Universidade Federal de São Carlos

Centro de Ciências Exatas e de Tecnologia

Departamento de Estatística

Estudo e comparação das técnicas de Validação Cruzada desenvolvidas para Séries Temporais

Daniel Simionato

Trabalho de Conclusão de Curso

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Estudo e comparação das técnicas de Validação Cruzada desenvolvidas para Séries Temporais

Daniel Simionato

Orientadora: Maria Sílvia de Assis Moura

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade Federal de São Carlos - DEs-UFSCar, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

São Carlos

Abril de 2022

FEDERAL UNIVERSITY OF SÃO CARLOS

EXACT AND TECHNOLOGY SCIENCES CENTER

DEPARTMENT OF STATISTICS

Study and comparison of Cross-Validation techniques developed for Time Series

Daniel Simionato

Advisor: Maria Sílvia de Assis Moura

Bachelors dissertation submitted to the Department of Statistics, Federal University of São Carlos - DEs-UFSCar, in partial fulfillment of the requirements for the degree of Bachelor in Statistics.

São Carlos

April 2022

Daniel Simionato

Estudo e comparação das técnicas de Validação Cruzada desenvolvidas para Séries Temporais

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Daniel Simionato e aprovado pela banca examinadora.
Aprovado em 18 de abril de 2022

Banca Examinadora:

- Profa. Dra. Maria Sílvia de Assis Moura (Orientadora)
- Prof. Dr. Luis Aparecido Milan
- Prof. Dr. Renato Jacob Gava

Dedico este trabalho aos meus pais, Juliana Pedroso Simionato e Osmair Donizeti Simionato por todo suporte

Agradecimentos

Agradeço aos meus pais e meu irmão, que sempre me apoiaram e incentivaram nos estudos.

Ao meu namorado Renato Nakahara, que jamais me negou apoio, carinho e incentivo. Obrigado por todo o companheirismo ao longo deste percurso.

A professora Dra. Maria Sílvia de Assis Moura, por ter sido minha orientadora e ter desempenhado tal função com dedicação e amizade.

A banca examinadora e todo corpo docente do curso de bacharelado de estatística da UFSCar, por todos os conselhos, pela ajuda e pela paciência com a qual guiaram o meu aprendizado.

Resumo

Testar a capacidade de generalização de um algoritmo obtido é crucial para qualquer metodologia de predição, para tal foram desenvolvidos os métodos de validação-cruzada. No entanto, quando estamos tratando com dados que tenham dependência entre as observações, como o caso de séries temporais, as metodologias usuais de validação-cruzada não são apropriadas. Atualmente não há uma maneira única e usual de se testar a capacidade de generalização de um modelo preditivo para séries temporais. À vista disso, nesse trabalho vamos estudar e comparar quatro diferentes variações de métodos para validação-cruzada, isso será realizado por meio simulações de séries temporais diversas.

Palavras-chave: *predição, séries temporais, validação-cruzada.*

Abstract

Testing the generalization capacity of an algorithm obtained is crucial for any prediction methodology, for which cross-validation methods were developed. However, when dealing with data that have dependency among observations, as in the case of time series, the usual cross-validation methodologies are not appropriate. Currently, there is no single and usual way to test the generalizability of a predictive model for time series. In this work we will study and compare four different variations of methods for cross-validation, this will be done through simulations of different time series.
Keywords: *prediction, time series, cross validation.*

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1

Introdução

Segundo Izbicki e dos Santos (2020), metodologias para predição surgiram há mais de dois séculos, possuindo na atualidade um papel central dentro da estatística. Na prática, é comum desenvolver diversos modelos a fim de buscar o que possui melhor poder preditivo, porém muitas vezes se corre o risco de realizar um sub-ajuste ou super-ajuste aos dados.

A fim de solucionar esse problema de sub ou super-ajuste de um modelo preditivo, foram desenvolvidas metodologias de validação-cruzada. Essas metodologias consistem basicamente em utilizar uma parte dos dados para treinar o modelo, enquanto o restante dos dados é utilizado para validar e testar a capacidade de generalização do modelo treinado. Contudo, quando o objetivo é realizar a predição de uma série temporal as metodologias padrões de validação-cruzada não podem ser aplicadas, uma vez que assumem que os dados são independentes e identicamente distribuídos.

Sendo assim, com o objetivo de conseguir comparar diversos modelos preditivos para séries temporais e testar a capacidade de generalização desses modelos, foram desenvolvidas algumas adaptações das metodologias de validação-cruzada. Neste trabalho, vamos estudar e comparar as metodologias *Modified cross-validation*, *Blocked cross-validation*, *HV-Blocked cross-validation* e *Markov cross-validation*, desenvolvidas por McQuarrie e Tsai (1998), Snijders (1988), Racine (2000) e Jiang e Wang (2017), respectivamente.

Para atender o objetivo de comparar as metodologias citadas, este trabalho está estruturado da seguinte maneira, no Capítulo 2 é apresentado a definição de séries temporais, dois modelos usuais para realizar a predição desse tipo de dados, métricas para calcular a perda de um modelo preditivo e as metodologias de validação cruzada desenvolvidas para esse tipo de dado a fim de estimar a perda do modelo. No Capítulo 3 utilizamos o método de Monte Carlo para realizar um comparativo das metodologias apresentadas realizando diversas simulações de séries temporais diversas e ordenando os resultados por postos. Por fim, no Capítulo 4 apresentamos a conclusão desse trabalho.

Capítulo 2

Material e Métodos

2.1 Séries temporais

De acordo com Morettin e Toloi (2006), uma série temporal é um conjunto de observações, em que cada observação é coletada em um tempo específico t. Normalmente o parâmetro t representa o tempo em uma série temporal, porém pode ser qualquer parâmetro físico, como espaço, altura, volume, entre outros. Uma característica importante das séries temporais é que as observações adjacentes são dependentes. Neste trabalho, vamos denotar uma série temporal de tamanho n como $Z = \{z_t\}, t = 1, \cdots, n$.

Se os valores de uma série temporal flutuam em torno de uma mesma média e possuem variância constante ao longo do período, ou seja, apresenta um equilíbrio estável, denominados a série de estacionária. Caso contrário, dizemos que ela apresenta tendência. (Morettin e Toloi (2006))

Dizemos que uma série possui sazonalidade caso o comportamento que ocorre durante um tempo se repita a cada período idêntico de tempo, ou seja, dizemos que uma série possui comportamento sazonal com período s, caso ocorra similaridade a cada s observações.

2.2 Algoritmos para predição de séries temporais

Vamos utilizar nesse trabalho em suma dois modelos de predição para séries temporais, a suavização exponencial e o modelo ARIMA.

Conforme Hyndman e Athanasopoulos (2021), a suavização exponencial consiste em gerar as predições a partir de uma média ponderada das observações anteriores, no qual quanto mais recente for a observação maior sera seu peso associado. Essa metodologia tende a ser rápida e gera estimativas confiáveis para uma grande variedade de séries temporais, por isso é frequentemente utilizada.

Enquanto a metodologia de suavização exponencial é baseada na tendência e sazonalidade da série, o modelo ARIMA busca descrever as autocorrelações das observações. O modelo ARIMA é definido pelos parâmetros (p,d,q), que correspondem a ordem da parte autoregressiva, grau de diferenciação necessária para tornar a série estacionária e a ordem da parte de média móveis, respectivamente.

2.3 Avaliação de performance de modelos preditivos

Um bom modelo de predição é aquele que realiza estimativas próximas do valor real da série, ou seja, com um erro pequeno. Conforme Morettin e Toloi (2006), podemos definir o erro de previsão h passos a frente no instante t como a diferença entre o valor real de série no instante t e a previsão deste valor feita h instantes antes disso.

$$\epsilon_h(t) = z_t - \hat{z}_{t-h}(h),$$

em que

- $\epsilon_h(t)$ é o erro de previsão h passos a frente no instante t;
- z_t corresponde ao valor real da série no instante t;
- $\hat{z}_{t-h}(h)$ é a previsão do valor da série no instante t realizada h instantes antes disso.

Para avaliar o quão boas foram as predições de um algoritmo iremos utilizar o acurácia do erro de predição absoluto. (Bergmeir *et al.*, 2018)

$$APAE_m = \left| \frac{\hat{g^m} - L^m}{L^m} 100 \right|,$$

em que

- $\hat{g^m}$ é a perda estimada a partir do conjunto de treinamento com um modelo preditivo m e a metodologia g;
- L^m é a perda calculada a partir do conjunto de teste com um modelo preditivo m.

Utilizamos como função de perda o erro quadrático médio, que é dado pela fórmula abaixo.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\epsilon_h(t))^2$$

2.4 Metodologias para Validação Cruzada

Conforme Izbicki e dos Santos (2020), se as métricas apresentadas na seção 2.3 forem utilizadas para seleção de modelos podem levar a um super-ajuste, uma vez que o algoritmo é construído de modo a ajustar bem a amostra $Z = \{z_t\}, t = 1, \cdots, n$.

Para solucionar esse problema, foram desenvolvidas metodologias de validação cruzada. Essas metogolodias visam avaliar a capacidade de generalização de um modelo de predição, verificando se existe um sub-ajuste ou super-ajuste aos dados utilizados para treino do algoritmo.

Uma das metodologias mais conhecidas de validação cruzada é a *k-fold*. Essa técnica é adequada quando podemos assumir que as observações são independentes e identicamente distribuídas (i.i.d.), logo não é ideal a aplicação em séries temporais.

Nesse trabalho vamos estudar e comparar algumas variantes da técnica de validação cruzada adequadas para séries temporais.

2.4.1 Validação Cruzada *K-fold* (CV)

Segundo Izbicki e dos Santos (2020), a metodologia consiste em primeiramente separar o conjunto de dados em k subconjuntos de aproximadamente o mesmo tamanho, posteriormente cada bloco é escolhido interativamente para o teste. O modelo é treinado nos *k-1* subconjuntos, enquanto o subconjunto restante é utilizando para estimar o erro.

2.4.2 *Out-Of-Sample Validation* (OOS)

Seja $Z_1, \cdots, Z_N$ uma série temporal, a validação fora da amostra (OOS) consiste em dividir a base em duas partes, o período inicial $Z_1, \cdots, Z_k$ será utilizado para treinar o algorítimo de predição enquanto o segundo período $Z_{k+1}, \cdots, Z_N$ será utilizado para testar e validar as métricas de performance.

Figura 2.1: Técnica de validação fora da amostra

Para este trabalho utilizamos a proporção de 70% dos dados para o conjunto de treino e 30% para o conjunto de teste.

2.4.3 *Modified cross-validation* (ModCV)

O método ModCV desenvolvido por McQuarrie e Tsai (1998), também denominado de *non-dependent cross-validation* por Bergmeir e Benítez (2012), consiste em inicialmente embaralhar suas obervações aleatoriamente e dividir em *k-folds* (lotes) de aproximadamente o mesmo tamanho, procedimento similar a validação cruzada *k-fold*. Posteriormente, são removidas as observações da base de treinamento que estejam dentro de uma determinada distância de tempo t das observações da base de teste, esse procedimento garante independência entre as bases de treino e teste. Usualmente, utiliza-se a ordem de autocorrelação da série (p) como a distância de tempo para remover as observações.

Um ponto de atenção para utilização desse método é que dependendo da quantidade de *lags* e do número de *k-folds* utilizados, a remoção das observações dependentes pode levar a uma perda significativa de dados, podendo ocasionar um sub-ajuste do modelo.

2.4.4 *Blocked cross-validation* (BCV)

O método BCV prosposto por Snijders (1988) é semelhante a validação cruzada *k-fold*, com a diferença de não realizar o embaralhamento inicial das observações.

Logo, a metodologia consiste em primeiramente separar o conjunto de dados em k subconjuntos de mesmo tamanho, posteriormente cada bloco é escolhido interativamente para o teste. O modelo é treinado nos *k-1* subconjuntos, enquanto o subconjunto restante é utilizando para estimar o erro. Bergmeir e Benítez (2012) recomendam utilizar essa metodologia quando a série temporal que se deseja realizar a predição é estacionária.

2.4.5 *HV-Blocked cross-validation* (hvBCV)

O método hvBCV proposto por Racine (2000) é uma modificação do BCV apresentando na Seção 2.4.4, a fim de aumentar a independência entre as bases de treino e teste.

O precedimento também é similar à validação cruzada*k-fold*, não havendo o embaralhamento das observações, além disso são removidas as observações adjacentes entre as bases de treino e teste a fim de aumentar a independência entre as observações. Conforme Cerqueira *et al.* (2021), para cada interação deve-se retirar as p observações adjacentes entre as bases de treino e teste, onde p corresponde a ordem de autocorrelação da série temporal em estudo.

Racine (2000) demonstra que a partir do hvBCV a chance de selecionar o modelo com a melhor capacidade preditiva converge para 1 conforme o número total de observações se aproxima do infinito, ou seja, a metologia é mais apropriada para séries temporais com grande número de observações.

2.4.6 *Markov cross-validation* (M-CV)

Segundo Jiang e Wang (2017), as adaptações da metodologia usual de validação cruzada apresentadas nas Seções 2.4.3, 2.4.4 e 2.4.5 não conseguem resolver os problemas de periodicidade, sobreposição e autocorrelação ao mesmo tempo. Visando atender a esses três critérios foi proposto o método M-CV, o procedimento consiste em primeiramente repartir a base de dados em alguns subconjuntos que representem a série original, posteriormente é aplicado a validação cruzada *k-fold*, com k igual a 2, em cada subconjunto.

Mais especificamente, dada um série temporal $Z = \{z_t\}, t = 1, \cdots, n$ com autocorrelação de ordem p, iremos primeiramente repartir os dados em dois subconjuntos (Z^+ e Z^-) a partir do algoritmo 1.

Algoritimo 1 Interação de Markov

Pré: Série temporal $Z = \{z_t\}, t = 1, \cdots, n$, com autocorrelação p

Pós: O estado (d) é iniciado aleatoriamente com algum dos casos:

$d_1 = +1, d_2 = +1;$

$d_1 = +1, d_2 = -1;$

$d_1 = -1, d_2 = +1;$

$d_1 = -1, d_2 = -1;$

para $t = 3, \cdots, n$ **faça**

  se $(d_{t-1} < 0)\&(d_{t-2} < 0)$ **então**

    $d_t = +1$

  senão se $(d_{t-1} > 0)\&(d_{t-2} > 0)$ **então**

    $d_t = -1$

  senão se $u > 0,5$ **então**

    $d_t = -1$

  senão

    $d_t = +1$

  fim se

fim para

Logo, podemos separar os dados nos dois subconjuntos seguindo a regra dada por

$$z_t \in \begin{cases} Z^+, & \text{se } d_t > 0 \\ Z^-, & \text{se } d_t < 0 \end{cases}$$

Posteriormente, cada subconjunto deve ser dividido em mais m subconjuntos, com m definido como

$$m = \begin{cases} 2p/3, & \text{se } p\%3 = 0 \\ 2\lfloor p/3 \rfloor + 2 & \text{caso contrário} \end{cases}$$

onde % é a operação módulo e $\lfloor . \rfloor$ é a função piso. Essa segunda divisão em *2m* subgrupos é realizada a partir do seguinte critério:

$$Id_t = t\%m + 1 + I(d_t > 0)m, \; I(.) \text{ é a função indicadora;}$$
$$Z_u = \{z_t | Id_t = u\}, u = 1, \cdots, 2m.$$

Por último, devemos dividir cada subconjunto Z_u em mais dois grupos Z_{up} e Z_{ui} de acordo com se a posição (número ordinal) da amostra for par ou impar, respectivamente. Completamos então o método M-CV aplicando a validação cruzada *2-fold* e cada $2m$ subconjuntos, utilizando a separação definida de Z_{up} e Z_{ui}.

Capítulo 3

Aplicação

3.1 Simulação de dados

Neste capítulo realizamos primeiramente a simulação séries temporais diversas, esses dados serviram para um melhor entendimento e avaliação das metodologias apresentadas no Capítulo 2. Cada série temporal foi simulada com o intuito de assumir características diferentes das demais, alternando no tamanho amostral, tendência e/ou composição. Aplicamos as metodologias predição para séries temporais apresentadas na Seção 2.2 para cada uma das metodologias de Validação Cruzadas estudadas.

Utilizamos o método de Monte Carlo a fim de comparar as metodologias de validação cruzada, simulando um conjunto de mil séries para cada combinação de características diferentes. Para cada combinação de modelo preditivo e série temporal foram ordenadas de menor para maior (ranking) as metodologias de validação cruzada conforme a métrica APAE. Calculamos então o ranking médio e o desvio padrão para cada metodologia, obtendo um indicativo de qual metodologia é o melhor estimador da perda naquela representação.

3.1.1 Séries Temporais Estacionárias

Séries simuladas:

- ARIMA(p=0.7,d=0,q=0), tamanho amostral n=50 (100AR50);
- ARIMA(p=0.7,d=0,q=0), tamanho amostral n=100 (100AR100);
- ARIMA(p=0.7,d=0,q=0), tamanho amostral n=1000 (100AR1000);

- ARIMA(p=0.7,d=0,q=0.7), tamanho amostral n=1000 (101AR1000);
- ARIMA(p=[0.7,-0.7],d=0,q=0), tamanho amostral n=1000 (200AR1000);

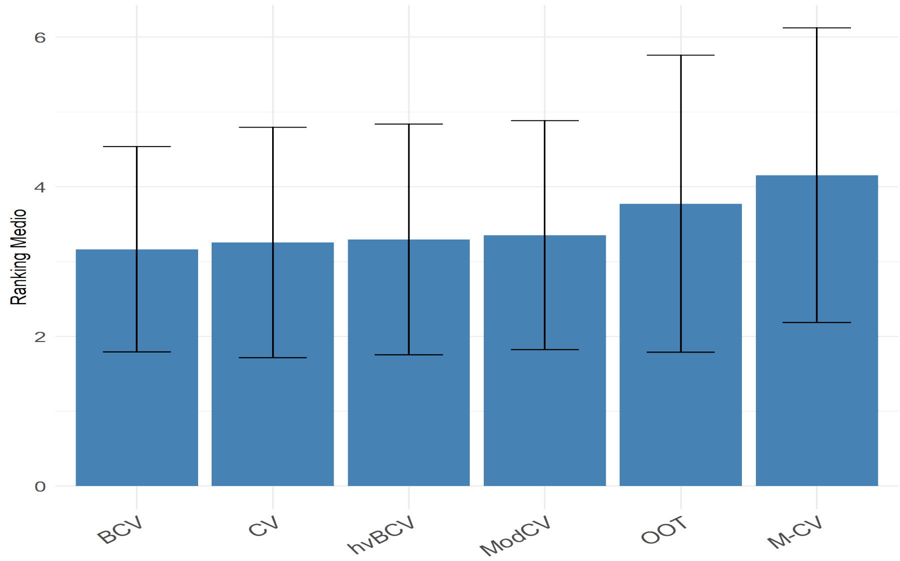

Figura 3.1: Ranking Médio - Suavização Exponencial, série ARIMA(1,0,0) n = 50

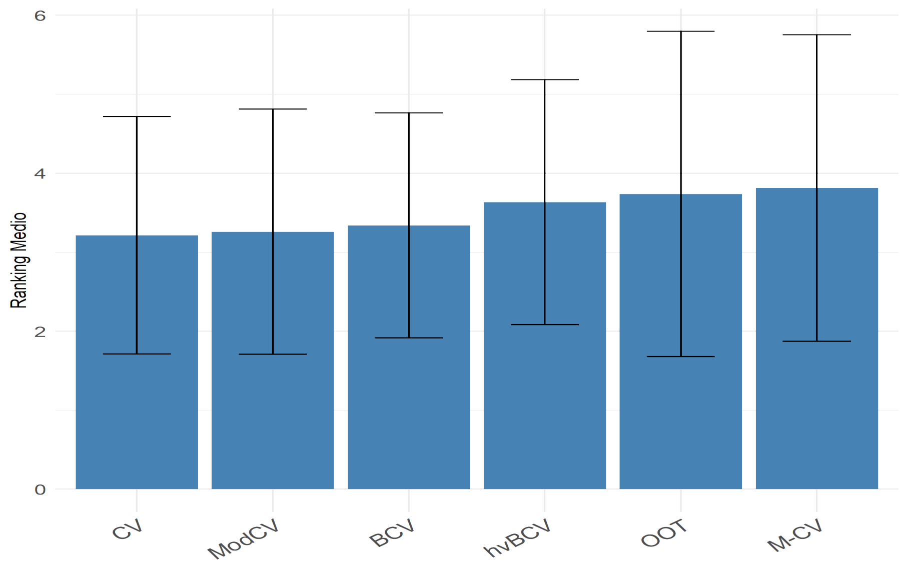

Figura 3.2: Ranking Médio - Modelo ARIMA, série ARIMA(1,0,0) n = 50

Para ambos os modelos as metodologias BCV, CV, hvBCV e ModCV ficaram entre o melhor estimador do erro, enquanto o M-CV ficou como o pior e OOT como o segundo pior. Em geral com essas configurações de dados podemos dizer que o CV se saiu melhor.

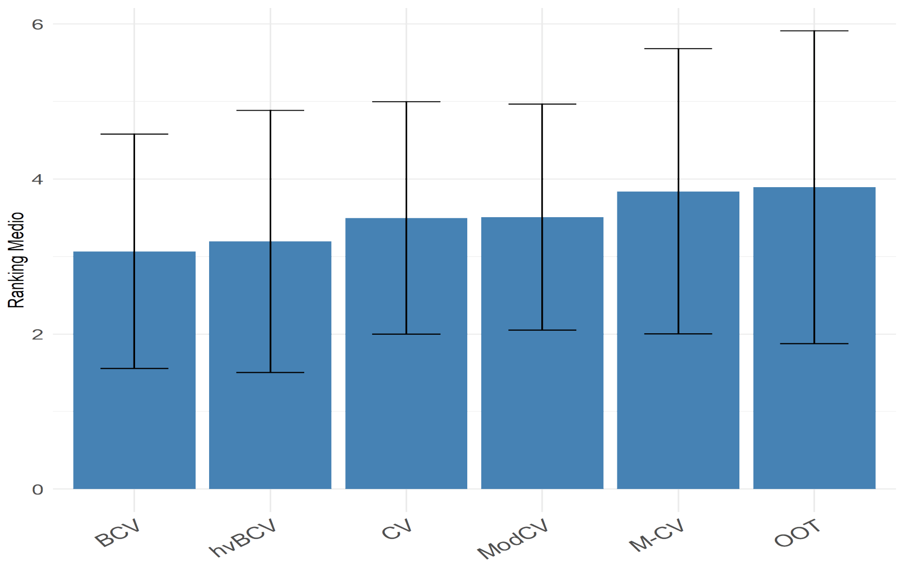

Figura 3.3: Ranking Médio - Suavização Exponencial, série ARIMA(1,0,0) n = 100

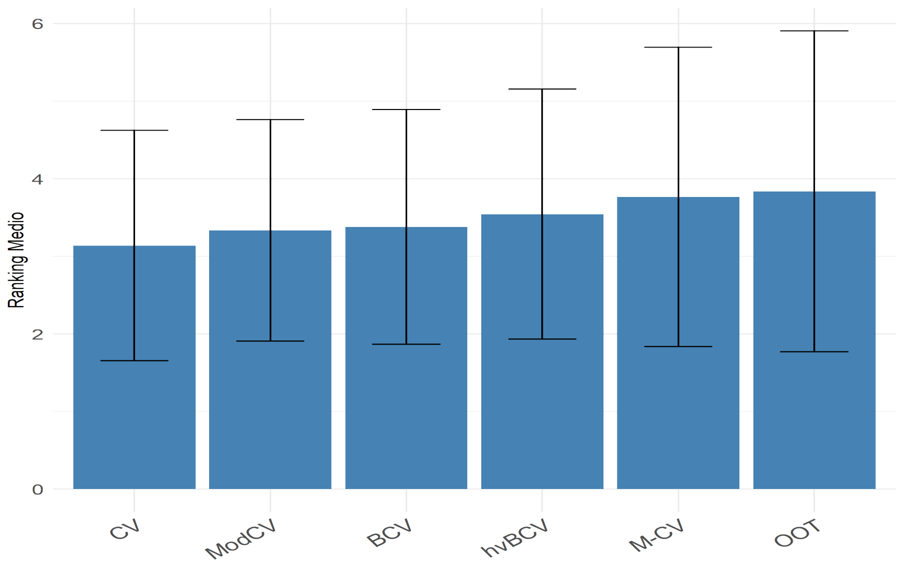

Figura 3.4: Ranking Médio - Modelo ARIMA, série ARIMA(1,0,0) n = 100

Novamente, para ambos os modelos as metodologias BCV, CV, hvBCV e ModCV ficaram entre o melhor estimador do erro, porém alterou a ordem do pior e segundo pior. Em geral com essas configurações de dados podemos dizer que o CV empatou com BCV.

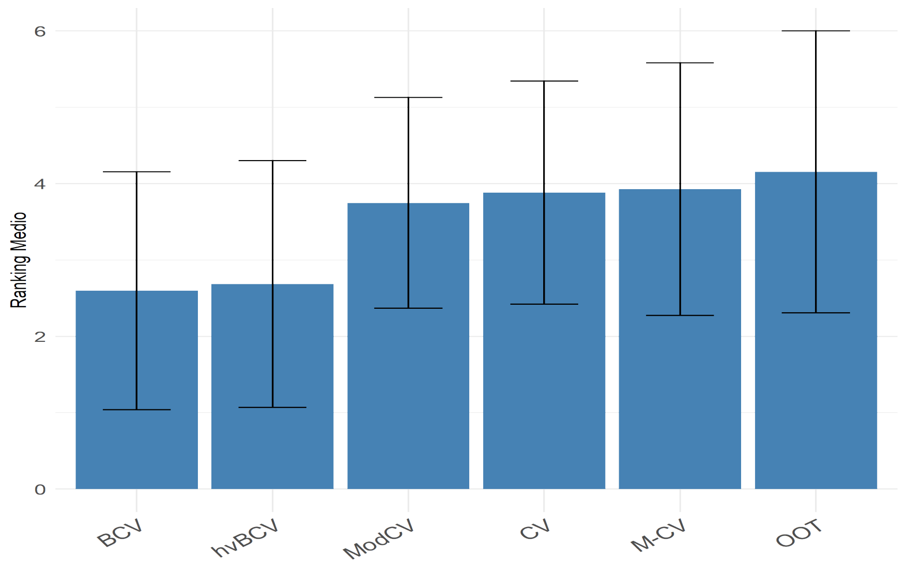

Figura 3.5: Ranking Médio - Suavização Exponencial, série ARIMA(1,0,0) n = 1000

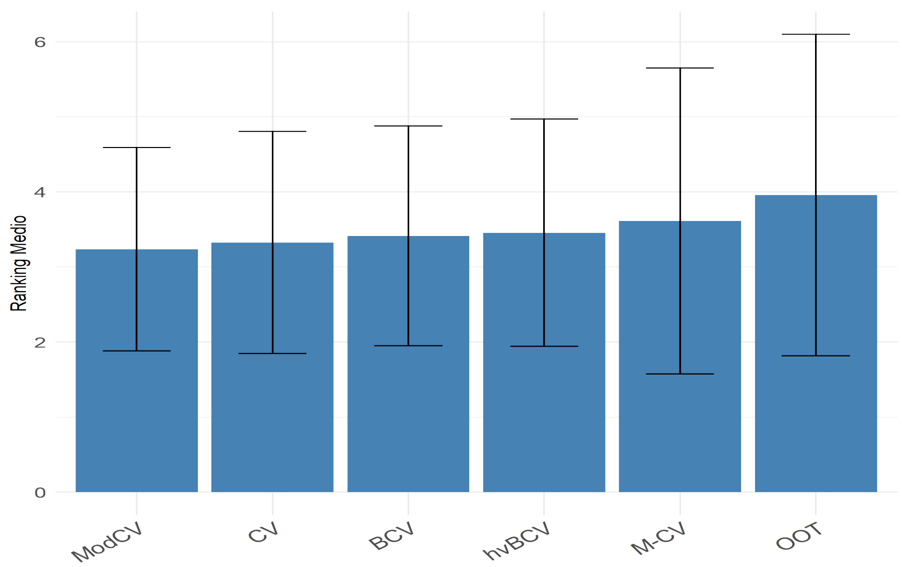

Figura 3.6: Ranking Médio - Modelo ARIMA, série ARIMA(1,0,0) n = 1000

Para essa simulação as metodologias M-CV e OOT novamente ficaram novamente entre as piores, porém quando foi utilizado o modelo preditivo de Suavização Exponencial é possível observar que as metodologias ModCV e CV ficaram mais próximas ao M-CV e OOT do que BCV e hvBCV.

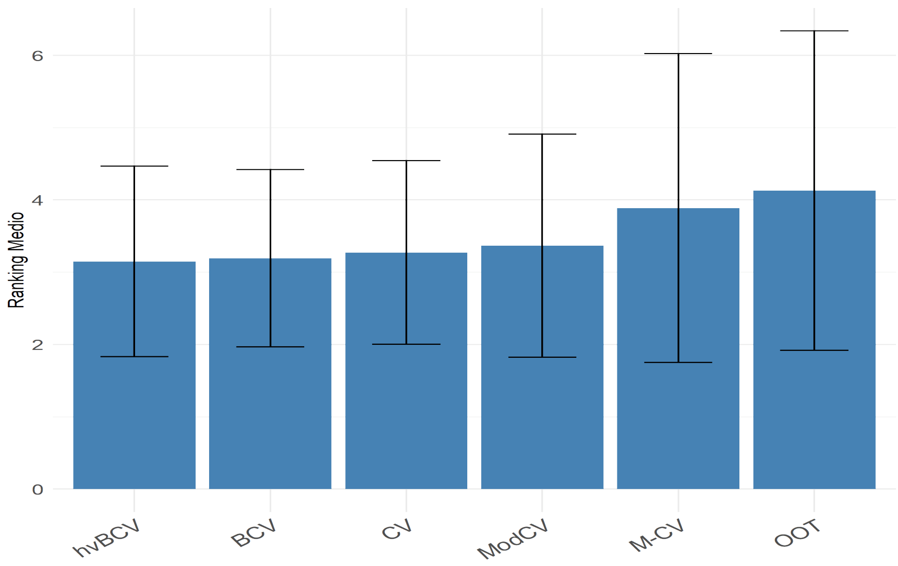

Figura 3.7: Ranking Médio - Suavização Exponencial, série ARIMA(2,0,0) n = 1000

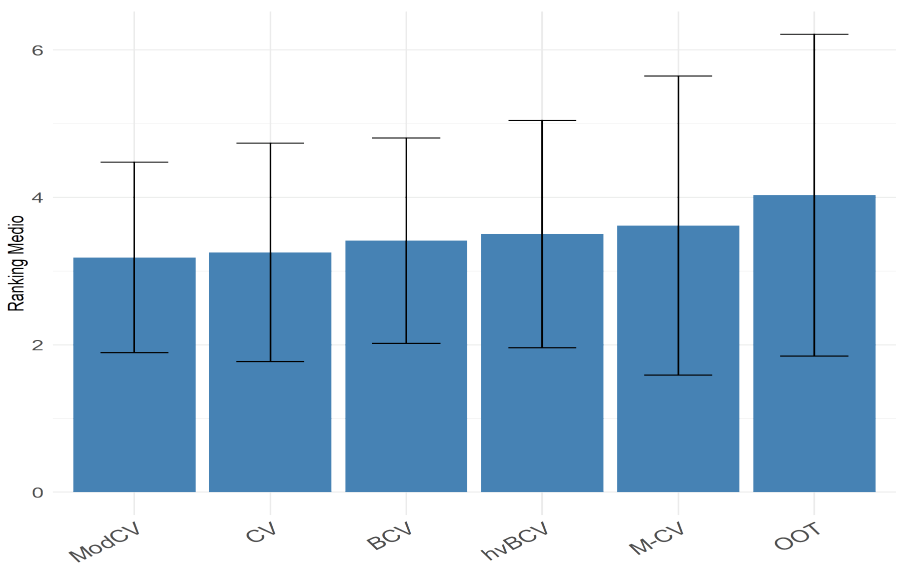

Figura 3.8: Ranking Médio - Modelo ARIMA, série ARIMA(2,0,0) n = 1000

Nesse caso, simulando séries temporais com autocorrelação de ordem 2 e tamanho 1000, obtivemos resultados similares a simulação de séries temporais com autocorrelação ordem 1 e tamanho 100. As metodologias BCV, CV, hvBCV e ModCV ficaram entre o melhor estimador do erro, enquanto OOT e M-CV entre as piores.

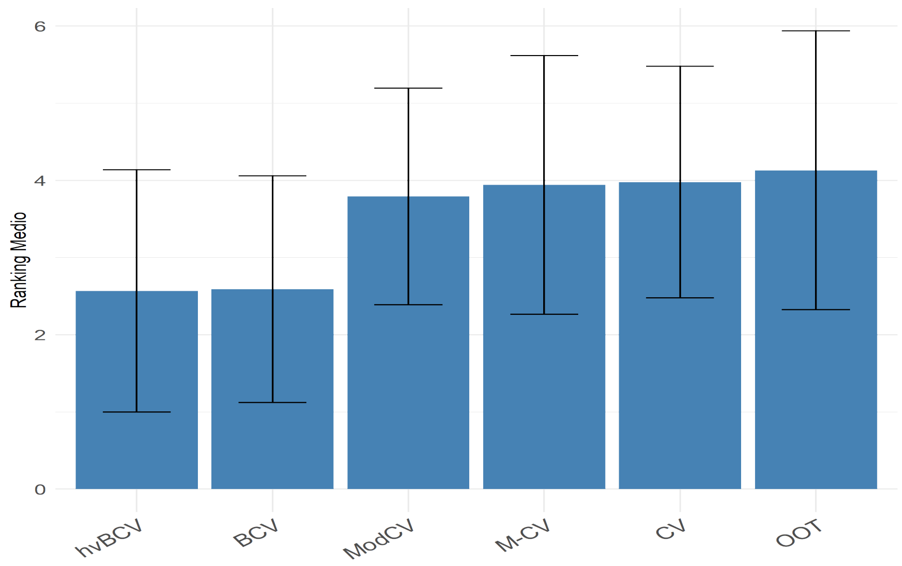

Figura 3.9: Ranking Médio - Suavização Exponencial, série ARIMA(1,0,1) n = 1000

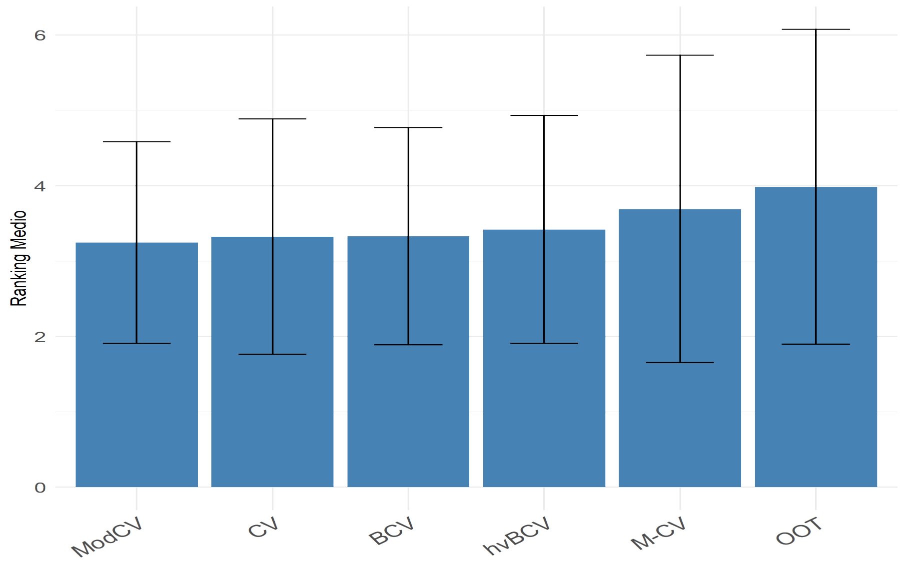

Figura 3.10: Ranking Médio - Modelo ARIMA, série ARIMA(1,0,1) n = 1000

Adicionando um componente de médias móveis nas séries temporais simuladas obtivemos resultados similares aos obtidos com as séries com tamanho 1000 e ordem de autocorrelação igual a 1, com as metodologias hvBCV e BCV se destacando com o modelo de Suavização Exponencial.

Agrupando as resultados pela média e ordenando as metodologias pelo *ranking*, é possível obter os seguintes resultados para as séries temporais estacionárias apresentados na Tabela 3.1.

Tabela 3.1: Valor ordinal do *ranking* médio para as Séries estacionárias

Dados	OOT	CV	ModCV	BCV	hvBCV	M-CV
Séries Estacionárias	6	4	3	1	2	5

3.1.2 Séries Temporais não Estacionárias

Séries simuladas:

- ARIMA(p=0.7,d=1,q=0), tamanho amostral n=50 (110AR50);
- ARIMA(p=0.7,d=1,q=0), tamanho amostral n=100 (110AR100);
- ARIMA(p=0.7,d=1,q=0), tamanho amostral n=1000 (110AR1000);
- ARIMA(p=0.7,d=1,q=0.7), tamanho amostral n=1000 (111AR1000);
- ARIMA(p=[0.7,-0.7],d=1,q=0), tamanho amostral n=1000 (210AR1000);

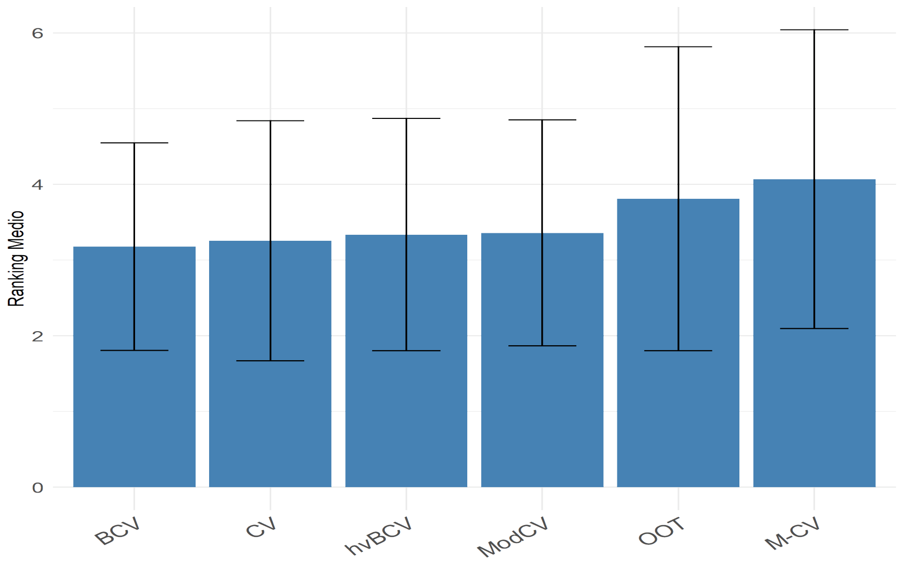

Figura 3.11: Ranking Médio - Suavização Exponencial, série ARIMA(1,1,0) n = 50

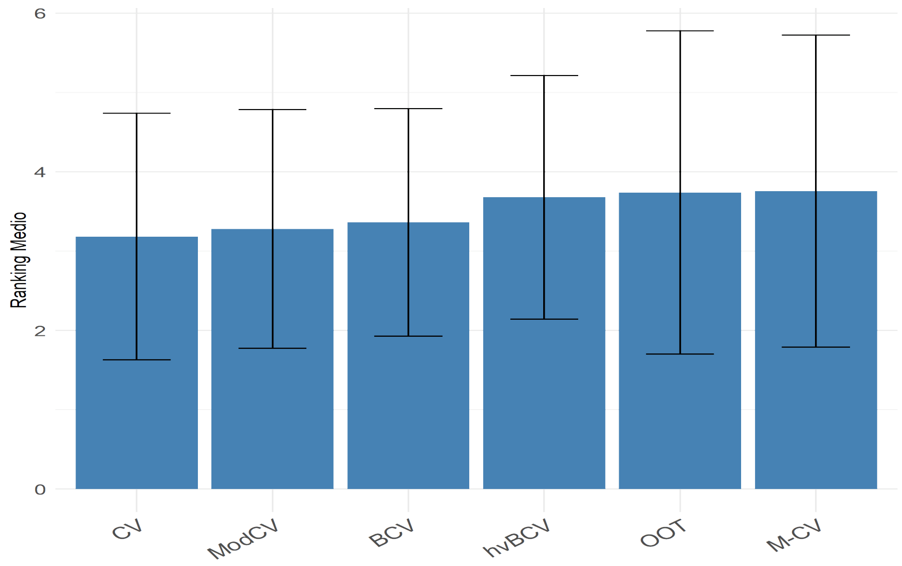

Figura 3.12: Ranking Médio - Modelo ARIMA, série ARIMA(1,1,0) n = 50

Neste caso a metodologia BVC apresenta os melhores resultados, mesmo comparando com as metodologias CV, hvBCV e ModCV, as quais apresentaram valores próximos. As metodologias OOT e M-CV apresentaram os piores resultados entre as comparações, com a metodologia M-CV foi a última entre as comparações de ranking.

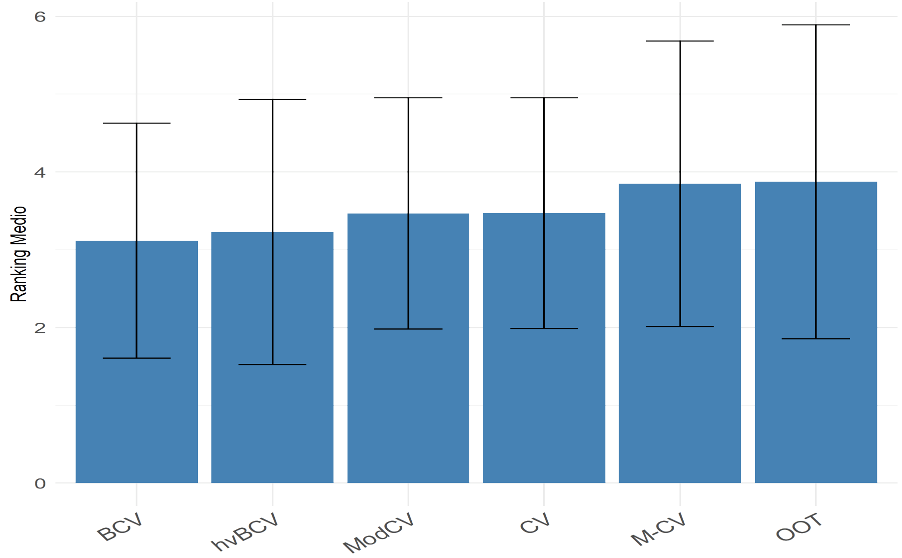

Figura 3.13: Ranking Médio - Suavização Exponencial, série ARIMA(1,1,0) n = 100

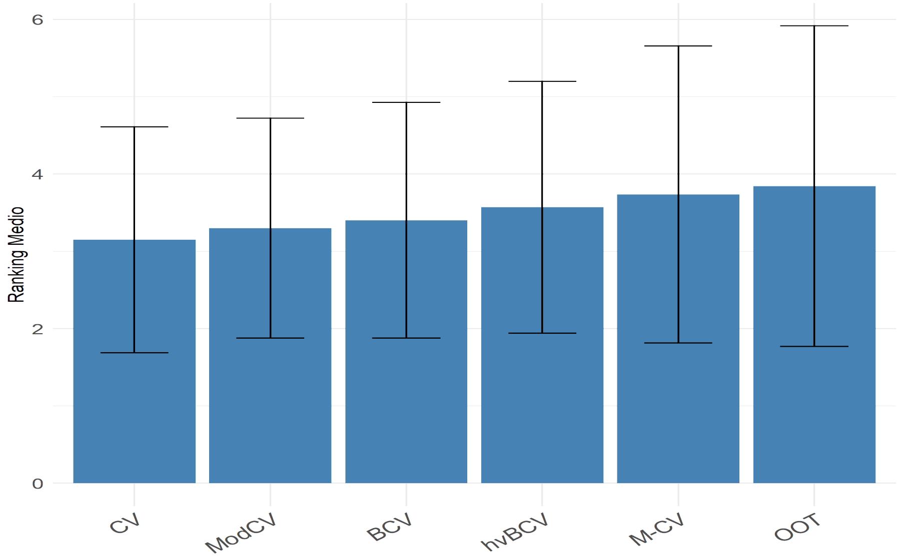

Figura 3.14: Ranking Médio - Modelo ARIMA, série ARIMA(1,1,0) n = 100

Analisando a Figura 3.13 e a Figura 3.14 é notório que obtivemos resultados semelhantes as simulações da série também de tamanho 100 e ordem de autocorrelação 1, porém sem tendência. Com as metodologias BCV, hvBCV, ModCV e Cv ficando próximas entre as melhores.

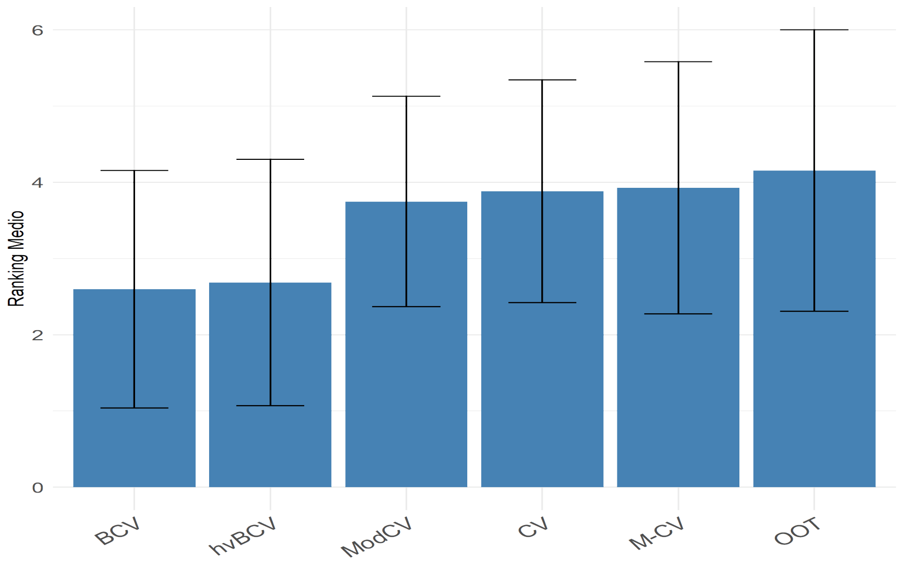

Figura 3.15: Ranking Médio - Suavização Exponencial, série ARIMA(1,1,0) n = 1000

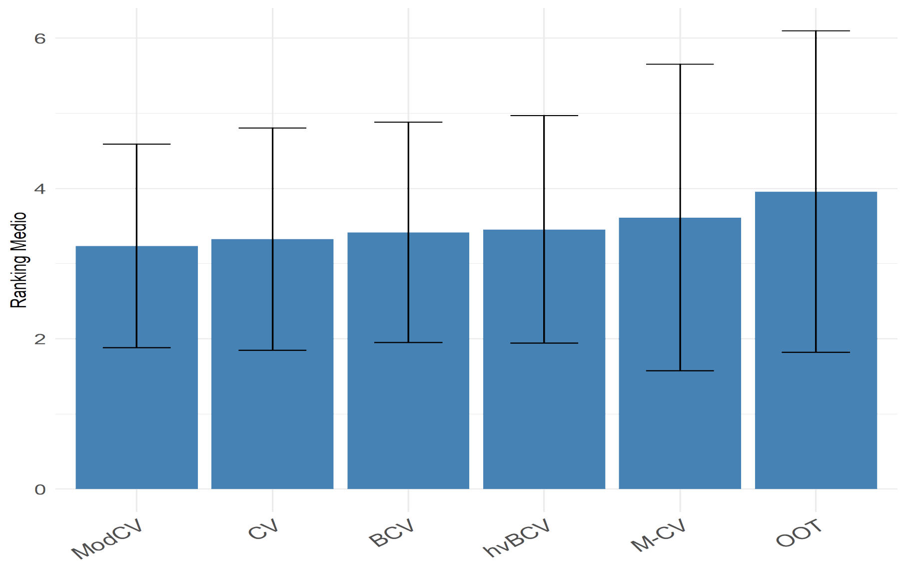

Figura 3.16: Ranking Médio - Modelo ARIMA, série ARIMA(1,1,0) n = 1000

Similar com o que ocorreu nas Figuras 3.5 e 3.6, quando analisamos a estimação do risco com modelo preditivo de Suavização Exponencial é possível observar que as metodologias ModCV e CV ficaram mais próximas ao M-CV e OOT entre as piores do que as BCV e hvBCV.

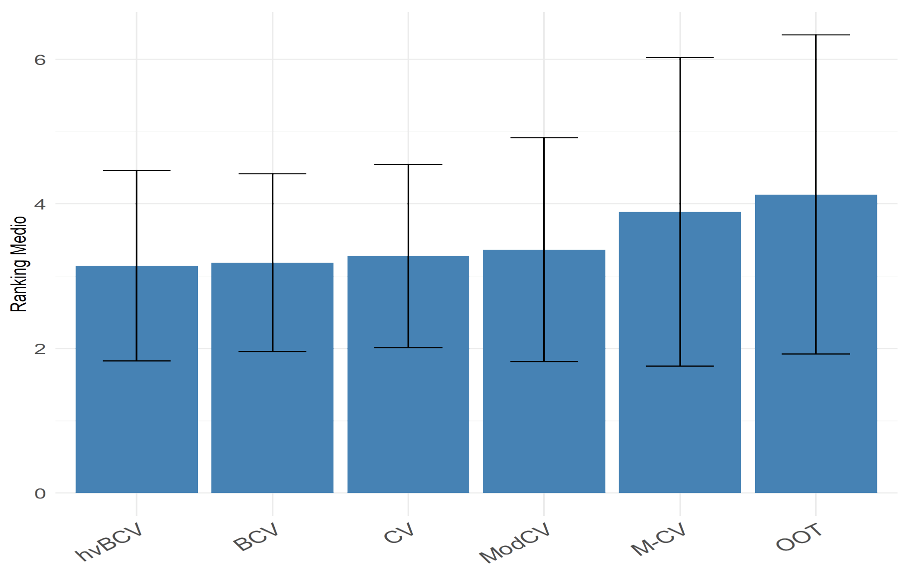

Figura 3.17: Ranking Médio - Suavização Exponencial, série ARIMA(2,1,0) n = 1000

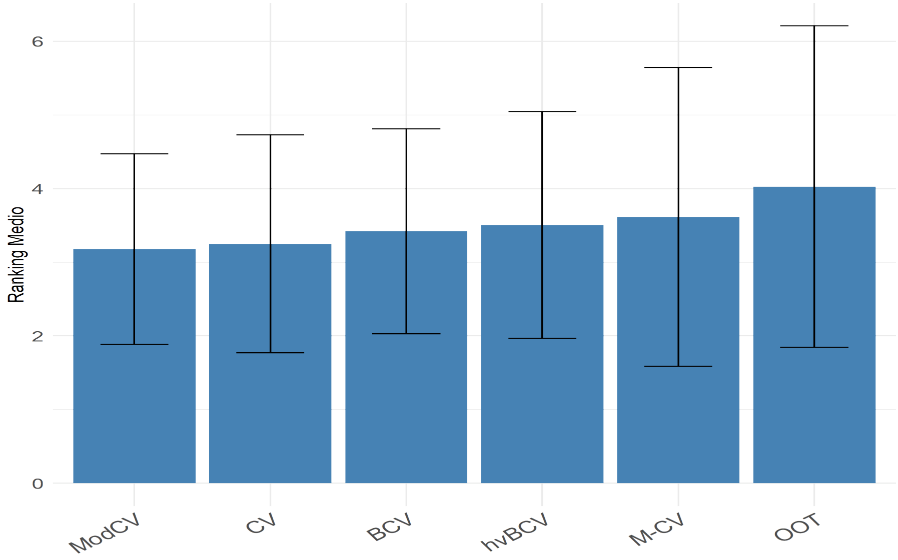

Figura 3.18: Ranking Médio - Modelo ARIMA, série ARIMA(2,1,0) n = 1000

Para ambos os modelos as metodologias BCV, CV, hvBCV e ModCV ficaram entre as melhores para estimador do erro preditivo, enquanto o OOT ficou como o pior e M-CV como o segundo pior.

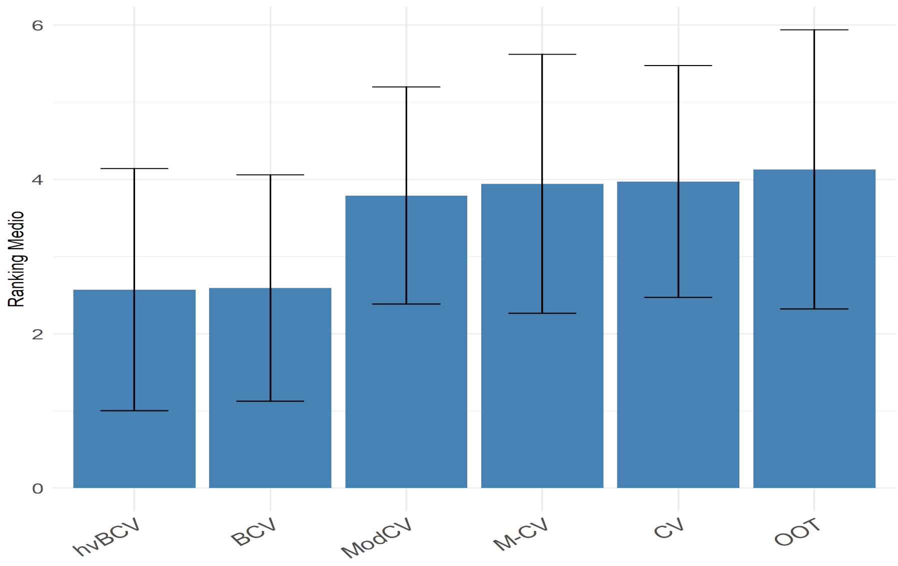

Figura 3.19: Ranking Médio - Suavização Exponencial, série ARIMA(1,1,1) n = 1000

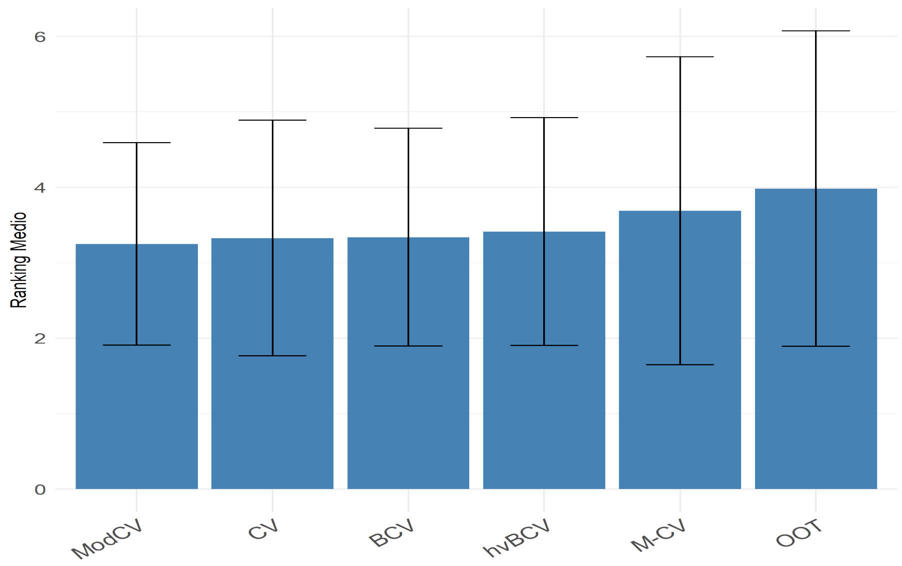

Figura 3.20: Ranking Médio - Modelo ARIMA, série ARIMA(1,1,1) n = 1000

Por último, simulando séries temporais ARIMA(1,1,1), obtivemos resultados similares a simulação de séries ARIMA(1,0,1), as metodologias hvBCV e BCV se destacando com o modelo de Suavização Exponencial.No geral para esse caso podemos dizer que a metodologia BCV se saiu melhor.

Agrupando as resultados pela média e ordenando as metodologias pelo *ranking*, é possível obter os seguintes resultados para as séries temporais não estacionárias apresentados na Tabela 3.2.

Tabela 3.2: Valor ordinal do *ranking* médio para as Séries não estacionárias

Dados	OOT	CV	ModCV	BCV	hvBCV	M-CV
Séries não Estacionárias	6	4	2	1	3	5

Capítulo 4

Conclusão

Nesse trabalho estudamos e comparamos diferentes metodologias de validação cruzada desenvolvidas para dados onde não é verdadeira a suposição de independência das obervações. Analisamos em específico como essas metodologias se comportam em diferentes séries temporais e a habilidade de estimar a perda dado um modelo preditivo usual para esse tipo de dados (ARIMA e Suavização Exponencial).

Agrupando as análises apresentadas no Capítulo 3 pela média e ordenando as metodologias pelo *ranking*, é possível obter os seguintes resultados apresentados na Tabela 4.1.

Tabela 4.1: Valores ordinais do *ranking* médio, resumo dos resultados obtidos

Dados	OOT	CV	ModCV	BCV	hvBCV	M-CV
Séries Estacionárias	6	4	3	1	2	5
Séries não Estacionárias	6	4	2	1	3	5
Total	6	4	2	1	3	5

Analisando a Tabela 4.1, é possível observar que a metodologia *Blocked cross-validation* foi a melhor para estimar a perda dado um modelo preditivo usual nas condições testadas no estudo, tanto para séries estacionárias quanto não estacionárias. A segunda melhor metodologia foi a *Modified cross-validation*, demonstrando que para esses casos de estudos apresentados metodologias menos complexas se saíram melhor que as demais, sendo elas simples modificações do método de validação cruzada *k-fold* comumente usando com modelos de aprendizado de máquina. No geral as metodologias desenvolvidas especificamente para séries temporais realmente se saíram melhores que metodologia de validação cruzada usual.

A metodologia apresentada por Jiang e Wang (2017) acabou tendo resultados parecidos da validação fora da amostra (OOS) também comumente utilizada, seria interessante realizar simulações de mais séries alterando seus devidos comportamentos a fim de testar mais essa metodologia.

Para estudos futuros, seria interessante realizar mais comparações com diferentes tipos de séries temporais, alterando mais o tamanho amostral dos dados e a ordem de autocorrelação da série. Além disso, pode-se testar o comportamento dessas metodologias em diferentes tipos de modelos preditivos, como por exemplo modelos de aprendizado de máquina.

Referências Bibliográficas

Bergmeir, C. e Benítez, J. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, **191**, 192–213.

Bergmeir, C. e Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, **191**, 192–213. Data Mining for Software Trustworthiness.

Bergmeir, C., Hyndman, R. J. e Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics Data Analysis*, **120**, 70–83.

Cerqueira, V., Torgo, L. e Soares, C. (2021). Model selection for time series forecasting: Empirical analysis of different estimators.

Hyndman, R. J. e Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, volume 3rd edition. OTexts: Melbourne, Australia.

Izbicki, R. e dos Santos, T. M. (2020). *Aprendizado de máquina: uma abordagem estatística*. ISBN 978-65-00-02410-4.

Jiang, G. e Wang, W. (2017). Markov cross-validation for time series model evaluations. *Information Sciences*, **375**, 219–233.

McQuarrie, A. D. R. e Tsai, C.-L. (1998). *Regression and time series model selection*. World Scientific.

Morettin, P. A. e Toloi, C. M. (2006). *Analise de Series Temporais*. ABE- Projeto Fisher e Editora Blucher.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics 99 (2000) 39-61*.

Snijders, T. A. B. (1988). On cross-validation for predictor evaluation in time series. Em T. K. Dijkstra, editor, *On Model Uncertainty and its Statistical Implications*, páginas 56–69, Berlin, Heidelberg. Springer Berlin Heidelberg. ISBN 978-3-642-61564-1.

Apêndice A

Códigos utilizados

Os códigos utilizados nesse trabalho estão disponíveis em

https://github.com/DanielSimionato/tcc