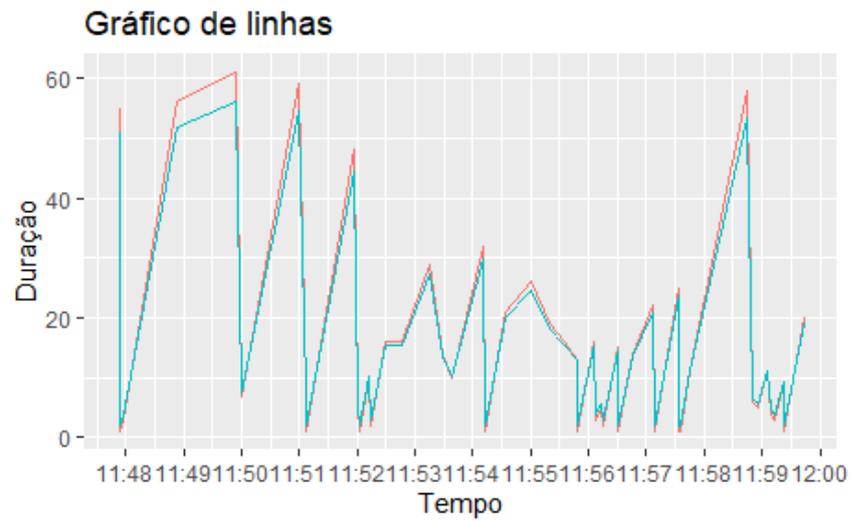


Figura 25: Previsão do Modelo GGACD comparada com dados reais, Bayer.

RMSE e MAPE.

3.3 BASF SE

3.3.1 Durações

Após a entrada dos dados e a aplicação da função, podemos visualizar as durações em segundos entre a mudança de preço acima de 0,01 dolares obtidas referentes a empresa BASF SE, figura 26, em seguida observa-se as mesmas durações retirando sua sazonalidade, figura 27.

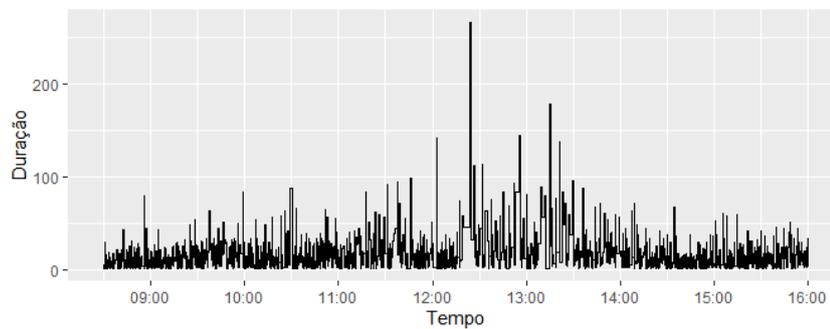


Figura 26: Durações observadas, BASF SE

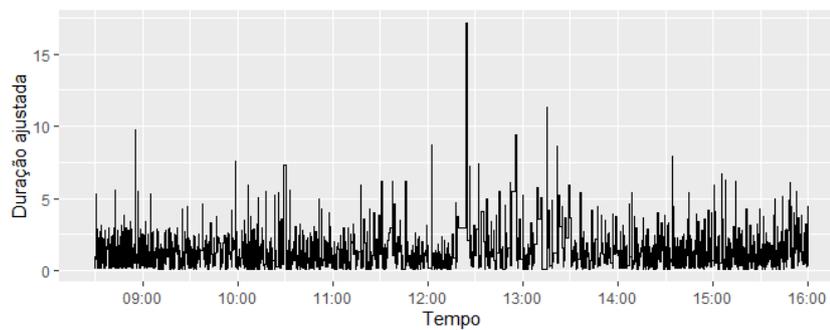


Figura 27: Durações sem sazonalidade, BASF SE

3.3.2 Modelos e Adequações

Utilizando as durações sem sazonalidade obtidas podemos partir para a estimação e verificação dos modelos. Os modelos para cada distribuição associada ao erro são estimados utilizando a função `acd.fit` e quando testados para correlação dos erros, gerão os seguintes resultados para os p-valores do teste Ljung-Box, bem como o AIC e BIC, na tabela 11 temos a distribuição atribuída ao erro e os resultados obtidos, para a interpretação dos resultados utilizaremos um nível de confiança de 0,95:

Tabela 11: Distribuições do erro e resultados do teste Ljung-Box associados

Erro	Lag=4	Lag=16	AIC	BIC
Exponencial	0.6475	0.01725	4546.18	4563.20
Weibull	0.639	0.01625	4547.83	4570.51
Burr	0.9313	0.1866	4457.48	4485.84
Gamma-Gen	0.7572	0.1062	4393.27	4421.63

Apenas em um dos casos nos obtivemos um p-valor abaixo de 0,05, logo não possuímos evidências para rejeitar a independência dos erros, observando os valores do AIC e BIC podemos concluir que os modelos apresentam resultados positivos.

Nas figuras 28, 29, 30 e 31, podemos ver os QQ-plots referentes aos erros de cada modelo sendo a linha de referência o equivalente ao QQ-plot de uma distribuição exponencial unitária, os resultados mostram que não houve um ajuste muito bom.

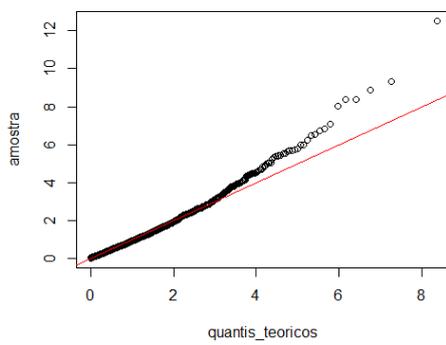


Figura 28: QQ-plot exponencial

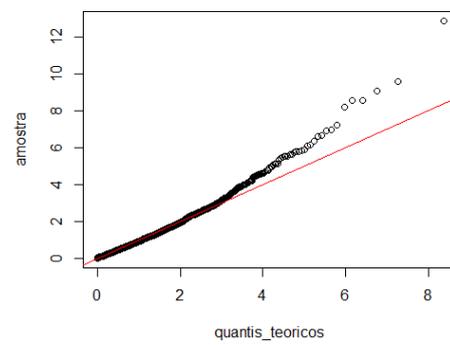


Figura 29: QQ-plot weibull

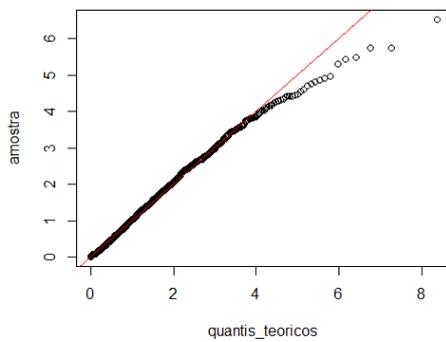


Figura 30: QQ-plot burr

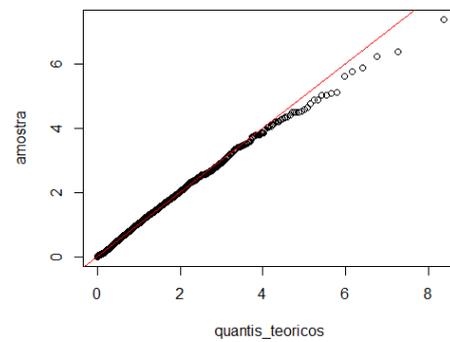


Figura 31: QQ-plot gamma-generalizada

Figura 32: Conjunto de QQ-plots para diferentes distribuições

3.3.3 Bagging

A aplicação do método bagging consiste em reamostrar a série de durações originais, estimar modelos e apartir destes modelos gerar previsões, tais serão combinadas por meio de uma média para cada passo das previsões. Neste trabalho a reamostragem será feita utilizando Moving Block Bootstrap com blocos de tamanho 5, 10, 20, 30, 40 e 50, sendo assim geradas 50 novas séries temporais para cada tamanho de bloco, que utilizaremos para gerar os modelos responsáveis pelas previsões. Com o intuito de facilitar a visualização do processo de reamostragem será gerado um gráfico, figura 33, com as 50 primeiras observações das 50 séries geradas pela reamostragem referentes aos blocos de tamanho 50, sendo a série original representada pela cor preta.

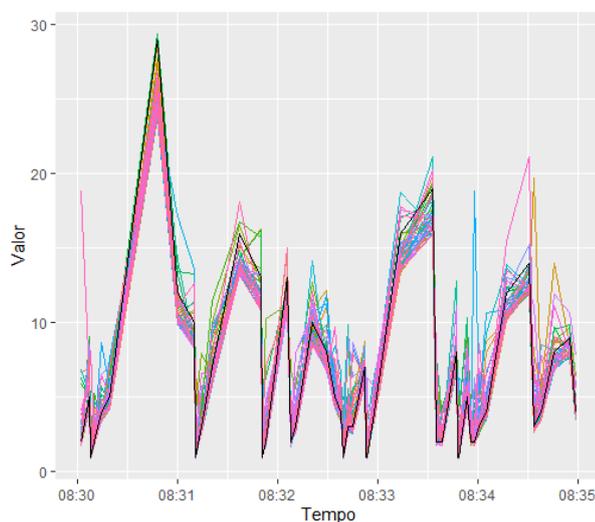


Figura 33: Durações reamostradas com blocos de tamanho 50, BASF SE

O gráfico de séries geradas pela reamostragem segue a série original, o que nos leva a acreditar em uma boa captação da sazonalidade.

3.3.4 Previsões

Para concluir o método bagging, após a reamostragem das séries é elaborado um modelo para cada nova série gerada, utilizando os novos modelos realizaremos previsões 50 passos a frente, tais previsões serão combinadas utilizando uma média entre elas para cada tamanho de bloco, para cada distribuição associada ao erro. Os modelos gerados pela série original também irão produzir previsões com o objetivo de uma comparação entre a predição dos modelos com e sem a utilização do método bagging.

A comparação entre os modelos será realizada observando os valores de MSE,

RMSE, MAAPE associados a cada previsão gerada, quando comparadas aos dados observados que foram deixados fora da amostra. Os valores comentados serão apresentados por meio de uma tabela para cada distribuição, contendo os resultados para cada tamanho de bloco e para a distribuição sem bagging, tabelas 12, 13, 14 e 15.

Para uma ilustração das previsões será apresentado um gráfico contendo os dados originais deixados fora da amostra para teste e as previsões realizadas utilizando o modelo sem bagging referente a cada distribuição, figuras 34, 35, 36 e 37.

Tabela 12: Estatísticas das previsões, modelos EACD, BASF SE.

Estatísticas	Tamanho dos blocos					
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
ME	0.41535	0.41548	0.41872	0.41711	0.42208	0.423903
RMSE	1.27763	1.27836	1.27939	1.27841	1.28002	1.28334
MAPE	11.13567	11.15567	11.12698	11.12057	11.07753	11.06872

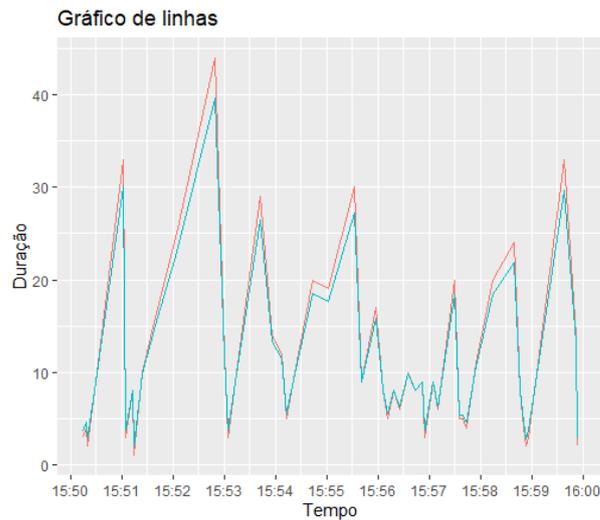


Figura 34: Previsão do Modelo EACD comparada com dados reais, BASF SE.

Tabela 13: Estatísticas das previsões, modelos WACD, BASF SE.

Estatísticas	Tamanho dos blocos					
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
ME	0.41477	0.41502	0.41824	0.41647	0.42136	0.43045
RMSE	1.27743	1.27819	1.27922	1.27819	1.27978	1.28319
MAPE	11.14076	11.15957	11.13144	11.12597	11.08401	11.00413

Tabela 14: Estatísticas das previsões, modelos BACD, BASF SE.

Estatísticas	Tamanho dos blocos					
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
ME	0.38932	0.39123	0.39397	0.39102	0.39632	0.40335
RMSE	1.26885	1.27040	1.27098	1.26967	1.27131	1.27391
MAPE	11.36950	11.37854	11.34754	11.35378	11.31030	11.24872

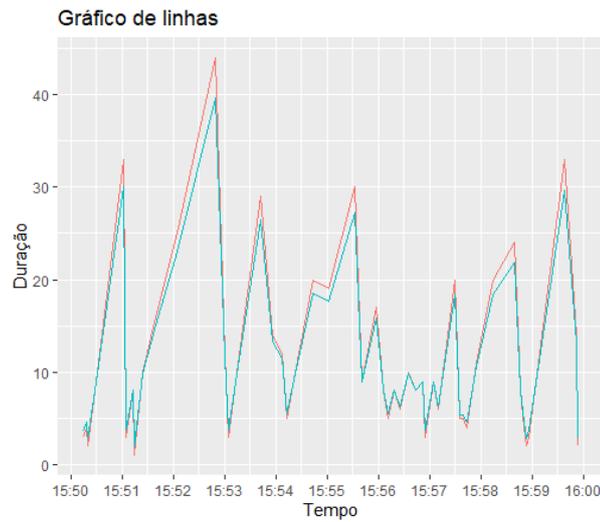


Figura 35: Previsão do Modelo WACD comparada com dados reais, BASF SE.

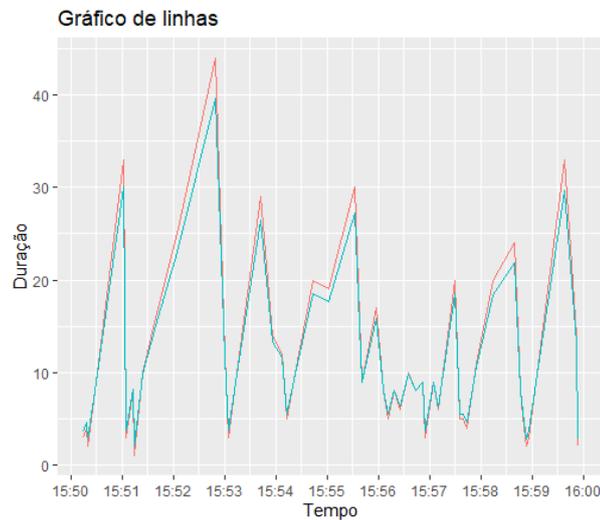


Figura 36: Previsão do Modelo BACD comparada com dados reais, BASF SE.

Tabela 15: Estatísticas das previsões, modelos GGACD, BASF SE.

Estatísticas	Tamanho dos blocos						
	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	
ME	0.41881	0.40310	0.40219	0.40312	0.40733	0.41635	0.40856
RMSE	1.27890	1.27434	1.27389	1.27373	1.27497	1.27827	1.27586
MAPE	11.11040	11.27229	11.27702	11.25004	11.21175	11.13196	11.21105

Analisando os resultados referentes à sua capacidade preditiva, todos os modelos apresentam baixos valores para ME, RMSE e MAPE, sendo que podemos observar os menores valores de MAPE para todos os modelos em comparação com os outros bancos de dados, o que pode evidenciar que a melhor capacidade preditiva tenha sido obtida neste banco de dados.

Ao comparar os resultados dos modelos com e sem a presença do método bagging,

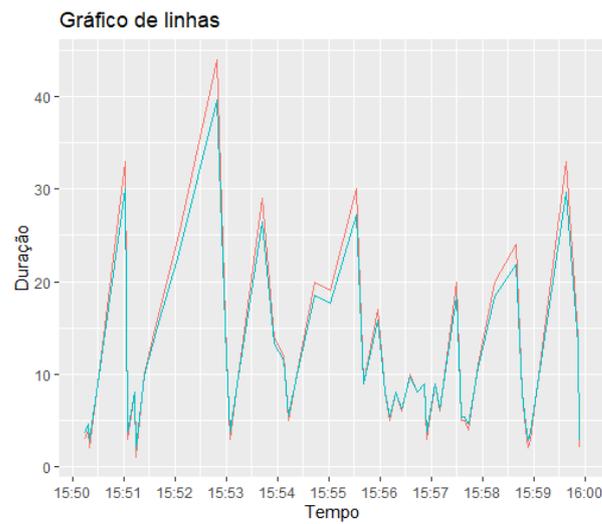


Figura 37: Previsão do Modelo GGACD comparada com dados reais, BASF SE.

não conseguimos chegar a uma conclusão, pois todos os modelos apresentam resultados muito semelhantes. A diferença observada neste banco de dados é que os modelos BACD obtiveram menores valores de ME quando comparados aos modelos que seguem distribuições diferentes para o erro. Mesmo assim, a diferença entre os ME obtidos é bem pequena, o que nos leva a acreditar em uma homogeneidade entre os resultados obtidos pelos modelos, não observando diferenças entre modelos, tamanhos de bloco ou presença e ausência de bagging.

4 Conclusão

O objetivo deste trabalho foi realizar uma comparação entre os modelos com e sem a utilização do método Bagging, verificando se algum dos modelos estabelecidos apresenta melhor desempenho.

A partir dos resultados apresentados, observando os dados em sua totalidade, não foi possível identificar a prevalência de um modelo em relação aos demais. Cada banco de dados analisado apresentou um modelo diferente com melhor desempenho em algum parâmetro específico. Isso evidencia que, para um banco de dados particular, um modelo pode gerar previsões mais precisas dependendo das características dos dados e dos critérios avaliados.

Para cada modelo desenvolvido, elaboramos sua versão utilizando o método Bagging para seis diferentes tamanhos de blocos. Contudo, em nenhuma das aplicações foi possível identificar um desempenho consistentemente superior dos modelos com Bagging, independentemente do tamanho do bloco utilizado. Observou-se uma alternância entre os modelos com e sem Bagging quanto ao melhor desempenho, sendo que, em algumas situações, o modelo sem Bagging superou os demais.

Concluimos que, para cada conjunto de dados, um modelo baseado em uma distribuição específica pode apresentar melhor desempenho, ainda que as diferenças observadas sejam geralmente pequenas. No entanto, em relação ao método Bagging, os resultados obtidos não indicam melhorias significativas na acurácia das previsões realizadas.

Referências

- BELFRAGE, M. *ACDm: Tools for Autoregressive Conditional Duration Models*. [S.l.], 2022. R package version 1.0.4.2. Disponível em: (<https://CRAN.R-project.org/package=ACDm>).
- BHATTI, C. R. The birnbaum–saunders autoregressive conditional duration model. *Mathematics and Computers in Simulation*, Elsevier, v. 80, n. 10, p. 2062–2078, 2010.
- BREIMAN, L. Bagging predictors. *Machine Learning*, Springer, v. 24, n. 2, p. 123–140, 1996.
- CAMARGO, J. S. Método bagging para aprimoramento de previsões de séries temporais. Universidade Federal de São Carlos, 2021.
- ENGLE, R. F.; RUSSELL, J. R. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, JSTOR, p. 1127–1162, 1998.
- GOURIEROUX, C.; MONFORT, A.; TROGNON, A. Pseudo maximum likelihood methods: Theory. *Econometrica: journal of the Econometric Society*, JSTOR, p. 681–700, 1984.
- GRAMMIG, J.; MAURER, K.-O. Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal*, Wiley Online Library, v. 3, n. 1, p. 16–38, 2000.
- MAYOROV, K. *Modelling trade durations with the Birnbaum-Saunders autoregressive model*. Dissertação (Mestrado), 2011.
- MORAIS, J. P. O. *Aplicação de Bagging em Modelo Gama Generalizado Autoregressivo de Duração Condicional*. [S.l.], 2023.
- SAULO, H.; BALAKRISHNAN, N.; VILA, R. On a quantile autoregressive conditional duration model. *Mathematics and Computers in Simulation*, Elsevier, v. 203, p. 425–448, 2023.
- TSAY, R. Autoregressive conditional duration models. In: _____. [S.l.: s.n.], 2009. p. 1004–1024. ISBN 978-1-4039-1800-0.
- ZHANG, M. Y.; RUSSELL, J. R.; TSAY, R. S. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, Elsevier, v. 104, n. 1, p. 179–207, 2001.

Apêndice

Link para o código: <https://github.com/14mateuskilson/TCC-bagging-ACDm>