

UNIVERSIDADE FEDERAL DO PARANÁ

ANDERSON ANDOLFATO FILHO

AVALIAÇÃO DE MODELOS DE SÉRIES TEMPORAIS PARA PREVISÃO DE
DEMANDA EM CADEIAS DE SUPRIMENTOS AUTOMOTIVAS: UM ESTUDO
COMPARATIVO DE CAPACIDADE PREDITIVA

CURITIBA

2024

ANDERSON ANDOLFATO FILHO

AVALIAÇÃO DE MODELOS DE SÉRIES TEMPORAIS PARA PREVISÃO DE
DEMANDA EM CADEIAS DE SUPRIMENTOS AUTOMOTIVAS: UM ESTUDO
COMPARATIVO DE CAPACIDADE PREDITIVA

Trabalho apresentado como requisito parcial
para a obtenção do título de Mestre em Métodos
Numéricos pelo Programa de Pós Graduação
em Métodos Numéricos do Setor de Exatas da
Universidade Federal do Paraná.

Orientador: Prof^o José Roberto Frega, DSc

CURITIBA

2024

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Andolfato Filho, Anderson

Previsão de demanda de peças de reposição utilizando modelos híbridos e técnicas de aprendizagem de máquina: um estudo de capacidade preditiva / Anderson Andolfato Filho. – Curitiba, 2024.

1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Métodos Numéricos em Engenharia.

Orientador: José Roberto Frega

1. Aprendizado do computador. 2. Análise de séries temporais. I. Universidade Federal do Paraná. II. Programa de Pós-Graduação em Métodos Numéricos em Engenharia. III. Frega, José Roberto. IV . Título.

Bibliotecário: Leticia Priscila Azevedo de Sousa CRB-9/2029

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ANDERSON ANDOLFATO FILHO** intitulada: **Previsão de Demanda de Peças de Reposição Utilizando Modelos Híbridos e Técnicas de Aprendizagem de Máquina: Um Estudo de Capacidade Preditiva**, sob orientação do Prof. Dr. JOSÉ ROBERTO FREGA, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 30 de Setembro de 2024.

Assinatura Eletrônica

24/10/2024 14:48:39.0

JOSÉ ROBERTO FREGA

Presidente da Banca Examinadora

Assinatura Eletrônica

06/10/2024 18:52:52.0

CLAUDIMAR PEREIRA DA VEIGA

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ - PPGOLD)

Assinatura Eletrônica

17/10/2024 15:31:33.0

ANDERSON LUIZ ARA SOUZA

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

*“Não faças como se tu fosses viver dez mil anos.
A morte paira sobre ti.
Enquanto vives, enquanto está em teu poder, sê bom.”
(Marco Aurélio; Meditações IV,17.)*

RESUMO

O presente trabalho tem o intuito de explorar e identificar modelos eficazes entre várias técnicas, focado em séries temporais de demanda e seus vários tipos que podem ser observados, sendo estacionários ou não. Aplica-se técnicas paramétricas, semi paramétricas e não paramétricas, modelos tais como Suavização exponencial, Arimax, GLMNET, SVR, Cubist, Random Forest, e esta pesquisa procura não apenas comparar a eficácia dos modelos usando diversas métricas de erro, mas também entender as nuances entre as demandas classificadas como suaves, erráticas, e intermitentes, bem como a distribuição dos hiperparâmetros nos modelos escolhidos como melhores. A proposta reside em uma análise comparativa, aplicando o método MIMO — Multi Input Multi Output — uma técnica para modelos não sequenciais, proposta por Bontempi (2008), bem como o método clássico iterativo, avaliando a performance por tipo de demanda, estacionaridade e autocorrelação das séries. Ademais, faz uso de modelos híbridos com ARIMA e suavização exponencial modelando a parte linear das séries temporais, e os modelos de aprendizagem de máquina atuando nos resíduos dos modelos anteriores, sendo a atuação na parte não linear. Esta técnica híbrida melhora a acurácia da previsão, combinando a técnica iterativa de predição na parte linear, e a técnica MIMO na parte não linear, provendo uma perspectiva flexível para predição.

Palavras-chaves: Séries Temporais; Modelos Híbridos; Previsão; Aprendizagem de Máquina; MIMO.

ABSTRACT

The present study aims to explore and identify the most effective forecasting techniques and models for various demand types and time series, whether stationary or not. By applying parametric, semiparametric, and nonparametric models—such as Exponential Smoothing, ARIMAX, GLMNET, MARS, SVR, Croston, XGBoost, NNetAR, MLP, Cubist, and Random Forest—this research seeks not only to compare the efficacy of these models using multiple error metrics, but also to deeply understand the nuances of demands classified as smooth, intermittent, erratic, and irregular, alongside the hyperparameters selected for each model. The dissertation proposes a comparative analysis employing Multi-Input Multi-Output (MIMO) techniques for non-sequential models, as proposed by Bontempi (2008), and classic iterative approaches to evaluate performance by demand type and stationarity characteristics of time series. Additionally, the hybrid modeling approach integrates linear components handled by models like Exponential Smoothing and ARIMA, while the residuals, representing the nonlinear component, are addressed using machine learning techniques. This hybrid approach improves forecast accuracy by combining iterative predictions for the linear part and MIMO for the non-linear part, providing a flexible and integrated perspective for final predictions.

Key-words: Time Series; Hybrid Models; Forecasting; Machine Learning; MIMO.

TABELA 1 – Tabela Consolidada de Símbolos

Símbolo	Descrição	Capítulo
y_t	Valor real da série temporal no tempo t .	2, 4
\hat{y}_t	Valor previsto pela suavização exponencial ou ARIMA.	2, 4
ϕ	Parâmetro autoregressivo em modelos ARIMA.	2, 4
θ	Parâmetro de médias móveis em modelos ARIMA.	4
σ^2	Variância do ruído branco em ARIMA.	4
L_t	Componente de tendência (linear) em modelos Prophet.	4
X_i	Variável explicativa i no modelo.	3, 4
W	Pesos sinápticos entre os neurônios.	4
b	Bias (deslocamento) nos neurônios.	4
$f(x)$	Função de previsão ou transformação aplicada às variáveis de entrada.	3, 4
λ	Parâmetro de regularização.	3, 4
α	Parâmetro de suavização em métodos exponenciais.	3, 4
C	Parâmetro de custo (trade-off entre erro e margem).	4
ϵ	Insensibilidade ao erro (epsilon-tube).	4
η	Taxa de aprendizado em XGBoost.	4
$\gamma(k)$	Autocovariância no lag k .	2, 4
ρ	Coeficiente de correlação entre variáveis ou séries temporais.	2, 4
T	Número de árvores no modelo (XGBoost) ou tamanho total da série temporal.	3, 4
$h(x)$	Função de base (spline) em MARS.	4
β	Coeficientes associados às funções de base.	4
$F(x, t)$	Função de distribuição cumulativa para processos estocásticos.	2
m_1, m_2	Momentos de primeira e segunda ordem (média e variância).	2
$\rho_{ij}(\ell)$	Correlação cruzada no lag ℓ .	2
Γ_ℓ	Matriz de covariância cruzada no lag ℓ .	2
\mathbf{w}	Vetor de pesos em modelos lineares.	3, 4
\mathbf{x}	Vetor de variáveis explicativas.	3, 4
ξ	Variável de folga para margens suaves em SVR.	3, 4
$K(\mathbf{x}_i, \mathbf{x}_j)$	Kernel para medir similaridade em SVR.	3, 4
ν	Parâmetro de regularização em MARS.	3
ℓ	Função de perda utilizada para otimização.	3, 4
k	Número de divisões no método de validação K-fold.	4
\mathcal{D}	Conjunto de dados utilizado na validação cruzada.	4
$\mathcal{L}(y, \hat{y})$	Função de erro ou perda entre valor real y e valor previsto \hat{y} .	4
\mathbf{h}	Vetor de hiperparâmetros avaliados.	4
p	Parâmetro de probabilidade em Grid Latin Hypercube Sampling.	4

LISTA DE ILUSTRAÇÕES

FIGURA 1 – Exemplos de Demanda - Peças de reposição	18
FIGURA 2 – Tipos de Demanda	27
FIGURA 3 – Metodologia Utilizada	50
FIGURA 4 – Validação - K-fold - Exemplo	61
FIGURA 5 – Validação - Janela Expansiva - Exemplo	62
FIGURA 6 – Demanda - Intermitente	73
FIGURA 7 – Demanda - Errática	73
FIGURA 8 – Demanda - Suave	74
FIGURA 9 – Distribuição de zeros - Intermitente	76
FIGURA 10 – Densidade Conjunta - Arima SVR - Estacionariedade - Intermitente	77
FIGURA 11 – Densidade Conjunta - Arima SVR - Autocorrelação - Intermitente	78
FIGURA 12 – Densidade Conjunta - Glnet Poisson - Estacionariedade - Intermitente	81
FIGURA 13 – Densidade Conjunta - Glnet Poisson - Autocorrelação - Intermitente	82
FIGURA 14 – Densidade conjunta - SVR - Estacionariedade - Intermitente . . .	84
FIGURA 15 – Densidade conjunta - SVR - Autocorrelação - Intermitente	85
FIGURA 16 – Densidade conjunta - ArimaRF - Estacionariedade - Suaves . . .	89
FIGURA 17 – Densidade conjunta - ArimaRF - Autocorrelação - Suaves	89
FIGURA 18 – Densidade conjunta - ArimaXGboost - Estacionariedade - Suaves	92
FIGURA 19 – Densidade conjunta - ArimaXGboost - Autocorrelação - Suaves .	93
FIGURA 20 – Densidade conjunta - ArimaSVR - Estacionariedade - Suaves . .	95
FIGURA 21 – Densidade conjunta - ArimaSVR - Autocorrelação - Suaves . . .	96
FIGURA 22 – Densidade conjunta - Glnet Poisson - Estacionariedade - Erráticas	99
FIGURA 23 – Densidade conjunta - Glnet Poisson - Autocorrelação - Erráticas	100
FIGURA 24 – Densidade conjunta - XGboost - Estacionariedade - Erráticas . .	103
FIGURA 25 – Densidade conjunta - XGboost - Autocorrelação - Erráticas	103
FIGURA 26 – Densidade conjunta - SVR - Estacionariedade - Erráticas	106
FIGURA 27 – Densidade conjunta - SVR - Autocorrelação - Erráticas	107
FIGURA 28 – Log(Cost) - Estacionariedade - Intermitente	116
FIGURA 29 – Log(Cost) - Autocorrelação - Intermitente	116
FIGURA 30 – Log(RbfSigma) - Estacionariedade - Intermitente	117
FIGURA 31 – Log(RbfSigma) - Autocorrelação - Intermitente	117
FIGURA 32 – Mixture - Glnet Poisson - Estacionária - Intermitente	118
FIGURA 33 – Mixture - Glnet Poisson - Autocorrelação - Intermitente	118

FIGURA 34 – Penalty - Glmnet Poisson - Estacionária - Intermitente	119
FIGURA 35 – Penalty - Glmnet Poisson - Autocorrelação - Intermitente	119
FIGURA 36 – Log(Cost) - SVR - Estacionariedade - Intermitente	120
FIGURA 37 – Log(Cost) - SVR - Autocorrelação - Intermitente	120
FIGURA 38 – Log(RbfSigma) - SVR - Estacionariedade - Intermitente	121
FIGURA 39 – Log(RbfSigma) - SVR - Autocorrelação - Intermitente	121
FIGURA 40 – Minn - Arima RF - Autocorrelação - Suave	122
FIGURA 41 – Minn - Arima RF - Estacionariedade - Suave	122
FIGURA 42 – Árvores - Arima RF - Autocorrelação - Suave	123
FIGURA 43 – Árvores - Arima RF - Estacionariedade - Suave	123
FIGURA 44 – Minn - Arima XGb - Autocorrelação - Suave	124
FIGURA 45 – Minn - Arima XGb - Estacionariedade - Suave	124
FIGURA 46 – Árvores - Arima XGb - Autocorrelação - Suave	125
FIGURA 47 – Árvores - Arima XGb - Estacionariedade - Suave	125
FIGURA 48 – Cost - Arima SVR - Autocorrelação - Suave	126
FIGURA 49 – Cost - Arima SVR - Estacionariedade - Suave	126
FIGURA 50 – RBF - Arima SVR - Autocorrelação - Suave	127
FIGURA 51 – RBF - Arima SVR - Estacionariedade - Suave	127
FIGURA 52 – Mixture - Glmnet Poisson - Autocorrelação - Errático	128
FIGURA 53 – Mixture - Poisson - Estacionariedade - Errático	128
FIGURA 54 – Penalty - Glmnet Poisson - Autocorrelação - Errático	129
FIGURA 55 – Penalty - Glmnet Poisson - Estacionariedade - Errático	129
FIGURA 56 – Minn - Xgboost - Autocorrelação - Errático	130
FIGURA 57 – Minn - Xgboost - Estacionariedade - Errático	130
FIGURA 58 – Árvores - Xgboost - Autocorrelação - Errático	131
FIGURA 59 – Árvores - Xgboost - Estacionariedade - Errático	131
FIGURA 60 – Cost - SVR - Autocorrelação - Errático	132
FIGURA 61 – Cost - SVR - Estacionariedade - Errático	132
FIGURA 62 – RBFSigma - SVR - Autocorrelação - Errático	133
FIGURA 63 – RBFSigma - SVR - Estacionariedade - Errático	133

LISTA DE TABELAS

TABELA 1 – Tabela Consolidada de Símbolos	7
TABELA 2 – Sumário dos tipos de relações entre séries temporais.	25
TABELA 3 – Caracterização da Demanda	26
TABELA 4 – Exemplos de Modelos Sequenciais	31
TABELA 5 – Engenharia de variáveis para previsão de y_t	72
TABELA 6 – Descritivo das Peças	
Resumo sobre características da série temporal por tipo de demanda	74
TABELA 7 – Rank Médio - Intermitentes	75
TABELA 8 – Modelo Campeão - Intermitentes	75
TABELA 9 – Estrutura Temporal - Intermitentes	76
TABELA 10 – Parâmetros SVR - Autocorrelação - Intermitente	79
TABELA 11 – Parâmetros - Estacionariedade - Intermitente	80
TABELA 12 – Resumo - Glmnet Poisson - Intermitente	83
TABELA 13 – Parâmetros SVR - MIMO - Autocorrelação	85
TABELA 14 – Parâmetros SVR - MIMO - Estacionariedade	86
TABELA 15 – Ranking dos Modelos	87
TABELA 16 – Ranking dos Modelos - Contagem	88
TABELA 17 – Estrutura Temporal - Suaves	88
TABELA 18 – Parâmetros RF - Autocorrelação - Suaves	90
TABELA 19 – Parâmetros RF - Estacionariedade - Suaves	91
TABELA 20 – Parâmetros XGboost - Autocorrelação - Suaves	93
TABELA 21 – Parâmetros XGboost - Estacionariedade - Suaves	94
TABELA 22 – Parâmetros ArimaSVR - Autocorrelação - Suave	96
TABELA 23 – Parâmetros ArimaSVR - Estacionariedade - Suave	97
TABELA 24 – Ranking dos Modelos	98
TABELA 25 – Rank - Contagem	98
TABELA 26 – Estrutura Temporal - Suaves	99
TABELA 27 – Parâmetros Glmnet Poisson - Autocorrelação - Erráticas	101
TABELA 28 – Parâmetros Glmnet Poisson - Estacionariedade - Erráticas	101
TABELA 29 – Parâmetros XGboost - Autocorrelação - Erráticas	104
TABELA 30 – Parâmetros XGboost - Estacionariedade - Erráticas	105
TABELA 31 – Parâmetros SVR - Autocorrelação - Errática	107
TABELA 32 – Parâmetros de Cost e RBF - Estacionariedade - Séries Temporais	108

TABELA 33 – Resumo dos parâmetros SVR para séries intermitentes, erráticas e suaves.	109
TABELA 34 – Resumo dos efeitos da estacionariedade e autocorrelação nos parâmetros SVR.	110
TABELA 35 – Resumo dos efeitos em parâmetros do Glnet Poisson para séries intermitentes e erráticas.	110
TABELA 36 – Resumo dos efeitos da autocorrelação e estacionariedade nos modelos baseados em árvores.	111

SUMÁRIO

1	INTRODUÇÃO	15
1.1	ESTRUTURA DO TRABALHO	16
2	SÉRIES TEMPORAIS	18
2.1	EXEMPLOS DE SÉRIES TEMPORAIS	18
2.2	SÉRIES TEMPORAIS E CADEIAS DE SUPRIMENTOS	19
2.3	PROCESSOS ESTOCÁSTICOS	20
2.4	FUNÇÕES DE AUTOCOVARIANÇA E AUTOCORRELAÇÃO	22
2.4.1	Autocorrelação Parcial	22
2.4.2	Correlação Cruzada	23
2.5	COMPORTAMENTO DA SÉRIES TEMPORAIS - SUPPLY CHAIN	25
3	MODELOS DE APRENDIZAGEM EM SÉRIES TEMPORAIS	30
3.1	MODELOS SEQUENCIAIS	31
3.1.1	Suavização Exponencial	31
3.1.2	Auto Regressivo, Integrado de Médias Móveis	33
3.1.3	NNAR	34
3.1.4	Considerações Finais - Modelos Sequenciais	35
3.2	MODELOS NÃO SEQUENCIAIS	36
3.2.1	Modelos Lineares Generalizados - Séries temporais	37
3.2.2	MLP	38
3.2.3	GLMnet	39
3.2.4	SVR - Support Vector Regression	41
3.2.5	XGboost	43
3.2.6	Random Forest	44
3.2.7	Cubist	46
3.2.8	MARS	47
3.3	MODELOS HÍBRIDOS	49
3.4	MÉTODOS DE PREDIÇÃO	52
3.4.1	Método Iterativo	52
3.4.2	Métodos Diretos	53
3.4.3	Método MIMO	54
4	MÉTODOS DE VALIDAÇÃO CRUZADA	58
4.1	K-FOLD	58
4.2	JANELA EXPANSIVA	61
4.3	GRID LATIN HYPERCUBE	62
4.4	MÉTRICAS DE ERRO	66
5	RESULTADOS	69
5.1	TÉCNICAS DE ENGENHARIA DE VARIÁVEIS	70
5.2	TRANSFORMAÇÕES DE ESCALA:	70
5.3	CRIAÇÃO DE LAGS:	71
5.4	ESTATÍSTICAS DE JANELA:	71

5.5	INTERMITENTES	74
5.5.1	Arima - SVR	76
5.5.1.1	Parâmetros SVR - Híbrido	78
5.5.1.2	Comentários Finais - ArimaSVR	80
5.5.2	Glmnet - Poisson	81
5.5.2.1	Parâmetros Glmnet - MIMO	82
5.5.2.2	Comentários Finais - GLMNET Poisson	83
5.5.3	Support Vector Regression	84
5.5.3.1	Parâmetros SVR - MIMO	85
5.5.3.2	Comentários Finais - SVR	87
5.6	SÉRIES TEMPORAIS - SUAVES	87
5.6.1	Arima - Random Forest	88
5.6.1.1	Parâmetros RF - Híbrido	90
5.6.1.2	Comentários Finais - ArimaRF	91
5.6.2	Arima - XGBoost	92
5.6.2.1	Parâmetros XGboost - Híbrido	93
5.6.2.2	Comentários Finais - ArimaXGboost	94
5.6.3	Arima - SVR	95
5.6.3.1	Parâmetros SVR - Híbrido	96
5.6.3.2	Comentários Finais - ArimaSVR	97
5.7	SÉRIES TEMPORAIS - ERRÁTICAS	98
5.7.1	Glmnet - Poisson	99
5.7.1.1	Parâmetros Glmnet Poisson - MIMO	101
5.7.1.2	Comentários Finais - Glmnet Poisson	102
5.7.2	Arima - XGBoost	102
5.7.2.1	Parâmetros XGboost - Híbrido	104
5.7.2.2	Comentários Finais - ArimaXGboost	105
5.7.3	Arima - SVR	106
5.7.3.1	Parâmetros SVR - Híbrido	107
5.7.3.2	Comentários Finais - Arima SVR	109
5.8	RESULTADOS CONSOLIDADOS - MODELOS	109
6	CONCLUSÃO	113
A	APÊNDICE A - BOXPLOTS E DENSIDADES	116
A.1	INTERMITENTES	116
A.1.1	Arima SVR	116
A.1.2	Glmnet Poisson	118
A.1.3	SVR	120
A.2	SUAVES	122
A.2.1	Arima Random Forest	122
A.2.2	Arima XGboost	124

A.2.3	Arima SVR	126
A.3	ERRÁTICOS	128
A.3.1	Glmnet Poisson	128
A.3.2	Arima XGboost	130
A.3.3	Arima SVR	132
	REFERÊNCIAS	134

1 INTRODUÇÃO

Na era em que a sociedade está progressivamente se digitalizando e grandes volumes de dados têm sido gerados, a capacidade preditiva na previsão de séries temporais tornou-se um ponto crucial para a otimização das cadeias de suprimento.

As cadeias de suprimentos desempenham um papel vital tanto na economia nacional e global quanto, numa perspectiva microeconômica, dentro das próprias firmas. McAfee et al. (2012) demonstraram que empreendimentos com cadeias de suprimentos *data-driven*¹ tendem a ser mais lucrativos que seus competidores. Corroborando esse resultado, Perrey et al. (2013) apontam que as decisões baseadas em dados incrementam o retorno sobre o investimento (ROI). Em uma perspectiva de sustentabilidade, tanto para a firma quanto para a economia em geral, Cheng et al. (2021) notaram uma influência positiva do uso de análises estatísticas em dados bem estruturados sobre a economia circular², promovendo benefícios de longo prazo tanto para o meio ambiente quanto para a indústria.

No contexto das cadeias produtivas, Wang et al. (2016) assinalam que as análises preditivas têm sido cada vez mais adotadas em todos os níveis — sejam eles operacionais, táticos ou estratégicos — para auxiliar na tomada de decisões. Essas análises utilizam dados históricos para antecipar padrões e prever futuros resultados e atividades da cadeia de suprimentos. Riahi et al. (2021), em um estudo bibliométrico sobre o uso de aprendizado de máquina em cadeias produtivas, valendo-se do banco de dados Scopus³, avaliaram 136 estudos, indicando o uso de aprendizado supervisionado, como o *Support Vector Machine*, aprendizado não supervisionado, por meio de algoritmos de clusterização, e aprendizado por reforço. Ademais, os autores identificaram um aumento gradual desde 2010 nos estudos de cadeia de suprimentos e o uso de técnicas estatísticas e de aprendizado de máquina. Sharma et al. (2020), em um estudo de revisão sobre *supply chain* na agricultura, identificaram uma vasta gama de aplicações para esses métodos estatísticos e de aprendizado de máquina, incluindo a predição de safras, a determinação de propriedades do solo, previsão climática, previsão de demanda e de transporte, dentre outras etapas da cadeia produtiva. Os autores concluem sua revisão de 93 artigos apontando que existem diversas aplicações do uso do aprendizado de máquina em diferentes fases da cadeia produtiva, o que

¹ Decisões tomadas com base em análise de dados, direcionadas aos dados.

² A economia circular é um modelo econômico e de negócios que se baseia na ideia de eliminar o conceito de "fim de vida útil" dos produtos e materiais. Em vez de seguir um modelo linear de "produzir, usar e descartar", a economia circular promove a reutilização, reciclagem e regeneração de recursos para prolongar seu ciclo de vida e reduzir o desperdício.

³ Uma das maiores e mais abrangentes bases de dados bibliográficas e de resumos do mundo, que abrange uma ampla variedade de disciplinas acadêmicas.

indica um melhor retorno sobre o investimento e a minimização de perdas.

Diante deste contexto, o presente estudo tem como objetivo explorar e identificar as técnicas e modelos de previsão mais eficazes para diferentes tipos de demanda e para séries temporais, sejam elas estacionárias ou não. Por meio da aplicação de modelos paramétricos, semi-paramétricos e não paramétricos, como Suavização Exponencial, ARIMAX, GLMNET, MARS, SVR, Croston, XGBoost, NNetAR e MLP, Cubist e Random Forest, esta pesquisa busca não apenas comparar a eficácia destes modelos utilizando várias métricas de erro, mas também compreender em profundidade as nuances das demandas classificadas em suaves, intermitentes, erráticas e irregulares e os hiperparâmetros escolhidos em cada modelo.

Esta dissertação propõe uma análise comparativa, empregando técnicas de previsão Mult Input Multi Output — para os modelos não sequenciais —, propostas por Bontempi (2008) e clássicos iterativos para avaliar seu desempenho por tipo de demanda e características de estacionariedade das séries temporais. Ao fazer isso, tem a intenção de contribuir com uma meta-análise para a literatura existente, fornecendo *insights* práticos e teóricos para a gestão avançada de cadeias de suprimentos através de técnicas de previsão de séries temporais. Além disso, faz uso de uma abordagem híbrida na modelagem, na qual a componente linear das séries é tratada com modelos como Suavização Exponencial e ARIMA, enquanto os resíduos — representando a componente não linear — são abordados com técnicas de aprendizagem de máquina. Esta abordagem híbrida promete não apenas melhorar a precisão das previsões por meio da combinação de previsões iterativas para a parte linear e MIMO para a parte não linear, mas também fornecer uma perspectiva integrada e flexível para a previsão final.

Ao integrar modelos híbridos na análise, espera-se oferecer contribuições à literatura existente, proporcionando um estudo empírico sobre os modelos utilizados e os melhores hiperparâmetros e o âmbito da previsão estatística e a gestão avançada de cadeias de suprimentos. Este enfoque sublinha a importância de estratégias de previsão adaptativas e multifacetadas, refletindo a complexidade e a dinâmica das séries temporais modernas em cadeias de suprimentos.

1.1 ESTRUTURA DO TRABALHO

O trabalho aqui apresentado está organizado da seguinte forma:

- **Revisão da literatura e Conceitos:** Inicialmente, tratamos de conceitos teóricos como estacionariedade, autocovariância, autocorrelação, ruído branco, passeio aleatório e o comportamento de séries temporais dentro das cadeias produtivas, bem como as métricas de erro utilizadas.

- **Séries Temporais e Usos para Predição:** É feita uma revisão teórica sobre os modelos de séries temporais clássicos como Suavização exponencial e Arima, processos auto-regressivos e de médias móveis, até modelos dinâmicos auto-regressivos e para dados de contagem. Após isso, segue com a revisão de modelos não sequenciais e de aprendizagem de máquina como SVR e XGboost, por exemplo, indicando a mecânica por trás destas técnicas e referências de utilização destas técnicas no âmbito da predição de séries temporais em cadeias de suprimentos e também em outras áreas do saber.
- **Métodos de Predição:** Os métodos de predição serão discutidos em termos de abordagens iterativas, diretas e MIMO (Multiple Input, Multiple Output). Serão exploradas as vantagens e limitações de cada método, fornecendo insights importantes para a seleção adequada de técnicas de predição em diferentes contextos.
- **Modelos Híbridos:** Neste capítulo serão apresentados os modelos híbridos que combinam diferentes técnicas de predição para melhorar a acurácia e robustez das previsões. Serão discutidas as estratégias de integração e os benefícios decorrentes da abordagem híbrida.
- **Validação Cruzada:** Nesta seção de Validação Cruzada, serão abordadas as técnicas de avaliação de desempenho dos modelos de predição, com foco nas abordagens K-fold e Janela Expansiva. Será discutida a importância da validação cruzada para garantir a confiabilidade e generalização das previsões.
- **Engenharia de variáveis:** Nesta seção será abordada algumas referências históricas e técnicas no âmbito das transformações de variáveis para previsão da série temporal, comentando sobre trabalhos pioneiros dos anos 70 até trabalhos mais recentes utilizando tais técnicas em modelos de aprendizagem de máquina.
- **Resultados:** Por fim, são apresentados os resultados por tipo de demanda, ou seja, pela classificação da série temporal em três tipos distintos, indicando os modelos campeões em cada cenário, a distribuição de seus hiperparâmetros e a distribuição de degradação da métrica de erro quando se comparando um horizonte de previsão de 6 meses.

2 SÉRIES TEMPORAIS

Por definição, as séries temporais são dados que se sucedem no indexador tempo, onde a ordem tem um papel importantíssimo (SLUTSKY, 1937). Ou seja, são observações feitas em pontos ou períodos de tempo sucessivos, num delimitado intervalo de tempo. Em diversos âmbitos do estudo da estatística temos a tarefa de trabalhar com dados observados no tempo, por exemplo: o mercado acionário, indicadores macroeconômicos, número de nascimentos, demanda por peças de reposição, sendo o último o objeto de estudo deste trabalho.

Ademais, fazendo uso de apenas uma amostra de tamanho T , $\{y_1, y_2, \dots, y_T\}$, e a série temporal $\{Y_t\}_{t=-\infty}^{\infty}$ é identificada pela descrição do seu t -ésimo elemento. Sendo assim, cada elemento Y_t representará uma observação em um ponto específico no tempo t . A notação $t = -\infty$ indica que a série temporal se estende infinitamente no passado, enquanto $t = \infty$ indica que se estende infinitamente no futuro; em suma, cada Y_t representa uma observação em um instante de tempo específico.

2.1 EXEMPLOS DE SÉRIES TEMPORAIS

No exemplo a seguir temos o exemplo com as próprias séries de demanda aqui estudadas.

Algumas apresentam tendência, seja de alta ou de baixa, outras são intermitentes e outras inflacionadas de zeros, conforme ilustrado na figura 1

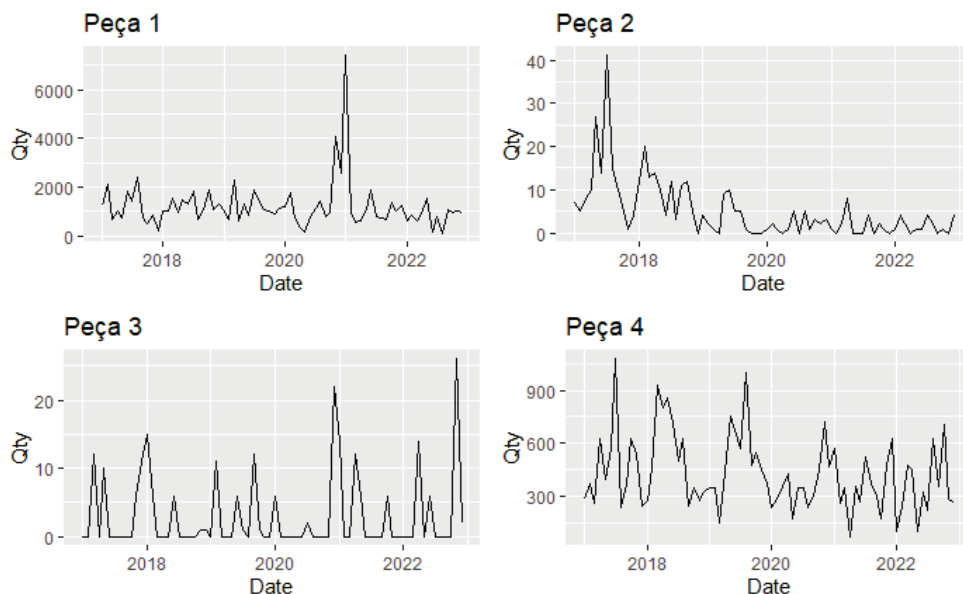


FIGURA 1 – Exemplos de Demanda - Peças de reposição

Fonte: Elaboração própria

Os gráficos 1 têm como eixo y as vendas mensais em quantidades e o eixo x a data em meses. Percebe-se que as séries temporais aqui estudadas têm a mesma periodicidade - mensal, contudo, de comportamentos distintos, algumas tendo diversos zeros e outras mantendo um maior nível de vendas. Com ou sem tendência, demandas intermitentes e demandas mais próximas da estacionariedade.

2.2 SÉRIES TEMPORAIS E CADEIAS DE SUPRIMENTOS

Em quase todos os campos das ciências, os dados são medidos com base no tempo. Ademais, os assuntos tratados nesta dissertação, sendo as séries temporais de contagem e o estudo da demanda de cadeias de suprimentos, têm sido objeto de estudo nos últimos anos.

Quanto ao estudo de cadeias de suprimentos e séries temporais, no artigo *Forecasting spare part demand with installed base information: A review*, Van der Auweraer et al. (2019), foi feita uma revisão de artigos que concernem a predição de séries temporais em supply chain, foram avaliados 5 livros e 38 artigos de diversos periódicos, e os autores indicam que quase todos os estudos consideram o histórico de vendas e o histórico de descartes destes produtos, sem antecipar o potencial futuro percentual de mudança no nível da série no tempo. Ademais, que uma pequena parte dos estudos de *forecasting* relacionados foca em relações causais às demandas, e que seria interessante utilizar diversas covariáveis que sejam os balizadores da demanda e não apenas os métodos de dissecação das séries temporais e que há fortes indícios de grande potencial da combinação de técnicas de *forecasting*, ou seja, *ensembled*.

O padrão da demanda entre as peças é a primeira característica, e provavelmente a mais importante. Enquanto algumas peças têm demandas mais comportadas, estáveis, outras peças têm demandas intermitentes, com sequências de zeros seguidas, gerando uma demanda errática denotada como *lumpy*, de acordo com Boylan e Syntetos (2010). De forma que essas intermitências geram caudas de distribuição mais pesadas nos dados, resultando em dados não normais que são assunções de diversos modelos; configurando esta, um dos desafios do procedimento de *forecasting*.

Vanderput (2017), em seu livro, *Data Science for Supply chain Forecasting* demonstra um apanhado de técnicas, medidas de erro, clusterização e demais resultados pertinentes a modelagem estatística e seus resultados ao caso de *supply chain*; fazendo uso de técnicas paramétricas e não paramétricas, bem como modelos para dados contínuos nesta atividade que demonstra ter dados discretos.

Ademais, temos um ponto da discussão, que concerne à seleção automática de modelos, uma vez que se pode observar uma gama muito grande de séries temporais - produtos, diferentes a serem modeladas no âmbito do *supply chain*, em especial das

peças de reposição de veículos.

Villegas e Pedregal (2021), demonstram, entre outros temas, a seleção de modelos de forma automática devido à grande quantidade de produtos que devem ser estimados; ademais, os autores indicam que essa seleção automática tem recebido grande atenção no meio da predição de séries temporais. No estudo em questão, foram avaliadas 166 séries temporais e utilizou-se a métrica AvgRelMAE, como visto mais adiante na seção 4.4, para qualificar os modelos; os resultados indicaram que, apesar dos modelos de componentes não observados terem melhor qualificação, todos apresentam resultados similares, os modelos ARIMA e ETS têm performance bastante similar, dado que o conjunto de dados utilizado não tem praticamente *outliers* e, como esperado, quanto maior for o horizonte de predição, pior fica a predição.

2.3 PROCESSOS ESTOCÁSTICOS

Os processos estocásticos, por definição, são séries de variáveis aleatórias ordenadas por um indexador, neste caso sendo o suporte a cronologia dos dados em séries temporais.

Seja um processo estocástico discreto no tempo, sendo o tempo $\{0, 1, 2, 3, 4\}$ e a variável aleatória denotada por X_0, X_1, X_2, \dots , seguindo a notação $\{X_n : n = 0, 1, 2, \dots\}$. Para o caso contínuo, o processo estocástico $\{Y_t : t \geq 0\}$, com o índice contínuo $\{t \geq 0\}$. Os índices n e t são frequentemente referidos como o tempo, portanto, X_n é um processo discreto no tempo e Y_t um processo contínuo no tempo.

Considere um processo estocástico no tempo discreto, para cada t , X_t será uma variável aleatória que seguirá a função distribuição - sendo a descrição de 1ª ordem do processo :

$$F(x, t) = P(X_t \leq x), -\infty \leq x \leq \infty$$

Entretanto, para dados como os de séries temporais, que se ordenam subsequentemente um ao outro, e não são independentes uns dos outros, como nos casos cronológicos, tem-se o interesse em casos mais elevados dos processos. Sendo sua descrição para qualquer X_{t1}, X_{t2} :

Entretanto, para dados de séries temporais, que se ordenam sequencialmente e exibem dependências temporais, como em registros cronológicos, há interesse em analisar processos estocásticos de ordem superior. Isso inclui a consideração de dependências complexas entre observações em diferentes momentos, modeladas por processos como X_{t1}, X_{t2} que captam essas dinâmicas temporais.

$$F(x_1, x_2; t_1, t_2) = P(X_{t_1} \leq x_1; X_{t_2} \leq x_2), -\infty \leq x \leq \infty$$

Permitindo-se estabelecer, arbitrariamente no tempo, a autocovariância e correlação entre as variáveis aleatórias X ; as estatísticas mencionadas são muito importantes na análise e descrição de processos temporais, demonstrando a independência do processo. O próximo passo é dado com os seguintes parâmetros de média e variância - e seus momentos, onde m_1 e m_2 são o primeiro e segundo momento, respectivamente:

$$E[X(t)] = \mu \Rightarrow m_1 = \bar{X} = E[X] \quad (2.1)$$

$$m_2 = VAR[X(t)] = \sigma_x^2 = E[(X - \bar{X})^2] \quad (2.2)$$

E sua autocovariância (σ_{xx}) e autocorrelação (ρ_{mn}) serão :

$$\sigma_{xx} = E[(X_n - \mu_x)(X_{n+m} - \mu_x)]$$

$$\rho_{mn} = E[(X_n - \mu_x)(X_{n+m} - \mu_x)]/\sigma_x^2$$

Türkmen et al. (2021) em estudo probabilístico relacionando tipos de demanda, introduzindo a demanda intermitente como um processo de renovação discreta, é feita a argumentação que os tempos entre as chegadas da demanda são independentes e identicamente distribuídos; os autores sugerem que a utilização de distribuições mais flexíveis pode melhorar a previsão das demandas intermitentes e erráticas, conectando com técnicas como as redes neurais, para capturar as características estocásticas intrínsecas destes dois tipos de demanda. Outros artigos também fazem esta discussão entre tipos de demanda e características estocásticas, como em Dolgui et al. (2005), o artigo aborda a demanda intermitente com foco em padrões quase-periódicos, utilizando distribuições beta-binomial generalizadas e inferência bayesiana — e incorporando incerteza — para modelar padrões de demanda, sendo métodos para capturar a natureza estocástica deste tipo de demanda. Em Sandmann e Bober (2010), os autores fazem críticas a modelos determinísticos para demandas intermitentes, advogando a favor de ser um processo estocástico e necessitar de técnicas de mesma ordem; utilizam-se de estatísticas sumárias e estruturas de correlação para identificar as propriedades estocásticas dos dados e também de distribuições geométricas para

os tempos entre as demandas. Sendo abordagens que enfatizam o uso de técnicas estocásticas em séries temporais de grande incerteza e padrões irregulares.

Bontempi (2008) indica que o método MIMO captura as dependências estocásticas entre os valores futuros da série temporal, quando os métodos iterativos podem sofrer com a propagação do erro. A discussão indica que prever múltiplos valores num horizonte h qualquer pode ser mais precisa do que prever um valor de forma subsequente de cada vez, uma vez que as dependências entre os valores futuros são consideradas. Através desta técnica, com intuito de capturar as dependências estocásticas, resultando em uma estimativa da distribuição conjunta $p(Y|X)$, onde $Y = \phi_{t+1}, \phi_{t+2}, \dots, \phi_{t+H}$, sendo que essa complexidade é abordada com uma modelagem direta entre os diversos valores futuros.

2.4 FUNÇÕES DE AUTOCOVARIÂNCIA E AUTOCORRELAÇÃO

Estudando as sucessões cronológicas discretas, as funções de autocovariância e autocorrelação têm um papel preponderante, conforme as próximas elucidações do tema.

Se $\{X_n\}$ é um processo estacionário - estacionariedade "fraca", terá sua média e variância definidas pelas equações 2.1 e 2.2, então a covariância de X_n e X_{n+k} dependerá apenas do operador de retardo k , o que acarreta a seguinte definição de autocovariância :

$$\gamma(k) = cov(X_{n+k}, X_n)$$

Para cada k , a função de autocovariância mede a força com que os pares de valores covariam separados por um intervalo de tempo de amplitude k , denominado *lag*,

A função de autocorrelação (FAC) é definida pela seguinte expressão:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

A expressão anterior, simplesmente indica a correlação entre X_n e X_{n+k} ; outra interpretação de $\rho(k)$ seria o valor ótimo de X_n para a função de preditor de X_{n+k} , tal valor, α que minimiza $E(X_{n+k} - \alpha X_n)^2$,

2.4.1 Autocorrelação Parcial

A função de autocorrelação parcial é formalmente definida para uma série temporal estacionária com média zero, seja X_{t+k} para $k \geq 2$, e que é representada pela

regressão de X_{t+k} em $\{X_{t+k-1}, X_{t+k-2}, \dots, X_{t+1}\}$, que pode ser escrita como:

$$X_{t+k} = \phi_1 X_{t+k-1} + \phi_2 X_{t+k-2} + \dots + \phi_{k-1} X_{t+1}$$

O intercepto não foi incluído, dado que o exemplo se referia a uma série de média zero. Ademais, o que o processo de autocorrelação parcial denota é a correlação entre X_t e X_{t+h} , depois de todos os efeitos intermediários produzidos pelas variáveis que se encontram entre t e $t+k$. A função de correlação parcial dada por ϕ_{kk} é encontrada pela seguinte solução, Madsen (2007):

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \quad (2.3)$$

onde ϕ_{hh} é a autocorrelação parcial no lag k , e a matriz 2.3 pode ser escrita como :

$$P_k \phi_k = \rho_k$$

Ademais, ϕ_{kk} :

$$\phi_{kk} = \frac{\det \mathbf{V}_k}{\det \mathbf{P}_k},$$

Onde \mathbf{V}_k é obtida substituindo a k -ésima coluna de \mathbf{P}_k por ρ_k .

2.4.2 Correlação Cruzada

Uma vez que a análise será feita utilizando-se de método MIMO, o tema da correlação cruzada é bastante pujante para a análise; uma vez que é importante considerar as relações entre múltiplas séries temporais ou múltiplos lags e variáveis utilizadas na previsão de y .

Para séries temporais multivariadas, consideramos r_{it} como uma coleção de k séries temporais. Podemos definir as matrizes de correlação cruzada que medem a dependência linear entre essas séries em diferentes defasagens (lags).

Matriz de Correlação Cruzada de Lag Zero Tsay (2013):

$$\rho_0 \equiv [\rho_{ij}(0)] = D^{-1}\Gamma_0 D^{-1}$$

onde D é uma matriz diagonal contendo os desvios padrão das séries individuais r_{it} e Γ_0 é a matriz de covariância de lag zero.

Matriz de Correlação Cruzada de Lag ℓ :

$$\rho_\ell \equiv [\rho_{ij}(\ell)] = D^{-1}\Gamma_\ell D^{-1}$$

onde Γ_ℓ é a matriz de covariância cruzada de lag ℓ .

A função de autocovariância $\gamma(k)$ é um caso especial da covariância cruzada, onde a série temporal $\{X_n\}$ é correlacionada consigo mesma. Podemos ver isso na definição da matriz de covariância cruzada Γ_ℓ , onde a diagonal de Γ_0 representa a autocovariância de cada série temporal individual. Portanto, a função de autocovariância definida anteriormente pode ser vista como:

$$\gamma(k) = \Gamma_{ii}(k)$$

para a série X_i em lag k .

Relacionando Autocorrelação com Correlação Cruzada. De maneira similar, a função de autocorrelação $\rho(k)$ é um caso especial da correlação cruzada. A autocorrelação é a correlação da série consigo mesma em diferentes lags:

$$\rho(k) = \frac{\gamma(0)}{\gamma(k)} = \frac{\Gamma_{ii}(0)}{\Gamma_{ii}(k)}$$

Tsay (2013) indica que as matrizes de correlação cruzada de processos fracamente estacionários, têm-se fora da diagonal principal, em $l = 0$, as medidas de relação linear entre r_{it} e r_{jt} no mesmo momento temporal e para o caso $l > 0$ indicando se r_{it} é correlacionado com valores passados de r_{jt} . Aqui, as variáveis podem não ter uma relação linear, podem ter correlação concorrente ou unidirecional e pode existir a ausência da relação *lead-lag* e ainda existir uma relação de *Feedback*. No contexto MIMO, as matrizes de correlação cruzada ajudam a modelar as influências entre as múltiplas entradas e saídas, no intuito de capturar as relações dinâmicas da série temporal.

Tipo de Relação	Condição	Interpretação
Sem Relação Linear	$\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ para todos $\ell \geq 0$	As séries r_i e r_j não têm relação linear.
Correlação Concorrente	$\rho_{ij}(0) \neq 0$	As séries r_i e r_j estão correlacionadas no mesmo momento temporal.
Sem Relação de Lead-Lag	$\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ para todos $\ell > 0$	As séries r_i e r_j não têm relação de lead-lag.
Relação Unidirecional	$\rho_{ij}(\ell) = 0$ para todos $\ell > 0$, mas $\rho_{ji}(v) \neq 0$ para algum $v > 0$	Há uma relação unidirecional de r_i para r_j .
Relação de Feedback	$\rho_{ij}(\ell) \neq 0$ para algum $\ell > 0$, e $\rho_{ji}(v) \neq 0$ para algum $v > 0$	Há uma relação de feedback entre r_i e r_j .

TABELA 2 – Sumário dos tipos de relações entre séries temporais.

2.5 COMPORTAMENTO DA SÉRIES TEMPORAIS - SUPPLY CHAIN

As séries temporais abordadas neste estudo foram classificadas em 4 conceitos: Suave, errática, intermitente e irregular; sendo que a última tem uma característica de inflação de zeros.

Ao longo da evolução da análise de dados, diversos autores discutiram métodos de caracterizar demandas. Croston (1972) discute sobre demandas intermitentes, sua variação do tamanho da demanda e aleatoriedade e esporadicidade destas séries - ademais, o autor identificou que modelos de suavização exponencial geram grandes erros, causando estoques desnecessários. Williams (1984) considera a variância do número de pedidos e seu tamanho dado um tempo de entrega específico, classificando os itens em cinco categorias em relação à esporadicidade de demanda alta/baixa e tamanho.

Eaves e Kingsman (2004) retomam a discussão feita por Williams (1984), e reclassificam as peças em cinco categorias. A categorização é dada por frequência, quantidade, *lead-time*¹ e categoria; sendo as categorias: suave, irregular, baixo giro, esporádica leve e esporádica forte.

Johnston e Boylan (1996) introduzem a média de intervalo da demanda (ADI) que posteriormente foi complementada por Syntetos et al. (2005), conforme a seguir:

¹ lead-time é o tempo que cada produto, ou peça de reposição demora para sair do fornecedor e chegar ao destino, ou centro de distribuição, podendo ser medido em semanas ou dias ou meses,

$$P = \frac{\text{Períodos totais}}{\text{Intervalos totais de demanda}}$$

$$CV = \frac{\text{Desvio Padrão da Demanda}}{\text{Demanda média}}$$

O CV^2 é utilizado — onde o coeficiente de variação é utilizado com exponencial 2 — por Syntetos et al. (2005), ambos conceitos permitem que seja categorizada a demanda em quatro quadrantes, suaves (*smooth*), erráticas (*erratic*), intermitentes (*intermittent*) e irregulares (*lumpy*). Sendo que as de características suave e erráticas têm menor variabilidade quando comparadas com as demais, contudo a errática tem maior variabilidade que a suave. Já as intermitentes e irregulares têm maior variabilidade e demandas oscilantes na ótica temporal; bem como entre as erráticas e suaves, as irregulares têm maior variabilidade no tamanho da demanda quando comparadas com as intermitentes.

Ressalto que neste trabalho não serão tratadas as séries irregulares, uma vez que o custo computacional é alto para uma série com grande quantidade de zeros.

As demandas das peças de reposição para este trabalho foram categorizadas conforme Syntetos et al. (2005), a categorização se dá conforme os critérios indicados no artigo referenciado. Tais critérios seguem conforme a tabela a seguir:

TABELA 3 – Caracterização da Demanda

P	CV^2	Tipo de Demanda
< 1,32	< 0,49	Suave
\geq 1,32	< 0,49	Intermitente
< 1,32	> 0,49	Errática
\geq 1,32	\geq 0,49	Irregular

A seguir, um exemplo dos visuais dos quadrantes :

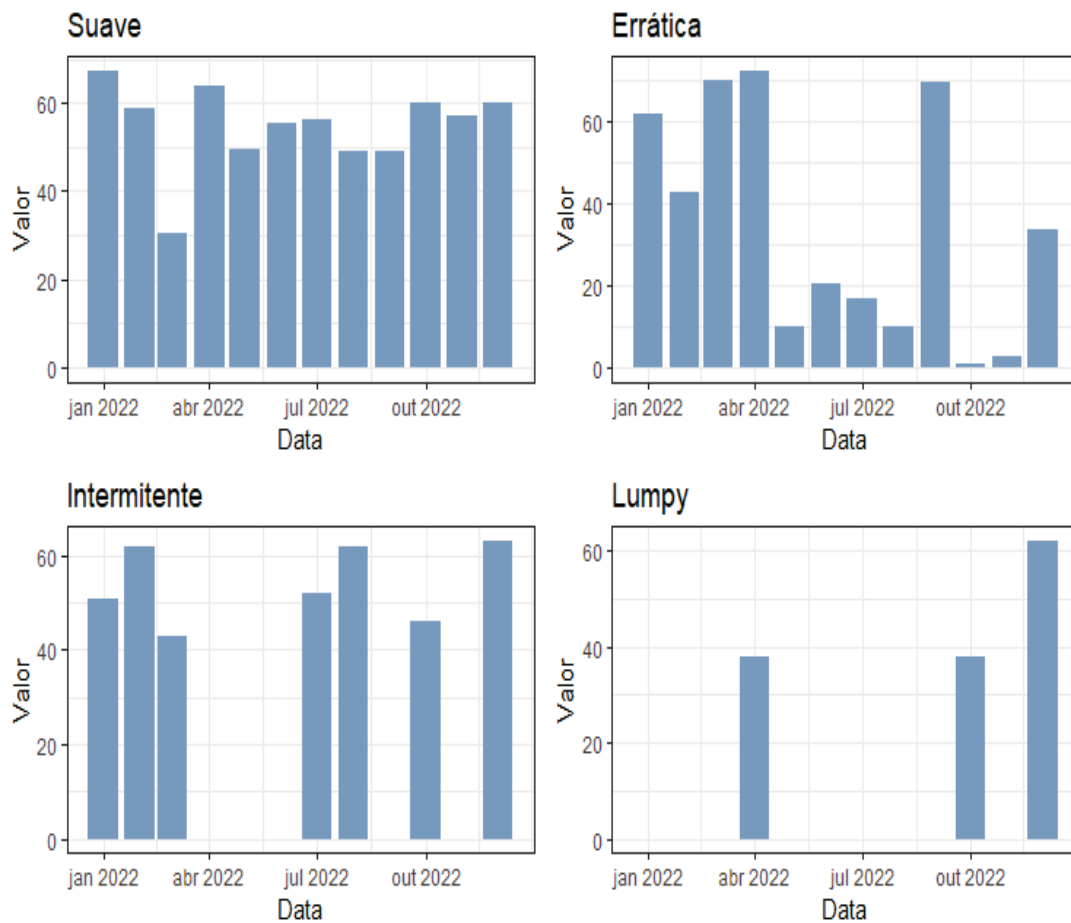


FIGURA 2 – Tipos de Demanda

Fonte: Elaboração própria

As séries temporais suaves têm uma característica de padrão consistente e leve variação ao longo do tempo, não apresentando grandes dificuldades para a atividade de previsão. Já as erráticas apresentam flutuações imprevisíveis, sem um padrão claro e em linha; as séries intermitentes têm a característica de apresentarem zeros em determinados meses ou de forma aleatória. A série temporal mais desafiadora destas três é a que contém intermitência e tem sido estudada desde os anos 70, iniciando os estudos com Croston (1972), contudo, as primeiras propostas de Croston têm suas limitações, Teunter et al. (2011a) indicou que o método era problemático devido à obsolescência e propôs um novo método, considerando dois vetores: sendo um de probabilidade da demanda e o segundo de tamanho da demanda e o motor do modelo se mantém focado em suavização exponencial; na literatura de séries intermitentes, o modelo singular de suavização exponencial se mantém bastante popular Syntetos et al. (2005), Teunter et al. (2011a), Wallström e Segerstedt (2010) e Willemain et al. (2004), dentre outros trabalhos. Nesta dissertação, será feita também a comparação com o modelo de suavização, contudo, com grande visibilidade para os modelos não sequenciais e de aprendizagem de máquina e os híbridos propostos,

Erjiang et al. (2022) fizeram um estudo segregando os tipos de demanda e fazendo escolhas dinâmicas de quais modelos seriam mais interessantes para cada tipo de demanda, contudo, com modelos clássicos para previsão temporal. Entretanto, indicaram que tipos de demanda diferentes requerem estratégias diferentes, resultando em uma estratégia chamada de DWS-A² para os intermitentes e DWS-B para demanda suaves e erráticas. Ressaltando a importância dessa avaliação de forma segregada,

Zhuang et al. (2022) também no intuito de estudar as demandas intermitentes e mapear seus testes, optaram por uma abordagem híbrida para previsão, contudo, entre modelos de classificação e previsão. Os resultados mostraram que a metodologia IDCF melhorou significativamente a precisão das previsões em comparação com métodos tradicionais, como o método SBA (Syntetos-Boylan Approximation) e técnicas de *bootstrap*. Indicando uma estratégia que envolve aprendizagem de máquina, classificação e regressão, indicando que a estratégia tende a reduzir custos operacionais.

Hu et al. (2021) em um estudo sobre as demandas intermitentes, propõe uma abordagem híbrida entre dois modelos de aprendizagem de máquina, um sendo a parte classificativa e o outro sendo usado para prever os valores da demanda. A precisão da proposta foi avaliada pela métrica SMAPE — comentada na seção 4.4 — demonstrando bons resultados para a proposta híbrida.

Teunter et al. (2011b) ressaltam que a classificação das demandas por tipo é de grande utilização nos estudos de séries temporais em demandas de diversos produtos, alimentos, indústria pesada, automotiva, dentre outras indústrias. Percebe-se que o maior interesse reside nas demandas intermitentes, uma vez que são inflacionadas por zeros, muitas vezes, e, portanto, mais difíceis de se prever, enquanto demandas como as suaves podem ser similares a dados econômicos, séries monetárias, entre outras séries com ausência de zeros.

A problemática da previsão de demanda em cadeias produtivas é um tema de longa data, iniciando junto com a criação de métodos produtivos clássicos como fordismo e Toyotismo. (OHNO, 1988) indica que a previsão de demanda sempre esteve no cerne dos sistemas produtivos, como o toyotismo, no qual a eficiência é atingida ao ter um bom ajuste entre a produção e a demanda real, de forma a minimizar estoques; em linha, (HOPP; SPEARMAN, 2008), enfatizam também a previsão de demanda no âmbito da gestão empresarial, e da transição entre o fordismo e a flexibilidade do toyotismo, o qual busca alinhar a produção com a demanda real.

Seguindo a problemática, no início dos anos 2000, (KIM; RYAN, 2003) realizou um estudo de previsão em cadeias de suprimento, indicando que técnicas subótimas como suavização exponencial podem ter um efeito potencializador no efeito chicote,

² estratégia dinâmica de pesos, sendo o melhor modelo em termos de redução de erro para os intermitentes, e uma ponderação entre todos modelos para os demais tipos

elevando os custos em toda a cadeia de suprimentos e argumenta que modelos mais avançados podem reduzir os custos excessivos de estoques. A precisão na previsão de demanda é essencial para mitigar o efeito chicote, um fenômeno amplamente estudado devido aos seus impactos negativos na cadeia de suprimentos, como aumento nos custos de inventário e transporte. Enquanto (CHEN et al., 2024) destacam que previsões inadequadas amplificam a variabilidade ao longo da cadeia, resultando em custos operacionais mais elevados, (RAFATI, 2022) apontam que o efeito chicote pode elevar os custos de inventário em até 30%. Por sua vez, (YANG et al., 2021) reforçam que abordagens de previsão que consideram fatores comportamentais e cognitivos são cruciais para reduzir a propagação de variações, promovendo maior estabilidade e eficiência nos sistemas logísticos.

(CHINTAN PATIL, 2024) avaliaram 763 empresas de diferentes setores, e apresentaram uma relação entre a variância do capital de giro e a variância da demanda, indicando que 91% das empresas têm esta relação superior a 1, indicando um capital de giro volátil. Ainda, reportaram que há empresas com variância no capital de giro de até 200%. O efeito chicote financeiro (Cash Flow Bullwhip Effect) amplifica a variância no capital de giro ao longo da cadeia de suprimentos, aumentando significativamente os custos financeiros, como despesas com juros e riscos operacionais, além de dificultar a previsão e o planejamento do fluxo de caixa, especialmente em setores como manufatura, onde as oscilações são mais intensas.

A escolha dos modelos foi baseada em uma abordagem híbrida que combina métodos clássicos, amplamente utilizados na literatura, como ARIMA e suavização exponencial, com modelos de aprendizado de máquina, como XGBoost e Random Forest. Essa decisão foi tomada considerando a diversidade de padrões das séries temporais estudadas (suaves, intermitentes e erráticas) e a necessidade de lidar tanto com componentes lineares quanto não-lineares. O uso do framework tidymodels permitiu a implementação de um pipeline padronizado e eficiente, integrando múltiplos modelos e garantindo a robustez do processo por meio de validação cruzada e otimização de hiperparâmetros. Além disso, a literatura demonstra que tais modelos apresentam resultados robustos e são frequentemente aplicados na previsão de demanda em cadeias de suprimentos.

3 MODELOS DE APRENDIZAGEM EM SÉRIES TEMPORAIS

A atividade de predição de séries temporais em *supply chain* e na indústria de forma geral tem diversas aplicações, como em logística, manutenção preditiva, dentre outros. Há evidências na literatura estatística, econométrica e de finanças, de forma geral, que modelos paramétricos como ARIMA têm bons resultados; contudo, outros estudos também demonstram que técnicas de aprendizagem de máquinas têm resultados similares e, às vezes, melhores.

Aamer et al. (2021), em um estudo de revisão sobre os trabalhos feitos no âmbito da previsão de demanda para *supply chain*, em diversos ramos da microeconomia, dentre eles agricultura, energia, tratamento de água, transportes, dentre outros. Nesta revisão foram observados 21 artigos que empregaram redes neurais, 17 de redes neurais artificiais, 13 de SVM — modelo comentado na seção 3.2.4, 4 de *random forest* — modelo comentado na seção 3.2.6, e em menor grau, demais artigos utilizando as técnicas XGBoost, regressão linear, KNN¹, dentre outros. Nesta avaliação, foi concluído que estes modelos, para este fim em específico, em diversos casos apresentam acurácia melhor quando comparados com modelos tradicionais; entretanto, outros estudos serão elencados a seguir com outras perspectivas.

Em um estudo comparativo entre o modelo auto-regressivo, integrado de médias móveis (ARIMA) e modelos de *deep learning* para predição de demandas de curto prazo, Bousqaoui et al. (2021) identificaram que os modelos com melhor performance em termos de RMSE - raiz quadrada do erro médio, foram os modelos de deep learning, por mais que a diferença tenha sido marginal em 6,25 por cento do que o modelo com menor RMSE, o qual foi o CNN - rede neural convolucional. Os autores concluem que, em termos práticos, se o usuário tiver uma restrição de custo energético e tempo, a melhor saída seria o ARIMA, uma vez que o tempo de treino e o consumo energético são na ordem de 0,41 em relação ao modelo com menor RMSE.

Ademais, em outro estudo comparativo, relativo ao *forecast* no âmbito das cadeias de fornecimento, Kannan et al. (2020), discorreram sobre demandas esporádicas e intermitentes, comparando métodos tradicionais e de aprendizagem de máquina. Neste estudo, além das avaliações das métricas de erro, também foi avaliado o ganho econômico líquido do forecast de cada modelo, para cada produto dentro da amostra. Para as demandas erráticas, os modelos naive tiveram a melhor performance; contudo, para demandas suaves, o LGBM² com variáveis exógenas e o prophet demonstram os

¹ O K-Nearest Neighbors (KNN) é um algoritmo de aprendizado de máquina supervisionado, amplamente usado para classificação e regressão. A ideia básica por trás do KNN é classificar ou prever um ponto de dados com base na maioria das classes dos K pontos de dados mais próximos a ele.

² O LightGBM (Gradient Boosting Machine) é um algoritmo de aprendizado de máquina supervisionado

melhores resultados. Reporta-se que os modelos de aprendizagem de máquina têm performances marginalmente melhores; contudo, em termos de ganho econômico, os métodos tradicionais têm melhor performance.

3.1 MODELOS SEQUENCIAIS

Os modelos sequenciais levam em consideração a ordem temporal dos dados, assumindo que valores passados da série afetam os próximos valores. Shumway e Stoffer (2010) descrevem algumas características das séries temporais como a tendência, sazonalidade, o componente cíclico e o erro. A tendência, sendo de inclinação positiva ou negativa — representando a característica sequencial — e produzindo mudanças graduais de longo prazo; a sazonalidade, sendo movimentos submetidos a subida ou queda em determinado período de tempo; a principal diferença com o componente cíclico reside no fato de que a primeira é de fácil predição, enquanto a segunda não; como, por exemplo, uma crise macroeconômica, poderia ser um componente cíclico de difícil previsão.

Morettin e Toloi (2018) em se tratando da variabilidade dos dados — uma vez que este trabalho divide as séries em quatro tipos — classificam as séries temporais em estacionárias, caso não mude suas características ao longo do tempo, e ergódicas, séries às quais são estáveis, ou seja, a média e a autocovariância podem ser estimadas com uma única realização da série.

TABELA 4 – Exemplos de Modelos Sequenciais

Modelo	Conjunto de Treinamento	Hipótese	Exemplo
ARIMA	$Z_i = \{(Y_{t-1}, Y_t) : p \leq t \leq T\}$	$h_i : Y_p \rightarrow Y$	Séries temporais univariadas
NNAR	$Z_i = \{(Y_{t-1}, Y_t) : p \leq t \leq T\}$	$h_i : Y_p \rightarrow Y$	Rede neural autoregressiva

A tabela 4 demonstra um resumo de dois modelos sequenciais, bem como o conjunto de treinamento e a hipótese por trás destes modelos sequenciais. O conjunto de treinamento, basicamente denotando um período de tempo no intervalo de p a T , com pares de pontos, sendo um em período ligeiramente anterior ao atual. A hipótese aqui reside no objetivo de prever o valor atual com dados passados, ou seja, tentando prever o valor futuro com p valores passados.

3.1.1 Suavização Exponencial

A suavização exponencial foi inicialmente proposta no meio acadêmico na década de 50. Este método consiste em fazer as predições futuras com base em médias ponderadas das observações passadas, sendo que observações mais distantes do período atual têm menor peso.

baseado em árvores que pertence à família de métodos ensemble.

O caso mais simples desta família de modelos é a suavização exponencial simples, e este método tem bons resultados em séries que têm tendência e sazonalidades bem delimitadas. Fazendo uso do método de médias, todas as previsões futuras serão iguais à média dos dados observados:

$$y_{T+h|T} = \frac{1}{T} \sum_{t=1}^T y_t \quad (3.1)$$

Considerando-se um modelo de suavização com médias ponderadas, no sentido do decaimento da importância das observações, da mais atual à mais antiga, e considerando um modelo hipotético baseado na última observação :

$$\text{Previsão:} \quad \hat{y}_{t+h|t} = \ell_t \quad (3.2)$$

$$\text{Parâmetro de Suavização:} \quad \ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} \quad (3.3)$$

Onde α representa o parâmetro de suavização, e seu domínio é $0 < \alpha < 1$, em suma, na representação anterior a projeção $t + 1$ será o nível a suavização do período ligeiramente anterior somados tais períodos dados o decaimento balizado pelo valor do α .

Subsequente, temos os métodos *Holt Winters* de tendência linear, que modificam os modelos de tendência amortecida, adicionando a tendência ao parâmetro de suavização e, portanto, previsão (Gardner Jr., 2006).

$$\text{Previsão} \quad \hat{y}_{t+h|t} = \ell_t + hb_t \quad (3.4)$$

$$\text{Nível} \quad \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (3.5)$$

$$\text{Tendência} \quad b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}, \quad (3.6)$$

Onde b_t pode ser interpretado como um parâmetro de taxa de crescimento. Este método tende a extrapolar o futuro, seja entre uma tendência de baixa ou de alta, causando efeitos indesejáveis para mais ou para menos na previsão (Gardner jr., 2006), como forma de ajuste, e inclui o parâmetro $0 < \phi < 1$, ou parâmetro de amortecimento da tendência.

$$\text{Previsão} \quad \hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t \quad (3.7)$$

$$\text{Nível} \quad \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}) \quad (3.8)$$

$$\text{Tendência} \quad b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}, \quad (3.9)$$

A sazonalidade entra nesta família de modelos representada pelo símbolo I_t , podendo ser aditiva, multiplicativa, amortecida ou inexistente. A seguir, representam-se os casos de sazonalidade com tendência amortecida aditiva e multiplicativa, sendo as formas mais complexas, para esse parâmetro, nesta família de modelos, como de tendência aditiva e sazonalidade aditiva ou multiplicativa, tendência multiplicativa e sazonalidade aditiva ou multiplicativa.

3.1.2 Auto Regressivo, Integrado de Médias Móveis

Sendo praticamente um benchmark³, os modelos ARIMA têm presença em diversos estudos estatísticos e, principalmente, na previsão de séries temporais. Temos na literatura sua importância, dado seu uso extensivo.

Hasan et al. (2021) utilizaram-se de modelos ARIMA univariados para comparação com modelos Prophet e LightGBM, a comparação foi feita em mais de 30 mil séries temporais do varejo, sendo classificadas como séries relacionadas a alimentação, domicílio e artigos utilizados para prática de atividades de lazer - dataset extraído do kaggle referente a produtos de um grande varejista americano. A comparação foi feita fazendo uso da métrica RMSE, os autores indicaram que os melhores modelos são o ARIMA e o LightGBM, e apenas no caso de alimentação o ARIMA teve uma diferença substancial em melhor previsão dentro de 28 dias; nos demais casos, as métricas são muito próximas, o modelo Prophet ficou em terceiro lugar com grande margem.

Abolghasemi et al. (2022), em estudo sobre previsão de demanda e POS data⁴, de forma a verificar a melhora de previsão nos modelos testados. Os autores utilizaram os modelos ARIMA e ETS como *benchmark* para este estudo; ademais, para acoplar as variáveis exógenas, os modelos ARIMAX e R-ARIMA foram usados, com intuito - no total, mas parcial - de captar os efeitos das promoções. A métrica utilizada neste estudo foi a RelMAE, comentada na seção 4.4; considerando oito semanas de previsão. Concluíram que os modelos que fizeram uso de variáveis exógenas tiveram melhores resultados, dada sua capacidade de utilizar mais variáveis, como preço dos produtos oito semanas à frente, e que, sendo em períodos de promoção ou não, os modelos ARIMAX têm melhor capacidade preditiva.

Van Belle et al. (2021), utiliza também informações exógenas ao processo gerador de y_t , para demandas de curto prazo em cadeias de suprimentos. Fazendo

³ referência ou padrão utilizado para avaliar o desempenho, eficiência ou qualidade de algo, geralmente em comparação com outros itens similares. O termo é amplamente utilizado em diversos contextos, como finanças, tecnologia, negócios e pesquisa

⁴ refere-se aos dados coletados no ponto de venda final, geralmente em uma loja física ou online, sobre as transações de vendas de produtos, Esses dados incluem informações sobre quais produtos foram vendidos, em que quantidade, a que preço, em que local e em que momento. Essas informações são registradas por meio de sistemas de ponto de venda, como caixas registradoras eletrônicas, sistemas de automação comercial ou plataformas de comércio eletrônico.

uma comparação entre modelos autoregressivos com variáveis exógenas, SVR, LASSO e ETS, reportaram que o uso de variáveis externas melhora a acurácia do forecast de curto prazo; a métrica escolhida para essa comparação foi AvgRelRMSE - uma modificação do RMSE já elencado, que é uma média geométrica do RMSE relativizado. O horizonte de curto prazo é de 5 períodos à frente, e em cada período foi calculada a métrica de erro, indicando que o modelo que tem os valores mínimos da métrica em mais períodos foi o LASSO, com diferenças marginais em relação ao SVR e o ARIMAX. Ademais, os autores indicam que sim, o uso de informação externa reduziu o erro, e que a melhora como um todo em relação aos modelos de *benchmark*, ARIMA e ETS, foi de 4,7 por cento.

Em suma, a utilização de modelos autoregressivos é extensiva na literatura de *forecasting*, seja como *benchmark* ou recebendo variáveis externas como proposta de uso no ambiente fabril.

3.1.3 NNAR

O modelo NNAR é um modelo de série temporal que combina redes neurais e autorregressão. Utilizado para prever valores futuros em uma série temporal com base em valores passados, utilizando redes neurais para modelar as relações não-lineares nos dados.

A estrutura básica de um modelo NNAR(p,k) segue conforme :

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + e_t \quad (3.10)$$

Onde:

- y_t é o valor da série temporal no tempo t ,
- p é o número de lags (retardos) utilizados como entradas para a rede neural,
- f é a função de mapeamento não linear aprendida pela rede neural,
- e_t é o erro ou resíduo no tempo t ,

Panigrahi e Behera (2017) em um estudo que fez uso de modelo híbrido, ETS-ANN⁵, o método híbrido consiste em modelos y_t com suavização exponencial e então modela-se o resíduo $y_t - \hat{y}_t = \epsilon_t$ com a rede neural artificial, por fim gerando uma previsão combinada entre a parte ETS e ANN, denotada por $y_t = \hat{C}_1 + \hat{C}_2$, sendo C1 a parte ETS e C2 a parte ANN. O método de comparação foi através do RMSE e SMAPE, indicando melhores resultados em 6 das 16 séries analisadas e 7 em outras

⁵ Suavização exponencial e Redes Neuras Artificiais

16 séries, respectivamente, trazendo à baila, para comparação, resultados do trabalho de Zhang (2003), Hyndman et al. (2008) e Babu e Reddy (2014) nas mesmas séries temporais.

Twumasi e Twumasi (2022) em estudo sobre previsão de demanda por sangue em um hospital geral em Gana, indicaram que particularmente, o modelo KNN e NNAR, aplicados recorrendo à técnica MIMO, tiveram um erro sobremaneira menor que o modelo clássico ARIMA. AS métricas utilizadas de comparação foram RMSE e SMAPE, indicando um trabalho mais intensivo em modificações do KNN, sendo este o que apresentou melhor desempenho.

Yin et al. (2021) avaliaram a performance de modelos como SARIMA, NNAR e LSTM, usando algumas métricas de erro e o coeficiente de correlação R. Em termos do coeficiente de correlação, o NNAR em exercício de previsão longo, em 50 dias, utilizando dados diários, indicou que o modelo NNAR com a técnica MIMO teve melhores indicadores em métricas de erro RMSE e MAPE, concluindo que o modelo NNAR teve a melhor performance média, e o estudo como um todo demonstra a aplicação de métodos clássicos e de aprendizagem de máquina com o método MIMO para previsão de séries temporais, com bons resultados e evitando propagação de erro em muitos casos.

3.1.4 Considerações Finais - Modelos Sequenciais

Os modelos sequenciais têm a característica de realizarem a previsão da série temporal de forma iterativa, contudo, com a utilização do método MIMO concomitante ao método iterativo, indicando as melhoras que podem ser trazidas ao processo de previsão combinando técnicas.

Li et al. (2023) em uma discussão acerca das técnicas iterativas e MIMO, indicando a previsão sobre o modelo RNN-Recursive⁶, no qual o modelo é treinado para prever um passo à frente e o valor previsto é reutilizado, sendo passível de acumular erros. A técnica MIMO entra no contexto como RNN-MIMO, treinando um único preditor para prever um vetor de saída, e resultados que demonstram para horizontes de previsão mais longos, os métodos sequenciais profundos, especialmente o RNN-MIMO, produzem os melhores resultados de forma consistente. Realizaram uma comparação em relação aos modelos : Naive, XGboost, SVR, RNN Recursivo e S2S. Ahmadi et al. (2023a) também realiza uma discussão acerca da propagação de erro e o horizonte de tempo, indicando que o método iterativo foi mais adequado para o curto prazo e o método MIMO para os horizontes mais longos; foram testados os modelos clássicos como ARIMA e SARIMA, métodos de aprendizagem de máquina

⁶ Recurrent Neural Network Recursive strategy

como LightGBM e *deep learning* como N-Beats e LSTM; ademais, os modelos de deep learning foram testados em ambos os métodos, iterativo e MIMO.

Sousa et al. (2023) tratam de uma nova estratégia a AEnbMIMOCQR⁷, a qual combina várias técnicas, incluindo *conformal prediction*, regressão quantílica e MIMO, para previsão de vários passos à frente. Os resultados apresentados no artigo indicam que o método proposto supera outros métodos já consolidados no mercado e na academia, em termos de cobertura observada e largura dos intervalos de previsão. Em termos da técnica MIMO, ressalta-se que contribui sobremaneira para a redução da largura dos intervalos, evitando a acumulação de erros. A comparação foi feita com os métodos : EnbPI⁸, EnbCQR⁹, ARIMA e MIMOCQR¹⁰.

Em suma, mesmo os modelos sequenciais têm sido objeto de estudo e alteração de forma muitas vezes intensiva, com técnicas como MIMO, regressões quantílicas, simulações, de forma a estressar a teoria e o poder computacional para comparar e encontrar as melhores formas e estratégias de utilização dos modelos para previsão de séries temporais.

3.2 MODELOS NÃO SEQUENCIAIS

Os modelos em questão não têm intrinsecamente em sua formulação a característica temporal, como um modelo AR, o qual dá um peso para a(s) observação(ões) ligeiramente anterior(es); portanto, não levam em consideração a ordem temporal dos dados. Bontempi et al. (2013) em artigo sobre técnicas de predição em séries temporais, discutem como o problema da predição de séries temporais é também um problema de regressão, quando utilizados os valores passados para previsão dos futuros; também argumenta-se sobre as técnicas de previsão — iterativa, direta e MIMO. Os autores argumentam acerca do *embedding*, técnica que os valores passados $(y_{t-1}, y_{t-2}, \dots, y_{t-n})$ para prever valores futuros de y , podendo o problema ser representado por:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}) + \epsilon_t$$

Sendo :

- y é a variável de saída que estamos tentando prever.
- x_1, x_2, \dots, x_n são as variáveis de entrada, também chamadas de características ou preditores.

⁷ Adaptive Ensemble Batch Multi-Input Multi-Output Conformalized Quantile Regression

⁸ Ensemble Batch Prediction Interval

⁹ Ensemble Batch Conformalized Quantile Regression

¹⁰ (Multi-Input Multi-Output Conformalized Quantile Regression) - uma versão simplificada do AEnbMIMOCQR sem adaptação dinâmica e bagging.

- f é a função que descreve a relação entre as variáveis de entrada e a variável de saída, Essa função pode ser linear ou não linear e pode ser definida de várias maneiras, dependendo do tipo específico de modelo usado.
- ϵ é o termo de erro, que representa a parte da variabilidade em y que não é explicada pelas variáveis de entrada.

Em uma revisão sobre métodos de previsão em séries temporais, Todorov e Sánchez Lasheras (2022) no escopo da previsão de energia e petróleo e gás, foi identificado o crescente uso de métodos de aprendizagem de máquina, como florestas aleatórias e máquinas de vetores de suporte, bem como a utilização de combinação de técnicas clássicas e de aprendizagem de máquina; o artigo avaliou trabalhos de 2017 até o fim de 2021.

Tian et al. (2023) fizeram, também, uma avaliação de diversas publicações no âmbito de previsões de séries temporais, aqui para energia fotovoltaica, sendo 115 artigos avaliados. Os autores destacam o uso de diversos modelos de aprendizagem de máquina não sequenciais, o uso de engenharia de características para adaptar o problema aos modelos referenciados, abordagens híbridas e agregações temporais.

3.2.1 Modelos Lineares Generalizados - Séries temporais

Os modelos lineares generalizados em séries temporais estendem a ideia de modelos lineares generalizados para dados de séries temporais, incorporando a estrutura temporal dos dados em sua formulação. Isso significa que esses modelos consideram a dependência temporal dos dados e, portanto, são capazes de modelar as flutuações e tendências ao longo do tempo.

Por exemplo, um modelo de regressão de Poisson em séries temporais pode ser usado para modelar a contagem de ocorrências de eventos em intervalos regulares de tempo. Este modelo pode incluir um termo de erro autorregressivo para capturar a autocorrelação das contagens de eventos em diferentes momentos do tempo, bem como um termo para a tendência temporal para modelar o aumento ou diminuição da taxa de ocorrência ao longo do tempo.

Kolassa (2016), realiza uma discussão acerca da avaliação de predição de dados de contagem no varejo. Apresentando uma análise comparativa de diferentes distribuições de dados de contagem, com o objetivo de identificar a melhor distribuição para modelar e prever as vendas no varejo. O autor avalia quão bem se adequam as distribuições comuns, como a distribuição de Poisson, a distribuição binomial negativa e a distribuição de Poisson inflacionada de zeros, usando dados de vendas reais de uma loja de varejo. O autor utiliza modelos de regressão de Poisson e modelos de regressão linear generalizados (GLMs) para ajustar as diferentes distribuições e fazer as previsões.

Os resultados mostram que o GLM é uma abordagem útil para modelar as flutuações e tendências nas vendas ao longo do tempo, e que o uso de diferentes distribuições pode levar a previsões mais precisas. Ademais, neste trabalho, é argumentado que tratamentos via *bootstrap* ou não, têm resultados similares, uma vez que a quantidade de dados é suficiente, a variabilidade de amostragem não causa grande interferência nas distribuições preditivas.

Rohaeti et al. (2021) faz uso de técnicas GLM para uma análise preditiva de séries temporais de contagem, em dados do mercado de turismo. Neste estudo também foram utilizadas distribuições Poisson e binomial negativa, selecionando os parâmetros autorregressivos e de médias móveis (P e Q), sendo escolhidos pelo critério de Akaike. Testando-se essas duas distribuições com função de ligação gaussiana e logit, os autores encontraram que, em termos de MAPE, os modelos desempenham de forma quase idêntica; contudo, trazendo à baila AIC e BIC, a binomial negativa com função de ligação log é melhor que os demais modelos.

Os exemplos apresentados ilustram como os modelos GLM têm sua representatividade na previsão de séries temporais em diferentes campos, destacando a versatilidade e a eficácia desses modelos para lidar com problemas complexos de modelagem preditiva.

3.2.2 MLP

O Multi-layer perceptron é um modelo de redes neurais, sendo formado por unidades de processamento simples que, juntas, se tornam uma rede massiva e paralela que tem uma propensão natural em guardar conhecimento.

O modelo estima a saída \hat{y}_j como uma combinação linear das entradas $x_{j,i}$ e dos pesos \hat{w}_i , com i variando de 0 até $N - 1$, onde N é o número de características de entrada.

A seguir um exemplo da equação linear para uma observação j :

$$\hat{y}_j = w_0 + \sum_{i=1}^N w_i \cdot x_{j,i} \quad (3.11)$$

Sua representação matricial:

$$\hat{y} = \mathbf{X} \cdot \mathbf{w} \quad (3.12)$$

O seu uso em tarefas de previsão de séries temporais tem demonstrado resultados promissores em diversas tarefas, como previsões hidrológicas, turismo, cadeias produtivas, dentre outros temas. Qiao et al. (2020) Em um estudo de previsão para o

mercado de consumo de gás, por meio do uso do MLP e da métrica de erro RMSE, comparando técnicas derivadas deste modelo, indicou ótimos resultados para uma série temporal de característica suave. Zhang et al. (2022) Fez uso da técnica baseada em MLP, de amostragem contínua e intervalar, trabalhando com séries temporais de padrões curtos e longos, reportaram melhoras em previsões longas, pela ótica da métrica de erro MSE; reportando reduções de 9% a 34% no erro.

Sousa et al. (2023) em um estudo comparativo na atividade de previsão de demanda do fluxo de clientes no varejo, sendo a técnica MIMO aplicada ao MLP e LSTM, para previsões, e o desempenho em termos de RelMAE do modelo MLP, teve o menor erro em 0,818. O estudo concluiu que a combinação de previsões de diferentes modelos é interessante para lidar com séries complexas e intermitentes; ademais, o ensemble de modelos foi recomendado. Fernandes et al. (2020) realizaram um estudo sobre métodos de previsão, incluso MLP MIMO, para uso industrial, concluindo que a estrutura proposta tem potencial para ser utilizada para previsão em tempo real de parâmetros operacionais, empregando dados reais de geradores de vapor de usinas nucleares francesas.

3.2.3 GLMnet

Os modelos GLMnet são uma extensão dos modelos de regressão linear generalizados (GLM) que incluem regularização em sua estrutura. São modelos amplamente utilizados em análise de dados e *machine learning* para lidar com conjuntos de dados de alta dimensionalidade.

A regularização é uma técnica estatística que adiciona uma penalidade ao termo de perda de um modelo para evitar o sobreajuste¹¹. O GLMnet permite a aplicação de duas formas de regularização: L1 (Lasso) e L2 (Ridge). A regularização L1 tende a gerar modelos mais esparsos, com muitos coeficientes iguais a zero, enquanto a regularização L2 tende a gerar modelos com coeficientes menores; muito usada na presença de multicolinearidade — as variáveis independentes, correlacionadas, e neste sentido, as estimativas dos coeficientes também são correlacionadas e suas variâncias são muito grandes — e sobreajuste do modelo.

De forma a controlar a inflação dos coeficientes na presença de multicolinearidade, a regularização L2 traz o parâmetro λ ou penalização de encolhimento, que será adicionado na soma dos quadrados dos erros - MQO. Quando aplicado o método MIMO ao GLMNet, estamos trabalhando com múltiplas variáveis de saída simultâneas, o que exige uma penalização de forma a evitar sobreajuste e melhorar a generalização

¹¹ Ying (2019) Em aprendizagem de máquina supervisionada, pode ocorrer o problema de que o modelo não tenha uma boa performance em dados que são externos aos dados de treino. A performance é muito boa nos dados de teste, mas a generalização do modelo fica fraca.

do modelo.

A penalização L2 tende a deixar os coeficientes com valores menores, mas não os zera, o que é útil em presença de multicolinearidade. No contexto MIMO, temos a garantia de que os coeficientes sejam ajustados, porém, mantendo a variabilidade das variáveis preditoras. Já a penalização L1 ajuda a focar nas variáveis mais importantes e ter um modelo mais esparso. Portanto, este modelo tem alguns benefícios junto ao método de predição: a robustez na presença de multicolinearidade, a seleção das variáveis mais interessantes ao objetivo de previsão e a redução do *overfitting*.

No que tange a literatura no âmbito do *forecasting*, Vollmer et al. (2021) em um trabalho relacionando a previsão de séries temporais no âmbito da demanda por serviços hospitalares, utilizaram-se modelos paramétricos e não paramétricos, dentre eles o GLMNET para a comparação entre modelos. A previsão neste caso foi diária, em uma janela total de sete dias, e as métricas de erro utilizadas para tal foram MAE e MAPE. Primeiramente, pela seleção de importância de variáveis explicativas, os autores indicaram que a média móvel da semana ligeiramente anterior é a covariável de maior importância para explicar a semana seguinte, seguidos pelas dummies diárias e mensais. Ademais, é feita uma análise pelas métricas de erro MAE e MAPE, reportando resultados interessantes para o GLMNET na janela de 1 e 7 dias, indicando que os melhores resultados vieram de regressão penalizada e métodos de *ensemble* pela teoria do *wisdom of the crowd*.¹²

Cerqueira et al. (2019), no intuito de comparar métodos clássicos como ARIMA e suavização exponencial com métodos de aprendizagem de máquina, avaliaram uma série de modelos, inclusos GLMNET com distribuição gaussiana, em 90 séries temporais diferentes. Tais séries temporais são de diferentes campos, como saúde, economia, física dentre outros. No processo de predição, considerando uma janela de 18 observações, é feita a avaliação pela métrica de erro MASE - métrica bastante útil para comparar séries temporais de características diversas (HYNDMAN; KOEHLER, 2006), para comparação entre modelos; nesta janela os três modelos com melhores métricas foram o ARIMA, GLMNET e RBR. O principal ponto dos autores neste estudo é que o tamanho da amostra de treino importa quando se tratando de modelos de aprendizagem de máquina, uma vez que as séries utilizadas contêm mais de 500 observações.

Em suma, o modelo com penalização é bastante disseminado em diversas áreas, incluso em predições sequenciais como a predição de dados temporais, apre-

¹² Luo e Liu (2023) *Wisdom of the crowd* é uma termo que primeiramente foi cunhado para opinião de humanos, se mostrou bastante eficiente no âmbito da estatística, agregando opiniões de diversos modelos de aprendizagem supervisionada e semi supervisionada. Ho (1995) ressalta que um dos métodos mais populares de ensemble é o modelo Random Forest, que extrai o resultado da predição da maioria das árvores de decisão.

sentando bons resultados quanto ao uso das predições e na competição com outras técnicas.

3.2.4 SVR - Support Vector Regression

As máquinas de vetores de suporte compõem um modelo que pode ser utilizado tanto para regressão (SVR) - uma generalização do problema de classificação - como classificação; nesta dissertação trabalharemos à ótica da regressão. O trabalho seminal de (VAPNIK, 1995) é o precursor do SVR, e dos demais trabalhos que vieram após, com suas mudanças e meta análises sobre o potencial dos vetores de suporte.

Neste método, um hiperplano provê soluções ótimas através da expansão do vetor de solução nos vetores de suporte. Em um problema multidimensional, em suma, o problema é definido pela seguinte formulação:

$$\text{Minimizar: } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$\begin{aligned} \text{sujeito a: } & y_i - \mathbf{w}^T \mathbf{x}_i - b \leq \varepsilon + \xi_i, \quad i = 1, \dots, n \\ & \mathbf{w}^T \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*, \quad i = 1, \dots, n \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n \end{aligned}$$

Aqui temos que w é o vetor de coeficientes, b é o termo de viés, C é o parâmetro de regularização, n é o número de amostras de treino, x_i é o vetor de variáveis explicativas, y_i é o valor da i -ésima amostra da variável a ser explicada, ε é a margem de tolerância - parâmetro de folga, e os ε são as variáveis de folga associadas às amostras de treinamento e o termo $\frac{1}{2} \|\mathbf{w}\|^2$ é o termo L2, que penaliza a magnitude dos coeficientes.

Uma peça fundamental para uso da técnica MIMO neste modelo é a inserção da matriz de design, contudo, neste modelo de aprendizagem de máquina temos a presença do kernel; o uso desta técnica matemática é fundamental para transformar dados da matriz de design em um espaço de maior dimensionalidade, tal espaço o qual é mais fácil de encontrar uma solução linear para problemas não lineares, sendo denotado o kernel por $k(x_i, x_j)$. O kernel aqui utilizado — gaussiano, tanto para o SVR como o modelo híbrido com SVR nos resíduos, será :

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

onde γ é um parâmetro que define o alcance da influência de cada ponto de treinamento.

Conforme seção 3.4.3, o método MIMO faz a previsão em blocos e a implementação do método com SVR utiliza-se da matriz de design que incorpora múltiplas entradas e saídas conforme a engenharia de variáveis indicada na tabela 5, sendo variáveis aplicadas ao operador de retardo. O parâmetro γ do kernel gaussiano utilizado determina o alcance da influência de um único ponto de treino; quão menor for, maior o alcance, podendo capturar relações globais. Sendo aplicado tanto nos resíduos para a técnica híbrida como na própria série temporal.

O uso do modelo em questão para *forecasting*, e inclusive no âmbito da supply chain, já tem sido explorado há alguns anos. Por exemplo, Wang (2011) realizou um estudo comparativo entre SVR e RBF ¹³ em redes neurais, indicando que desde o início dos anos 2010, há interesse neste tipo de modelagem para dados de séries temporais.

Em estudo mais recente, Kilimci et al. (2019) estudaram 1500 séries de vendas em cadeias produtivas focadas em produtos de limpeza e *self care*, as métricas de erro escolhidas para avaliação dos modelos foram MAPE e MAD. Neste estudo em questão, os autores fazem uso de *ensemble* dos modelos - já identificados em outros trabalhos, como uma proxy pela teoria *wisdom of the crowd*, e identificaram ganhos de previsão em cerca de 3,5%.

Niu e Feng (2021) , fazem uma avaliação de diversos métodos de aprendizagem de máquina para previsão de vazão e gerência de recursos hídricos, aplicando a métrica de erro RMSE para comparação dos modelos. Separados por reservatórios, o modelo SVM, no primeiro reservatório, teve o segundo melhor RMSE na amostra de teste; já no segundo reservatório, teve a melhor métrica. Perdendo por uma diferença marginal apenas para o extreme learning machine¹⁴, no primeiro reservatório, em cerca de 2,8%.

(SELIM et al., 2020) em uma análise de propagação de erro na atividade de *forecasting*, indicou que o SVR na utilização do método MIMO apresentou redução na propagação de erros, comparando-se com métodos iterativos e diretos.

¹³ É um tipo particular de rede neural baseada em funções base radiais, consistente de três camadas e dependem da distância radial entre um ponto de entrada e um centro, A ideia principal por trás do modelo RBF é que essas funções podem ser usadas para representar e aprender relações não lineares entre os dados de entrada e saída.

¹⁴ Extreme Learning Machine (ELM) é um algoritmo de treinamento para redes neurais feedforward — ou seja, a informação flui em uma única direção — de única camada oculta (SLFN), que converge muito mais rapidamente do que métodos tradicionais e oferece desempenho promissor (WANG et al., 2022).

3.2.5 XGboost

Xgboost é um algoritmo de aprendizado supervisionado usado para regressão e classificação em grandes conjuntos de dados. Utiliza-se de árvores de decisão construídas de forma sequencial e superficial para fornecer resultados precisos, além de um método de treinamento altamente escalável que evita o sobreajuste.

A diferença nevrálgica entre o método de regressão e classificação deste método de aprendizagem de máquina, é a função perda (*loss function*). Sendo para o caso da regressão :

$$\text{Loss} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.13)$$

onde:

n é o número de exemplos no conjunto de dados de treinamento.

y_i é o valor verdadeiro (rótulo) da variável de saída para o i -ésimo exemplo.

\hat{y}_i é o valor previsto (saída do modelo) para o i -ésimo exemplo.

O modelo em questão usa esta função para criar suas árvores, minimizando a seguinte equação:

$$\text{Loss} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \gamma T + \frac{1}{2} \lambda O_{\text{value}}^2 \quad (3.14)$$

Em suma, é quantificado quão bom ou adequado é o modelo, por meio da função perda. Ademais, T nesta equação se refere às folhas da árvore e γ um termo penalizador fornecido pelo pesquisador, sendo bastante útil para evitar *overfitting* pois realiza o *pruning* - processo de remoção de nós ou subárvores de uma árvore de decisão durante o treinamento.

No contexto de predição MIMO, o xgboost bem como o modelo GLMNet, conta com um parâmetro de penalização L2 (γ) no número de folhas T , que encolhe os coeficientes, e a regularização pela complexidade da árvore mantém o modelo mais enxuto e com poder de generalização; ademais, considera a interação entre as variáveis preditoras e a dependência com a variável que se quer prever, uma vez que as árvores sequenciais são ajustadas de maneira a melhorar a previsão conjunta e de forma a considerar as correlações entre as variáveis da matriz de design X .

Panda e Mohanty (2023) em um estudo de *forecasting* em *supply chain* para cadeia de alimentos, com horizonte de previsão de dez semanas, fazendo uso de métodos de aprendizagem de máquina e tendo como balizadores as métricas de erro RMSE,

MAPE e MAE. Fazem uso de técnicas de *feature engineering*¹⁵ como os operadores de retardo da variável que se quer prever e aplica-se uma média móvel ponderada nestes dados, para atuar como preditora; ademais, os autores também utilizaram variáveis como preço base e preço por refeição. A comparação de modelos retornou que, especificamente para este estudo, o modelo LSTM¹⁶ foi melhor substancialmente que todos os outros em termos de RMSE e MAE, contudo não em termos de MAPE, enquanto XGBoost ficou muito similar aos demais.

Abolghasemi et al. (2019) exploram o uso de modelos como XGboost, SVR e ANN, no contexto de séries temporais hierárquicas. Os resultados mostram que os modelos de ML são competitivos e, em alguns casos, superam os modelos estatísticos tradicionais, especialmente quando se trata da desagregação de vendas para níveis inferiores da hierarquia. O XGBoost se destacou como o modelo mais eficaz em prever as proporções de vendas nos níveis inferiores, seguido pelo SVR. O estudo também aborda as abordagens tradicionais de previsão hierárquica, como *Top-Down* (TD), *Bottom-Up* (BU) e *Combinação Ótima* (COM), e introduz uma abordagem intermediária (*Middle-Out*)¹⁷ utilizando os modelos de ML para a desagregação dinâmica das séries temporais¹⁸. Este trabalho sugere que a incorporação de modelos de ML em previsões hierárquicas pode oferecer melhorias significativas na precisão das previsões, especialmente em contextos dinâmicos afetados por fatores exógenos.

3.2.6 Random Forest

Introduzido por Breiman (2001), o modelo Random Forest é um método não paramétrico, e pertence à categoria de métodos de conjunto, é muito comum ver este algoritmo associado às tarefas de classificação, contudo, é possível aplicar em problemas de regressão. No contexto de regressão, o objetivo é prever um valor contínuo.

A ideia principal por trás deste modelo é combinar diversas árvores de decisão em uma única predição, ou modelo, e cada árvore é treinada com uma amostra *bootstrap* dos dados e com subconjuntos aleatórios das variáveis; a combinação que

¹⁵ É o processo de criar novas variáveis ou transformar características existentes — exemplo : utilizar séries de fourier em dados temporais — com o objetivo de melhorar a performance de um modelo.

¹⁶ É um tipo de rede neural recorrente, de memória longa de curto prazo que tem o intuito de trabalhar problemas de dependências de longo prazo em dados sequenciais.

¹⁷ A abordagem intermediária (Middle-Out, MO) sugerida no artigo envolve a previsão inicial no nível intermediário da hierarquia, seguida pela agregação dessas previsões para os níveis superiores e pela desagregação para os níveis inferiores, Especificamente, o estudo foca em gerar previsões no nível dos varejistas (nível médio), para depois agregá-las e obter as vendas totais no nível superior e, em seguida, desagregá-las para prever as vendas nos níveis inferiores, como os centros de distribuição (DCs)

¹⁸ A desagregação dinâmica das séries temporais trata de decompor séries agregadas em níveis componentes inferiores, sendo bastante útil para prever vendas totais e depois desagregar essas previsões para produtos individuais ou regiões geográficas.

gera esta floresta tem a característica de reduzir o erro total, minimizando tanto a variância quanto o viés.

Em uma árvore de regressão, cada folha representa um valor numérico, diferente da árvore de classificação, a qual retorna *TRUE* ou *FALSE* ou alguma outra forma de categorização. O modelo em questão é uma coleção de árvores preditoras $h(x; \theta_k)$, e a previsão desta floresta se dá por :

$$\bar{h}(x) = \frac{1}{K} \sum_{k=1}^K h(x; \theta_k) \quad (3.15)$$

O método MIMO permite que apenas um único modelo Random Forest seja treinado para prever várias variáveis de saída simultaneamente, dado que as árvores produzirão vetores de saída, capturando a interdependência na saída. A capacidade das árvores de decisão de capturar relações não lineares entre as variáveis de entrada e saída é amplificada no contexto MIMO, observando o comportamento do passado recente da série temporal.

(DUDEK, 2022) fazendo uso do modelo em questão para um estudo comparativo em previsão de séries temporais com horizonte curto de previsão sendo o escopo. Apresentam um trabalho intensivo de tratamento das variáveis para previsão de demanda de energia, como estação do ano, dia e hora e três métodos de treinamento, como global, local e global estendido. O modo global estendido apresentou os menores erros MAPE combinando com padrões de entrada; esse método global estendido faz uso de toda a informação de padrões e variáveis que representam o padrão da série em lag.

Spiliotis et al. (2020) realizam uma comparação abrangente entre 11 métodos estatísticos e 7 de aprendizagem de máquina, para previsão de séries temporais diárias, sendo 3300, no âmbito do grande varejo. Foram usadas duas métricas de erro: RMSSE e AMSE; estas métricas são variações do Erro Absoluto Médio Escalonado (MASE), consideradas adequadas para avaliar a precisão e o viés das previsões em múltiplas séries temporais. Os quatro melhores métodos são todos de ML, sendo eles: Gaussian Process (GP), Random Forest (RF), Support Vector Regression (SVR) e k-Nearest Neighbors Regression (KNNR). Especificamente, Gradient Boosting Trees (GBT) e RF apresentaram o melhor desempenho geral, superando os métodos estatísticos tradicionais e outros métodos de ML, como redes neurais. Em suma, o artigo demonstra um modelo Random Forest utilizado com uma gama de propostas de variáveis exógenas e tratamentos diversos, demonstrando o potencial da técnica MIMO para previsão de séries temporais, neste caso, séries suaves de demanda por energia.

Punia et al. (2020) propõe um método híbrido de previsão que integra redes

LSTM e Random Forest (RF), projetado para capturar tanto suas características temporais quanto variáveis explicativas relacionadas à demanda; a avaliação foi feita dentro de previsão semanal, sendo 4 semanas à frente para avaliação. As redes LSTM são empregadas para modelar a sequência temporal dos dados de vendas, enquanto o RF é utilizado para capturar variações não temporais na demanda, como promoções, variações de preço, entre outros fatores exógenos. A comparação foi feita por meio do uso de três métricas de erro, ARME (*Average Relative Mean Error*), ARMAE (*Average Relative Mean Absolute Error*), e ARMSE (*Average Relative Mean Squared Error*). Os resultados demonstram que o modelo proposto — híbrido — foi melhor em todas as avaliações, indicando que o híbrido pode superar as limitações do LSTM e do RF, quando em conjunto.

3.2.7 Cubist

O modelo *Cubist* é um algoritmo de aprendizagem de máquina, desenvolvido principalmente para tarefas de regressão, sendo conhecido por combinar características de árvores de decisão e regressão linear, sendo um modelo híbrido. De acordo com Kuhn et al. (2014), a estrutura do modelo consiste em um componente condicional, que age como uma função segmentada, em conjunto com vários modelos de regressão linear.

O modelo é operacionalizado de forma que as árvores crescem, e a cada ponto final existe uma predição linear, enquanto os ramos ou galhos são considerados regras "if-then". A seguir, os principais pontos deste modelo:

- **Construção de Modelos Base:** Inicialmente, o Cubist constrói modelos de árvore de decisão, Cada nó terminal da árvore (ou seja, cada folha) está associado a um subconjunto de dados específico, com base nas condições "se-então" ao longo do caminho da árvore.
- **Regressão Linear em Nós Terminais:** Em vez de fazer uma previsão pontual no nó terminal, como é típico nas árvores de decisão, o Cubist ajusta um modelo de regressão linear aos dados associados a cada nó terminal. Isso permite que o modelo capture relações lineares entre as características e a variável alvo dentro de cada segmento de dados.
- **Regras de Comitê:** Uma característica distintiva do Cubist é a sua capacidade de combinar múltiplos modelos (chamados de comitês) para melhorar a precisão das previsões. Cada modelo no comitê pode ser aplicado a diferentes regiões do espaço de entrada, e as previsões são então ponderadas e combinadas de maneira a minimizar o erro geral.

- **Suavização:** Para lidar com as transições entre diferentes modelos de regressão linear (ou seja, entre diferentes folhas da árvore), o Cubist utiliza um processo de suavização. Isso ajuda a garantir que as previsões sejam mais consistentes e menos propensas a mudanças abruptas, o que é especialmente útil em casos de fronteiras de decisão complexas.

Bauer et al. (2020) apresentam um *framework* automatizado, focado em séries temporais sazonais; O conjunto de dados utilizado para a avaliação do *framework* é heterogêneo, composto por 150 séries temporais reais e publicamente disponíveis, coletadas de diversas fontes. A comparação é feita em vários métodos de aprendizagem de máquina, inclusive o Cubist, o qual apresentou em 55% dos *splits* de treinamento o menor erro médio de previsão, mas apenas em 17% dos *splits* de teste; sendo a métrica de erro utilizada o MAPE. Ressalta-se a necessidade do sistema de recomendação para escolher o melhor método de previsão neste *framework*.

Silva et al. (2020) propõe uma estratégia inovadora para previsão de vendas no varejo; a pesquisa foca na previsão de vendas de uma loja específica da Rossmann, utilizando uma estratégia de previsão para um, sete e quatorze dias à frente, aplicando a Decomposição Modal Empírica de Ensemble. Cada componente da série temporal é treinado separadamente com uma rede neural *feedforward*, um modelo Cubist e um modelo SVR, e as previsões são somadas para criar um modelo composto para cada abordagem. Para previsão de um dia à frente, o modelo EEMD-BRNN mostrou uma melhora significativa na precisão, já em horizontes de previsão de sete e quatorze dias, o modelo EEMD-Cubist demonstrou boa eficácia, apresentando melhorias de erro de 1,30% a 62,80% para sete dias e de 8,92% a 60,41% para quatorze dias, em comparação com os modelos não decompostos; ademais, a métrica utilizada foi o MAPE. Os resultados demonstraram boa performance do modelo Cubist quando se tratando de não estacionariedade.

3.2.8 MARS

O modelo MARS (*Multivariate Adaptive Regression Splines*) é uma técnica de aprendizado de máquina que combina elementos de regressão linear e splines ¹⁹ para criar um modelo flexível de regressão; sendo algumas vezes utilizada para modelar relações não lineares.

A técnica em questão é amplamente utilizada para modelar relações não lineares complexas nos dados. Este método divide o domínio das variáveis explicativas em regiões, de acordo com pontos específicos, chamados "nós", criando, assim, um

¹⁹ Splines são uma técnica matemática usada para aproximar funções complexas dividindo-as em segmentos menores e mais simples; o modelo mars utiliza-se de splines lineares em cada nó de sua árvore; os splines lineares conectam a função em cada nó, criando uma função contínua,

modelo que é uma peça linear em cada região. A técnica é similar às funções degrau, onde a relação entre a variável dependente e independente é alterada em cada "nó".

Exemplo de "nó":

$$Y = \begin{cases} a_1 + b_1X & \text{se } X < c, \\ a_2 + b_2X & \text{se } X \geq c, \end{cases}$$

Onde:

- a_1, a_2 são os interceptos para cada região.
- b_1, b_2 são os coeficientes angulares para cada região.
- c é o ponto de corte que define cada região.

O modelo MARS tem o seguinte formato:

$$\hat{f}(x) = \sum_{i=1}^k c_i B_i(x)$$

Sendo uma combinação linear de uma ou mais funções base, e cada função base tomará uma das três formas:

- Intercepto
- Função de dobradiça ou *hinge function*
- O produto de duas ou mais *hinge function*

O modelo MARS, aplicado junto ao método MIMO, com suas funções base, pode captar interdependências entre as variáveis de saída. Bem como os modelos anteriores, é um método capaz de modelar relações não lineares bastante complexas. Estes modelos, que são adaptativos, podem-se ajustar a características específicas das séries temporais, como séries mais complexas, erráticas ou intermitentes, e mesmo ser utilizado como auxiliar em um modelo híbrido.

Taylor et al. (2012) em um estudo de previsão de séries temporais que utiliza o modelo MARS e a técnica MIMO, para previsões no âmbito de dados financeiros de alta frequência, aplicando médias móveis e outras variáveis exógenas aos processos da série temporal. Neste estudo, MARS foi comparado com SVR, e duas redes neurais, uma BPN e outra CMACNN. A discussão teórica feita neste artigo enfatiza que o método

pode identificar variáveis preditoras importantes que podem influenciar o resultado das previsões e também a interpretabilidade do modelo. Os resultados indicaram que o MARS teve o melhor MAPE neste estudo em relação aos demais modelos.

Sahraei et al. (2021) em estudo sobre previsão de demanda de energia em transporte, contudo, aplicando cinco diferentes modelos MARS para comparação, com intuito de tanto prever quanto ter um nível de explicabilidade para o fenômeno da demanda de interesse. Mehdizadeh et al. (2020) avaliam a estimativa de velocidades do vento em escalas de curto e longo prazo, utilizando-se de modelos híbridos, especificamente modelos ARIMA MARS e ARIMA Random Forest, e fazem uma discussão acerca das vantagens e melhor desempenho que os modelos autônomos, capturando tanto os termos estocásticos quanto determinísticos; concluem que os modelos autorregressivos híbridos de MARS e Random Forest são melhores que apenas os modelos singulares. Menéndez-García et al. (2024) faz também um estudo comparativo entre modelos, contudo, para prever os preços do cobre, ressaltando a interpretabilidade do modelo MARS, contudo, não tão eficaz quanto o SVR e o MLP, especificamente nesta pesquisa.

Kim et al. (2020) testaram modelos MARS, M5Tree, KELM e BMA, demonstrando resultados promissores no que tange ao modelo BMA, um modelo bayesiano; intenção de ser usado o MARS neste contexto — de previsão de fluxo de água em dois rios distintos —. Os resultados indicam que o modelo MARS tem resultado comparável aos demais modelos em termos de RMSE; contudo, há uma superioridade da técnica bayesiana neste âmbito.

3.3 MODELOS HÍBRIDOS

Os modelos híbridos — são relativamente recentes, vide umas de suas primeiras aparições em Pai e Lin (2004) e posteriormente em trabalhos mais recentes como Büyüksahin e Ertekin (2019), Qiao et al. (2020), Panigrahi e Behera (2017) dentre outros. Tais modelos representam uma fusão entre o modelo clássico autorregressivo, ou suavização exponencial, aplicado à componente linear da série temporal e modelos de aprendizagem de máquina aplicados à parte não linear, ou seja, aos resíduos. O procedimento para a construção desses modelos híbridos envolve a combinação de técnicas MIMO (Múltiplas Entradas e Múltiplas Saídas) e um processo iterativo, onde o componente iterativo é aplicado à parte linear e o MIMO é aplicado aos resíduos da parte anterior, que é modelada com um dos modelos citados.

$$Z_t = Y_t + N_t$$

$$\varepsilon_t = Z_t - \tilde{Y}_t$$

$$\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + \alpha_t$$

$$\tilde{Z}_t = \tilde{Y}_t + \tilde{N}_t$$

Basicamente, o procedimento, tanto para o modelo de suavização quanto para o ARIMA, será o mesmo, conforme as equações anteriores e o fluxograma a seguir:

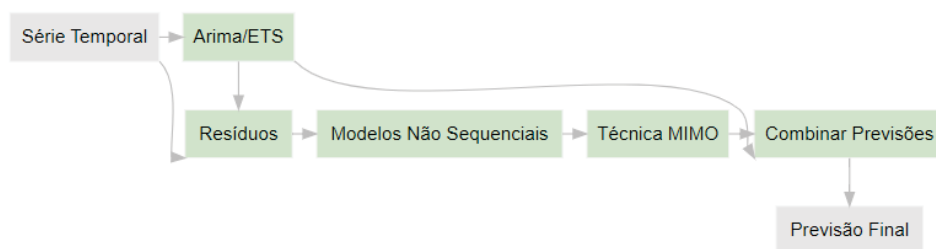


FIGURA 3 – Metodologia Utilizada
Fonte: Elaboração própria

Em linha, os modelos ARIMA e ETS foram utilizados por suas funções automáticas do software R, encontrando a melhor forma de cada modelo para os dados, sem interferência. Aqui, os modelos não sequenciais aplicados aos resíduos fizeram uso também da técnica MIMO, conforme exemplificada na seção 3.4.3; para tal, uma engenharia de variáveis nos resíduos é feita e mesmo nos valores observados.

De forma a poder utilizar a técnica MIMO, o operador de retardo é aplicado na variável de resíduo ε_t em seis períodos, então, faz-se uso de uma mediana móvel e também na própria variável Y_t . Aqui, o uso da mediana se dá pelo fato de ser robusta a *outliers*, sendo uma medida de tendência central, e é bastante útil quando a distribuição dos resíduos não é simétrica. A aplicabilidade em modelos não sequenciais como Random Forest e SVR pode ser benéfica uma vez que a mediana não é afetada pela possível característica sequencial dos dados.

Um estudo realizado por Guo et al. (2021), sobre modelos híbridos para previsão de demanda de produtos manufaturados, na China, tratando da técnica prophet²⁰ com SVR²¹. O modelo prophet é aplicado a componentes lineares, gerando previsões L_t e os resíduos e_t extraídos e aplicados ao modelo SVR; contudo, a previsão SVR

²⁰ modelo idealizado pelos colaboradores da empresa Facebook, é um modelo de série temporal baseado em uma abordagem aditiva.

²¹ Support Vector Regression

não fica especificada em iterativa, MIMO ou direta. Ademais, há indicativo de melhores métricas de erro em termos de MSE, RMSE e MAPE, quando comparado com Holt-Winters, LSTM, SARIMA, o próprio SVR e o Híbrido SARIMA-SVR.

Em linha com o estudo anterior, Abbasimehr et al. (2024), utilizaram-se também do prophet para modelar a parte linear, no que eles denominaram de séries temporais caóticas. O índice de Lyapunov máximo (MLYE) é calculado para verificar a presença de caos na série temporal. Um MLYE maior que zero indica comportamento caótico. Os resultados para a comparação mostram que o modelo híbrido prophet-EMD superou de forma consistente os modelos individuais em termos de MAPE, SMAPE e RMSE. O estudo conclui que esta técnica é bastante útil para séries temporais sazonais e caóticas; sendo as comparações feitas contra modelos SARIMA, LSTM, híbridos de prophet-SVR, ETS-ANN e redes neurais profundas.

Em Sina et al. (2023), foi realizado um estudo sistemático sobre os trabalhos que fizeram uso de modelos híbridos. Nesta varredura, os principais híbridos encontrados foram combinações de: ARIMA, SARIMA com LSTM, ANN e BPNN. Os modelos híbridos demonstraram um desempenho, em média, superior em termos de RMSE e MAPE, e especialmente com uso da decomposição por wavelet. Esta revisão sistemática selecionou 22 artigos dentre 1400, indicando que os métodos híbridos oferecem previsões, muitas vezes, mais precisas que os modelos individuais. De forma a denotar o forte uso de técnicas híbridas na literatura e pesquisas contínuas, relativas às técnicas, decomposições das séries e tipos das séries temporais, como estudado nesta dissertação.

Rotina para a metodologia utilizada

1. Dada a série temporal $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$,
2. Define-se o tamanho da amostra de treino l_{treino} (90%) e da amostra de teste l_{teste} (10%).
3. Os dados são utilizados na escala logarítmica, com exceção do Glnet Poisson.
4. Determina-se o melhor modelo (S)ARIMA(p,d,q) e ETS (E, T, S) usando os dados de treino $[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{l_{\text{treino}}}]^T$.
5. Calcula-se o SMAPE dos dados de treino.
6. Obtêm-se as previsões dos modelos ARIMA e ETS $[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{l_{\text{treino}}}]^T$.
7. Extraem-se os resíduos das séries, representando o segundo componente de interesse $e_t = y_t - \hat{y}_t$.

8. Os resíduos do passo anterior são inseridos, juntamente com as medianas móveis, nos modelos não sequenciais, para obtenção do modelo.
9. Em adição ao passo anterior, são utilizadas componentes de Fourier.
10. Realiza-se a otimização dos hiperparâmetros dos modelos não sequenciais com um grid de tamanho 100.
11. Por fim, faz-se a predição do modelo não sequencial $[\hat{e}_1, \hat{e}_2, \dots, \hat{e}_{t_{\text{treino}}}]^T$.
12. A previsão final é obtida combinando as predições ETS ou ARIMA \hat{Y}_t^1 , com as previsões dos modelos não sequenciais \hat{Y}_t^2 .

Relativo à escolha do tamanho da base de treino, no passo dois da rotina para a metodologia, Bergmeir et al. (2018) ressaltam que uma prática comum para capturar padrões e complexidades de forma efetiva em séries temporais se dá em alocar uma porção significativa de informação no treino. Ademais, ressaltam que separar 90% para o treino é uma prática comum.

3.4 MÉTODOS DE PREDIÇÃO

Os métodos de predição em séries temporais e em geral, predições de muitos passos à frente, têm sido estudados extensivamente em muitos campos da ciência; com isso, diversas técnicas têm sido discutidas e comparadas, dentre elas os métodos iterativos, MIMO²² e os métodos diretos. O trabalho aqui apresentado fez uso dos métodos iterativos e MIMO, com utilização de ambas as técnicas nos modelos híbridos. A seguir, serão elucidados os métodos, principalmente o método MIMO, com o intuito de cobrir a teoria e indicar casos de uso; método o qual é o mais jovem dentre os três, tendo sua proposta inicial em 2009 e, desde então, sendo bastante utilizado em estudos teóricos e na indústria como um todo, para atividades de previsão.

3.4.1 Método Iterativo

Esta estratégia de predição envolve a criação de um modelo de séries temporais minimizando os quadrados dos resíduos de uma etapa a frente; o modelo pode ser representado por:

$$\phi_{t+h} = f_h(x_h) + \omega_h, \quad \omega \text{ denota um ruído aditivo} \quad (3.16)$$

Aqui, D_h é definido como $\{(x_t, y_{t,h})\}_{t=d}^{N_t}$, onde $x_t = \{\phi_t, \dots, \phi_{t-d+1}\}$ e $y_{t,h} = \phi_{t+h}$,

²² Multi Input - Multi Output

Conforme indicado pela 3.16, nota-se que os valores de predição são retroalimentados no modelo para predições em $H + 2, H + 3, \dots, H + n$, portanto o nome de método iterativo. Um exemplo clássico deste método seria um modelo AR(1)²³, denotado por:

$$y_{t+1} = \alpha + \sum_{i=1}^p \phi_i y_{t+1-i} + \epsilon_t \quad (3.17)$$

onde:

- y_t é a variável de série temporal.
- α é uma constante.
- ϕ_i são os parâmetros do modelo AR.
- p é a ordem do modelo.
- ϵ_t é o termo de erro branco.

Para a previsão iterativa AR, os parâmetros na Equação (3.17) são estimados recursivamente por Mínimos Quadrados Ordinários (OLS), e as previsões de y_{t+h} são construídas recursivamente como:

$$\hat{y}_{t+h} = \hat{\alpha} + \sum_{i=1}^p \hat{\phi}_i \hat{y}_{t+h-i}$$

onde os chapéus indicam valores estimados ou previstos e $\hat{y}_j = y_j$ para $j \leq t$.

3.4.2 Métodos Diretos

Diferente da estratégia iterativa, a estratégia em questão envolve a criação de um modelo de previsão separado para cada horizonte de tempo que se deseja prever. Aqui a expectativa condicional $E[X_{n+h}|X_{1:n}]$ é o alvo desta abordagem para previsão e temos fórmulas matriciais associadas para a previsão e a covariância do processo de erro de previsão.

A expectativa condicional, denotada por $E[X_{n+h}|X_{1:n}]$, é central na abordagem direta de previsão.

²³ O processo Y_t é dito autorregressivo de ordem 1, com média nula e de forma abreviada AR(1), caso satisfaça a equação: $X_t = \phi_1 X_{t-1} + \epsilon_t$ e $(1 - \phi B)X_t = \epsilon_t$; Onde B é denominado operador de retardo, ou seja $BX_t = X_{t-1}$ e $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t = W_t$ é o polinômio regressivo de ordem p , Ademais, X_t é estacionário e $W_t \sim N(0, \sigma_w^2)$ é um ruído branco e ϕ_p são constantes,

Representação de dados para cada horizonte de previsão h :

$$D_h = \{(x_t, y_t, h)\}_{t=d}^{N_t}$$

onde

$$x_t = \{\phi_t, \dots, \phi_{t-d+1}\}$$

e

$$y_{t,h} = \phi_{t+h},$$

Um modelo é aprendido para cada h :

$$\phi_{t+h} = f_h(x_t) + \omega_h$$

A previsão, considerando um horizonte h é denotada por

$$\hat{\phi}_{t+h} = f_h(\phi_t, \phi_{t-1}, \dots, \phi_{t-d+1})$$

e

$$\rho_{ij} = \rho_{ij}$$

Esta metodologia foi primeiramente sugerida por Cox (1961), ademais Franses e Legerstee (2009) indica que a estratégia busca minimizar a soma dos erros quadrados em previsões de um passo à frente, mas não garante minimização para previsões h passos à frente, e a estratégia em questão faz h modelos, sendo que cada modelo é utilizado para prever o próximo passo e retroalimenta-se da previsão anterior feita pelo modelo h_{t-1} para o treino do modelo atual.

3.4.3 Método MIMO

O método consiste em utilizar uma gama de entradas para predição de vários passos à frente. O valor predito não é um escalar, porém, um vetor de valores futuros $(y_{n+1}, y_{n+2}, \dots, y_{n+h})$, de forma a preservar a dependência estocástica temporal oculta nos valores preditos da série — uma vez que os valores de y aplicados ao operador de retardo são utilizados no treino, teste e na previsão. Este método foi primeiramente proposto por Bontempi (2008) e tem sido utilizado de forma extensiva em diversos trabalhos, por exemplo: Sahoo et al. (2020), Shehzad et al. (2022), Ptotic et al. (2022), dentre outros.

$$D = \{(x_i, y_i) \in (\mathbb{R}^m \times \mathbb{R}^H)\}_{i=1}^N$$

$$x_i \subset \{\phi_i, \dots, \phi_{i+d-1}\}, y_i = \{\phi_{i+d}, \dots, \phi_{i+d+H-1}\},$$

$$\hat{y}_{t+h} = f(y_t, y_{t-1}, \dots, y_{t-d+1}) + w_t$$

onde:

- \hat{y}^{t+h} é o vetor de valores preditos h passos à frente.
- d é a ordem do modelo, ou seja, o número de lags considerados.
- w^t é o vetor de ruídos.

E pode ser sumarizada como :

$$\{y_i\} = f(x_i) + w_i \quad (3.18)$$

Em (3.18) temos o modelo de predição e o ruído do modelo (w_i), respectivamente, Ademais, para estimação dos valores futuros :

$$\{\hat{\phi}_{t+1}, \dots, \hat{\phi}_{t+h}\} = \hat{f}(\hat{\phi}_t, \dots, \hat{\phi}_{t-d+1})$$

A covariância do vetor de ruído W representa a incerteza ou erro inerente às previsões dos modelos. A seguir, a representação matricial de cada método e uma discussão acerca do que essas matrizes dizem sobre cada método:

$$W_{\text{iterado}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

$$W_{\text{direto}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

$$W_{\text{MIMO}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

onde:

- σ_i^2 é a variância do erro da i -ésima variável de saída.
- σ_{ij} é a covariância entre os erros das variáveis i e j .

Para o método iterado, temos uma matriz diagonal, sendo sigma a variância do i -ésimo termo de ruído e fora da diagonal, apenas zeros, uma vez que esta técnica assume que não há correlação entre os termos de ruído. Já nos métodos direto e MIMO, é observado o contraste, fora da diagonal temos os termos ρ_{ij} , que podem ser diferentes de zero, indicando a correlação dos termos de ruído.

No método direto, a matriz é simétrica, indicando que pode ou não haver correlação entre o ruído em diferentes períodos de tempo.

No método MIMO, a matriz de covariância não é necessariamente simétrica ou diagonal. De acordo com Taieb et al. (2010), a simetria da matriz implicaria em uma incerteza recíproca, indicando que as covariâncias entre as saídas i e j seriam as mesmas que entre j e i . No entanto, em sistemas complexos, como séries temporais macroeconômicas ou microeconômicas, as influências podem ser unidimensionais ou assimétricas, refletindo a natureza intrinsecamente não linear dessas relações.

A relação não-linear significa que a mudança em uma variável de saída devido a um erro em outra variável de saída ou entrada não segue um padrão proporcional ou até mesmo constante. Em muitos sistemas reais, as relações complexas entre variáveis podem envolver interações que não são bem descritas por modelos lineares; em linha com o tema, neste trabalho, abordamos modelos que podem captar essa característica não linear, como o SVR que pode lidar com relações não lineares usando kernels não lineares como o radial basis function (RBF), o GLMnet pode trabalhar bem com alta dimensionalidade e Xgboost que usa árvores de decisão e pode capturar relações não lineares complexas, dentre outros modelos.

Selim et al. (2020), em estudo sobre a propagação do erro relativo aos métodos de previsão, compara as três técnicas, iterativa, direta e MIMO. Em específico ao método MIMO, conclui-se que este é útil para evitar a redução da acumulação de erros e proporciona previsões bastante estáveis e precisas para múltiplos horizontes de

tempo; quanto ao método mais tradicional, o iterativo, os autores indicam que, embora eficaz no curto prazo, muitas vezes acumulam-se erros muito rapidamente. Especificamente, o modelo SVR e XGBoost utilizados neste estudo apresentaram redução da propagação dos erros e a tendência de capturar melhor as interdependências das variáveis explicativas de saída.

Ademais, o método MIMO tem similaridades e diferenças em relação aos processos AR e MA. Em relação ao processo AR, a dependência temporal é uma similaridade, onde os valores passados são considerados para previsão e também a linearidade, dependendo da escolha subjacente do modelo; as diferenças mais pujantes residem na característica multivariada e a predição em vetores. Quanto aos processos MA²⁴ (médias móveis), as semelhanças mais importantes são que o método em questão também pode incorporar os erros em suas predições, especialmente quando o resíduo é modelado e também a questão da linearidade como na comparação com o AR; e as diferenças são as mesmas das anteriores.

Em suma, a formulação do problema pode ser vista como uma regressão multivariada, na qual tem a intenção de minimizar uma função de erro, por exemplo, o MSE²⁵:

$$\min_f \sum_{t=1}^N \|y_t - f(y_{t-1}, y_{t-2}, \dots, y_{t-d})\|_2^2$$

Diversos métodos de estimação podem ser aqui usados, como mínimos quadrados, máxima verossimilhança, métodos de regularização como Ridge e Lasso, dentre outros métodos.

²⁴ $y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$ onde y_t é um vetor de observações no tempo t , ϵ_t é um vetor de erros no tempo t , e $\theta_1, \theta_2, \dots, \theta_q$ são os coeficientes do modelo MA, representando os efeitos dos termos de erro anteriores,

²⁵ Erro Quadrático Médio

4 MÉTODOS DE VALIDAÇÃO CRUZADA

A validação cruzada é uma técnica para avaliar um modelo estatístico ou de aprendizagem de máquina, testando sua performance. Essa avaliação é de grande importância na seleção automática e específica do modelo preditivo; em suma, a escolha dos hiperparâmetros que melhor generalizam tais dados. A técnica em questão consiste em dividir os dados em duas partes: uma de treino e uma de teste, treinar o modelo na primeira parte, validar na segunda parte e repetir esses passos diversas vezes.

Estas técnicas de reamostragem avaliam a habilidade de generalização e é uma forma de prevenir o modelo do sobreajuste — vício de aprendizagem causado nos dados de treino — Duda et al. (2000). Ademais, Berrar (2013), também indica que o sobreajuste é uma questão central no que tange a modelos supervisionados de aprendizagem de máquina e a necessidade de melhores resultados de acurácia advindos destes modelos. Como avaliar a capacidade de generalização do modelo? Utilizando novos dados provenientes da mesma população da qual os dados que foram utilizados para criar o modelo Simon (2003).

4.1 K-FOLD

A técnica em questão é um método que avalia e otimiza a performance de modelos de aprendizagem de máquina, o processo divide o conjunto de dados em k subconjuntos. Em cada iteração, uma subamostra é usada como teste e os demais dados como treino, assim, permitindo a seleção de hiperparâmetros que minimizam a soma dos erros quadráticos de previsão. Desta forma, é um método flexível para diversos modelos distintos.

Calcula-se a previsão em cada grupo para cada modelo:

$$\hat{y}_{(m)j}^{[-k]} = \int y f_m(y | \mathbf{X}_{(k-1) \times J+j}, \hat{\boldsymbol{\theta}}_{(m)}^{[-k]}) dy,$$

Avalia-se as previsões para todas as observações e para cada modelo, conforme :

$$\hat{\mathbf{Y}}_{(m)} = (\hat{y}_{(m)1}^{[-1]}, \dots, \hat{y}_{(m)J}^{[-1]}, \dots, \hat{y}_{(m)1}^{[-K]}, \dots, \hat{y}_{(m)J}^{[-K]}),$$

O critério para seleção dos hiperparâmetros é a minimização da função a seguir, neste caso o erro absoluto quadrático médio, porém outra métrica pode ser usada:

$$CV_k(\mathbf{w}) = \frac{1}{n} \|\mathbf{Y} - \hat{\mathbf{Y}}(\mathbf{w})\|^2,$$

As escolhas mais comuns para k , são entre 5 e 10, e quando $k = n$ temos a técnica leave-one-out¹

Zhang e Liu (2023) em artigo sobre o uso da validação cruzada k-fold — porém, não em séries temporais — para especificação errônea de modelos, demonstram que quando vários modelos são utilizados, e feito a predição média, é observado de forma assintótica que o risco de predição se torna o mínimo. A função risco, abordada no artigo, segue conforme :

$$R(\mathbf{w}) \equiv \mathbb{E} \left[\left(\hat{Y}_{n+1}(\mathbf{w}) - \mathbb{E}(Y_{n+1} | X_{n+1}) \right)^2 \right]$$

Para a convergência assintótica dos pesos de validação cruzada K-fold para minimizar o risco de previsão:

$$\frac{R(\hat{\mathbf{w}})}{\inf_{\mathbf{w} \in \mathbf{W}} R(\mathbf{w})} \rightarrow 1 \quad \text{em probabilidade,}$$

Para a convergência dos pesos para modelos corretamente especificados:

$$\tau \rightarrow 1 \quad \text{em probabilidade,}$$

Em se tratando de séries temporais, trabalhos do passado recente (2010 em diante) têm mostrado resultados promissores para utilização de validação cruzada em dados que possivelmente são sequenciais. Bergmeir e Benítez (2012) apresentaram um estudo comparativo em séries temporais estacionárias, e os resultados empíricos demonstrando evidência de que em tais condições, a técnica produz resultados mais acurados que os OOS².

Bergmeir et al. (2018) realizaram um estudo de simulação, comparando o k-fold padrão com OOS, utilizaram três processos de geração de dados e realizaram 1000

¹ procedimento que deixa apenas uma observação fora da amostra de treino, sendo a observação de teste

² Fora da Amostra, existe algumas estratégias diferentes dentro deste conceito.

rodadas de Monte Carlo em cada uma delas. Para cada rodada e processo generativo, uma série temporal estacionária é criada com 200 valores. Os resultados sugerem que, sistematicamente, a validação cruzada apresentou estimativas mais acuradas, indicando a correta especificação do modelo.

Cerqueira et al. (2020) fizeram um estudo de simulação aplicando a técnica k-fold e algumas variações da mesma, encontrando ótimos resultados de predição em séries estacionárias, indicando um melhor resultado do que o método OOS. A métrica de erro utilizada neste estudo foi o RMSE.

Raab et al. (2023) em um estudo espaço-temporal de predição em detecção de convulsão, aplicando modelos de aprendizagem profunda e aprendizagem de máquina para tal tarefa. Recorrendo ao k-fold de forma a encontrar os melhores hiperparâmetros de tais modelos, reportando 98,8% de acurácia na previsão dos eventos de interesse. Junaid et al. (2023) também estudando séries temporais, neste caso multimodais, e também para eventos da área da saúde, indicaram o uso do k-fold como forma de validação cruzada, no qual $k = 10$, como uma opção melhor de validação do que a opção "*leave-one-out*".

No âmbito das cadeias de suprimento e demandas de produtos, como é o caso da presente dissertação, tem-se em Taghiyeh et al. (2023), um estudo de *forecast* hierárquico, no qual também foi utilizada a técnica k-fold para previsão em modelos de aprendizagem de máquina, fazendo uso para otimização de hiperparâmetros. O artigo indica o uso da divisão dos dados em grupos cronológicos, garantindo que as previsões futuras sejam baseadas apenas em informações passadas, evitando vazamento de informações.

Lainder e Wolfinger (2022) fizeram uso do k-fold como validação cruzada, e duas variações da mesma técnica, focando na competição M5 Uncertainty. Em suma, o artigo fornece um caminho abrangente para o desenvolvimento de modelos de previsão de séries temporais de grande eficácia, destacando a importância da validação cruzada adequada, engenharia de recursos detalhada e técnicas de aumento para trabalhar com séries temporais de características diversas.

A seguir um exemplo do plano de validação:

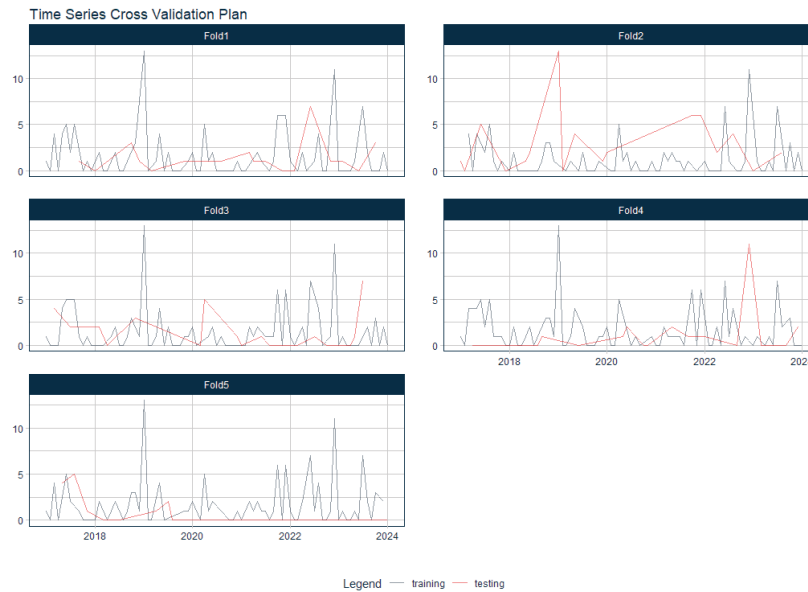


FIGURA 4 – Validação - K-fold - Exemplo

Fonte: Elaboração própria

O número de folds utilizados neste trabalho tem $k = 10$, seguindo uma gama de artigos que indicam este valor e inclusive Cerqueira et al. (2020), já explanado na comparação OOS e k-fold; Lines et al. (2018) também o usa em um estudo de classificação de séries temporais, Murray et al. (2022) utiliza-se da estratificação $k = 10$ em estudo comparativo de algoritmos de vanguarda para séries temporais. Ressalta-se que esta técnica é aplicada apenas em modelos não sequenciais.

4.2 JANELA EXPANSIVA

Neste trabalho, a técnica é aplicada para modelos sequenciais, ARIMA, NNETar, suavização exponencial, sendo uma técnica de validação cruzada muito comum para dados de série temporal. Neste tipo de validação, o conjunto de treinamento começa com um tamanho inicial e vai se expandindo, incluindo mais dados a cada iteração.

O conjunto de treinamento começa com um tamanho inicial e se expande ao incluir mais dados a cada iteração. O conjunto de teste é geralmente um único ponto no tempo ou um pequeno conjunto de pontos subsequentes.

De forma a obter uma estimativa de previsão robusta, Tashman (2000) recomenda aplicar esta estratégia em diversas janelas de teste. Cerqueira et al. (2020) fizeram um estudo, com base em Bergmeir et al. (2018), em séries temporais com no mínimo 500 observações e que não apresentaram valores faltantes, sendo um total distinto de 149 séries temporais. Em suma, neste estudo, o objeto é verificar qual método de estimação e validação melhor serve para diversos tipos de séries temporais, estacionárias ou não. Os autores reportaram que em dados sintéticos, os métodos de

validação cruzada, k-fold e modificações do método, funcionaram melhor do que os métodos OOS, com séries estacionárias; também, foi realizada a comparação com dados reais, obtendo resultados similares e indicando também as benesses pelo uso da validação cruzada. Se tratando de séries não estacionárias, as quais foram 77, observando fortes indicativos de que os métodos OOS — janela expansiva por exemplo — são melhores para estes casos.

Hewamalage et al. (2023) argumentam sobre a utilização de métodos mais orientados à ordem dos dados quando da não estacionariedade e indicam, através de uma revisão da literatura, que não há uma técnica ou uma abordagem única, dada a complexidade do problema; porém, destacam resultados positivos para o OOS quando no cenário discutido.

A seguir, uma representação do método, com seis avaliações cumulativas :

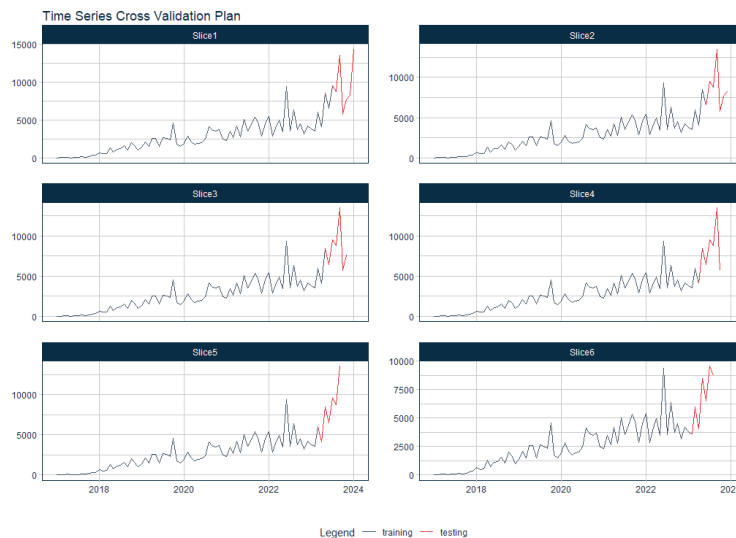


FIGURA 5 – Validação - Janela Expansiva - Exemplo
Fonte: Elaboração própria

4.3 GRID LATIN HYPERCUBE

Para realizar as validações cruzadas apresentadas, foi escolhido o Grid Latin Hypercube, o qual é uma técnica utilizada em experimentos de otimização e modelagem. Esta técnica combina duas outras, sendo a busca em *grid* e o Latin Hypercube.

O grid de busca é uma abordagem bastante tradicional na literatura — vide McKay et al. (1979), Stein e Boocher (2005) e Muehlenstaedt e Mattson (2006) , dentre outras referências —, para explorar o espaço de parâmetros de um modelo ou algoritmo. Esta técnica divide o espaço de busca com diversos pontos discretos para fazer as diversas avaliações.

Especificamente em McKay et al. (1979) temos a primeira proposta do método, sendo uma metodologia de geração de amostras, que tem o intuito de cobrir todo espaço amostral de forma a não gerar repetições; para que isso seja possível, dividi o intervalo de cada variável de interesse em N intervalos de mesma probabilidade, com base na distribuição escolhida para cada variável (por exemplo, uniforme, normal, etc.). O próximo passo reside em fazer uma amostra de cada intervalo, e as amostras são combinadas para formar um "hipercubo latino". Uma vantagem do LHS³ sobre a amostragem aleatória simples reside na garantia de uma cobertura mais uniforme do espaço de entrada. Isso pode reduzir significativamente a variância nas estimativas das propriedades estatísticas do modelo, como a média e a variância das saídas, com um número menor de simulações. Então, torna o LHS uma ferramenta útil em termos computacionais, ou seja, em situações onde as simulações são demoradas — o que é uma característica deste estudo, dada a quantidade das séries temporais por tipo de demanda.

Ademais, conceitos fundamentais do *Latin Hypercube* :

Divisão do Espaço de Entrada: Para cada variável de entrada X_i , o espaço de valor é dividido em N intervalos de mesma probabilidade, ou seja, uma variável no intervalo $\{0, 1\}$ pode-se subdividir em 5 intervalos iguais, por exemplo. Matematicamente, isso pode ser expresso pela função de distribuição cumulativa $F(x)$ da variável de entrada, onde cada intervalo $[a_j, b_j]$ é definido conforme:

$$F(b_j) - F(a_j) = \frac{1}{N} \quad \text{para } j = 1, 2, \dots, N$$

Seleção de Amostras: Para cada intervalo definido, uma amostra é selecionada. A seleção é feita de forma que cada intervalo contribua com exatamente uma amostra, garantindo que todas as partes da distribuição sejam representadas. Se x_i representa a amostra escolhida no i -ésimo intervalo, a seleção pode ser representada por uma função de distribuição inversa F^{-1} , que mapeia uniformemente distribuídos pontos de probabilidade P_i de volta ao espaço de entrada:

$$x_i = F^{-1}(P_i)$$

Essa equação garante que cada amostra x_i esteja representada dentro de seu respectivo intervalo de probabilidade.

Por exemplo, imaginemos uma demanda intermitente, que tenha quinze pontos — meramente hipotético —, conforme :

$$\text{demanda} = \{10, 20, 0, 0, 25, 100, 0, 0, 0, 10, 50, 0, 0, 0, 100\}$$

³ Latin Hypercube Sampling

E que esta demanda seja dividida em três grupos, sendo de baixa demanda, moderada e alta, conforme:

- Baixa demanda: demanda ≤ 10
- Demanda moderada: $10 < \text{demanda} \leq 50$
- Alta demanda: demanda > 50

O próximo passo será calcular a função de distribuição acumulada desta série temporal de demanda.

- Baixa demanda: $P_1 = \frac{\text{número de valores de demanda} \leq 10}{\text{total de valores de demanda}}$
- Demanda moderada: $P_2 = \frac{\text{número de valores de demanda entre 10 e 50}}{\text{total de valores de demanda}}$
- Alta demanda: $P_3 = \frac{\text{número de valores de demanda} > 50}{\text{total de valores de demanda}}$

Subsequente, vamos calcular os pontos de probabilidade P_i . É possível usar a função de distribuição inversa para fazer o mapeamento inverso, tendo os valores de demanda correspondentes em cada intervalo.

- Baixa demanda:

$$P_1 = \frac{8}{15} = \frac{8}{15}$$

Para a baixa demanda, os valores de demanda são: 0, 0, 0, 0, 0, 0, 0, 0

- Demanda moderada:

$$P_2 = \frac{4}{15} = \frac{4}{15}$$

Para a demanda moderada, os valores de demanda são: 10, 20, 25, 10

- Alta demanda:

$$P_3 = \frac{3}{5}$$

Para a alta demanda, os valores de demanda são: 100, 100, 50

O próximo passo consiste em usar a função de distribuição inversa para mapear esses espaços.

- Baixa demanda:

$$x_1 = F^{-1} \left(\frac{8}{15} \right) = \text{valor de demanda correspondente}$$

- Demanda moderada:

$$x_2 = F^{-1} \left(\frac{4}{15} \right) = \text{valor de demanda correspondente}$$

- Alta demanda:

$$x_3 = F^{-1} \left(\frac{3}{5} \right) = \text{valor de demanda correspondente}$$

Dessa forma, podemos obter amostras representativas de cada intervalo de demanda na série temporal proposta.

Permutação das Amostras: De forma a evitar a correlação entre as variáveis, uma permutação aleatória é aplicada aos pontos de amostra selecionados para cada variável. Isso garante que, ao construir o conjunto final de pontos de amostra no espaço de entrada multidimensional, cada combinação de intervalos entre as variáveis seja única. A permutação pode ser representada por uma função π_i aplicada ao conjunto de índices dos intervalos para cada variável i , conforme:

$$\pi_i : \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$$

Essa função é uma bijeção, garantindo que cada intervalo seja mapeado para exatamente outro intervalo, sem repetições.

Quanto à utilização da técnica, observamos na literatura uma gama de artigos aplicando tal técnica, Hill et al. (2023) utilizaram-se da técnica para sintonia de modelos baseados em árvore em pesquisa sobre a doença de Parkinson, especificamente, foi utilizado para encontrar a melhor combinação de hiperparâmetros usando 10 *resamples* de validação cruzada; o objetivo era otimizar os hiperparâmetros durante o treinamento, de forma a melhorar a precisão e eficácia dos modelos. Wang et al. (2023) trabalham no desenvolvimento de um modelo de otimização em dois níveis para a seleção de membros de uma planta de energia virtual, o método *Latin Hypercube Sampling (LHS)* foi utilizado para gerar um grande número de subsequências, ajudando no processo de otimização dos membros da planta de energia; os autores destacam o benefício de utilizar o método com programação inteira, sendo que o modelo proposto conseguiu

aumentar a receita média diária da planta em 12,806 RMB e elevar a proporção de geração de energia nova em média 1,8%.

Wapachi e Diab (2023) em um estudo de séries temporais para utilização em tempo real em uma usina nuclear, neste estudo foi destacada a eficácia dos modelos de aprendizado de máquina, especificamente RNNs, na previsão da resposta de uma NPP em condições de acidente, criando uma ferramenta para suportar a tomada de decisão rápida e reduzir custos computacionais. A inclusão da técnica Latin Hypercube na análise permitiu uma exploração eficiente do espaço de incertezas, contribuindo para a geração de uma base de dados abrangente e representativa para o treinamento dos modelos de ML. Aqui o método LHS foi utilizado em conjunto com Monte Carlo, para amostrar parâmetros incertos até que uma amostra estatisticamente significativa fosse alcançada.

4.4 MÉTRICAS DE ERRO

As métricas de erro são um tema bastante explorado na estatística, matemática, ciências econômicas, dentre outras áreas do saber; de forma a avaliar sistemas, modelos estatísticos, propostas de mudanças no "status quo". Começando pelo MAE - Mean Absolute Error:

$$\text{MAE} = \frac{1}{M} \sum_{t=1}^M |\hat{F}_t - F_t|$$

Sendo uma das métricas mais fáceis de se interpretar e computar, contudo, impossibilita a comparação entre séries temporais distintas; sendo um erro dependente da escala, Schaer et al. (2019) em um estudo de projeção de demanda para artigos eletrônicos - video games, aplicando modelos ETS, ARIMA com variáveis exógenas e sem variáveis exógenas, fez a comparação entre os modelos utilizando a métrica em questão e AvgRelMAE, que será discutida no decorrer da seção.

A métrica seguinte - RMSE, Root Mean Squared Error, muito utilizada também no âmbito da predição macroeconômica e não apenas da previsão em geral, é basicamente o MAE, mas leva em consideração os desvios quadráticos e, após isso, é feita a raiz quadrada:

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{t=1}^M (\hat{F}_t - F_t)^2}$$

Sendo também uma medida de erro absoluto, é notado que o RMSE é uma medida sensível a *outliers*, dado que o quadrado dos desvios tende a amplificar o impacto como um todo. Então, é sempre importante identificar o contexto dos dados.

Chai e Draxler (2014) argumentam sobre a vantagem de se usar o RMSE em detrimento do MAE, pois valores absolutos são bastante indesejáveis em diversos campos da matemática. Ademais, os autores indicam que dar maior peso a observações discrepantes pode ser melhor para diferenciar a performance entre modelos. Ainda existe a discussão sobre a importância de verificar mais de uma métrica de erro, pois uma métrica acaba enfatizando um aspecto do erro, enquanto métricas diferentes, combinadas, tendem a captar uma gama maior da característica dos erros.

A métrica de erro MASE - erro médio absoluto escalado, tem sua grande vantagem no cerne de poder comparar diferentes séries temporais. Essa métrica foi proposta por Hyndman e Koehler (2006), e é obtida escalando os desvios de previsão baseados na métrica de erro, MAE, da amostra de teste.

Kim e Kim (2016) fazem uma comparação entre métricas de erro sob a ótica de demandas intermitentes. A análise é feita em dados semanais de vendas de lubrificantes industriais - falarei mais sobre o caso dentro das métricas MAPE.

$$\text{MASE} = \frac{\frac{1}{T} \sum_{t=1}^T |e_t|}{\frac{1}{N-1} \sum_{i=2}^N |F_i - F_{i-1}|}$$

Outra métrica de erro bastante discutida em demandas mais problemáticas, como intermitentes e inflacionadas de zero, é a MAPE - erro médio percentual absoluto:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

Essa métrica expressa a magnitude média dos erros em termos da porcentagem - sendo independente de escala - do valor real; onde A_T é o valor observado e F_T o valor predito pelo modelo. Contudo, as revisões literárias indicam problemas com esta métrica de erro quando a avaliação se dá em valores muito próximos de zero. Kim e Kim (2016) indicam uma desvantagem significativa desta métrica, já comentada ligeiramente anteriormente, sobre os valores muito próximos de zero; ressaltando que em inúmeras áreas de pesquisa os dados contêm muitos zeros ou valores próximos, inclusive nas áreas de supply chain.

Nesta linha, uma modificação da métrica foi proposta por Makridakis (1993), o sMAPE - sendo o MAPE, porém, simétrico. Ademais, sua expressão se dá por:

$$\text{sMAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \times 100$$

Hyndman e Koehler (2006) indica que o MAPE dá maiores pesos para os erros positivos do que para os negativos, então o sMAPE vem para corrigir esse problema, contudo, essa métrica ainda envolve divisão por zero e o valor pode ser negativo, gerando uma interpretação ambígua. Ao longo dos anos houve bastante discussão acerca desta métrica, e diferentes autores acabam por usar diferentes variações da métrica, Hyndman (2014).

$$\text{MAAPE} = \frac{1}{n} \sum_{i=1}^n \left(\arctan \left(\left| \frac{F_i - \hat{F}_i}{F_i} \right| \right) \right)$$

A métrica de erro elencada anteriormente — erro percentual médio de arco tangente absoluto, é uma modificação, também, do MAPE. Também tem o intuito de mitigar algumas falhas já elencadas da métrica da qual advém. Desenvolvido por Kim e Kim (2016), os autores avaliaram a métrica em dois *datasets* reais e um simulado, obtendo diversos resultados, entre eles que o indicador em questão comparado com sMAPE, MAPE, MAE e MASE, tem resultados mais robustos na presença de zeros na série temporal e também é mais robusto a *outliers*. Por fim, em uma série de testes envolvendo dados do varejo americano, os autores concluem que a métrica proposta é mais parcimoniosa e consegue driblar os problemas do indicador antecessor.

A métrica de erro AvgRelMAE, média do Erro Absoluto Médio Relativo, advém da RelMAE, conforme a seguir:

$$\text{RelMAE}_i = \frac{|F_i - \hat{F}_i|}{F_i}$$

A métrica em questão utiliza-se do MAE fora da amostra, como um fator de escala. O AvgRelMAE indica melhorias em termos de funções de perda linear e quadrática; permitindo zeros e erros bastante próximos de zero, de previsões individuais e pode ser usada em uma gama grande de configurações, como tipos de demanda comentados anteriormente, na seção 2,8.

$$\text{AvgRelMAE} = \frac{1}{N} \sum_{i=1}^N \frac{|F_i - \hat{F}_i|}{F_i}$$

5 RESULTADOS

Neste capítulo será apresentada a engenharia de variáveis, um exemplo sobre a demanda de 6 séries temporais de peças de reposição de forma ilustrativa, e por fim, os resultados para os três tipos de séries temporais aqui trabalhados: Suave, Errático e Intermitente. Os resultados serão demonstrados na seguinte sequência:

1. Uma tabela resumo do Rank dos modelos por 4 métricas de erro.
2. Uma tabela sobre a contagem das vezes que o modelo foi o melhor.
3. Escolha dos três melhores modelos por contagem.
4. Análise dos hiperparâmetros dos melhores modelos, segregados por:
 - Séries temporais estacionárias ou não.
 - Séries temporais com autocorrelação ou não.

A análise dos hiperparâmetros incluirá uma visualização gráfica da densidade conjunta de dois hiperparâmetros para cada modelo, visando identificar variações na densidade entre diferentes tipos de séries temporais, com foco nas características de estacionariedade e autocorrelação. Esta abordagem facilita a identificação das melhores configurações de hiperparâmetros. A avaliação será complementada por uma tabela resumo que apresentará a média, o desvio padrão e a quantidade de séries temporais, segregadas por faixa de inflação de zeros no caso de séries intermitentes, e por coeficiente de variação nos demais casos. Adicionalmente, gráficos de boxplot e densidade, também segregados pelos tipos de séries temporais, estarão disponíveis no apêndice A, oferecendo uma análise granular da densidade conjunta.

A estrutura é iniciada com as intermitentes, as mais difíceis de se modelar, e subsequentemente as suaves e erráticas que têm características semelhantes. Sendo 2850 séries temporais intermitentes, 1982 séries temporais suaves e 1988 séries erráticas.

O processamento foi realizado em um ambiente Databricks na nuvem, utilizando 80 núcleos de processamento. Cada tipo de série temporal demandou cerca de 8 horas de processamento, devido à busca exaustiva no grid de hiperparâmetros, com aproximadamente 100 combinações testadas por modelo. O objetivo principal dessa busca intensiva é identificar as regiões do grid onde se concentram os melhores resultados, permitindo reduzir significativamente o espaço de busca em futuras otimizações.

5.1 TÉCNICAS DE ENGENHARIA DE VARIÁVEIS

As transformações das variáveis explicativas são passos essenciais para a tarefa de previsão, usando modelos de aprendizagem de máquina ou modelos puramente estatísticos. A discussão acerca dessas técnicas já tem mais de 50 anos, como em Box e Jenkins (1970) a qual é uma referência fundamental na área de séries temporais e aborda diversas técnicas para modelagem e previsão de séries temporais, incluindo métodos de transformação de variáveis e engenharia de características. Em Granger (1969) o economista Clive Granger introduziu o conceito de causalidade temporal e métodos para avaliar e investigar relações causais entre séries temporais. Contudo, o foco principal sendo em causalidade, o trabalho de Granger também discute técnicas para lidar com séries temporais, incluindo transformações de variáveis que podem ser relevantes para a engenharia de características.

Recentemente, após todos os avanços ocorridos nos últimos 50 anos, observamos diversos usos destas técnicas. A atividade de *forecasting* em muitos passos à frente por diferentes técnicas de aprendizagem de máquina tem sido atividade bastante explorada na literatura, vide Hond e Fan (2016), Chen et al. (2018), Nie et al. (2012). Ademais, Ahmadi et al. (2023b), faz um estudo usando modelos estatísticos e de aprendizado profundo, com técnicas MIMO e iterativas de previsão, sendo também focado em muitos passos à frente.

Em se tratando das técnicas para transformação das variáveis :

5.2 TRANSFORMAÇÕES DE ESCALA:

- Logaritmação: Reduz a assimetria e a heterocedasticidade da série.
- Raiz quadrada: Reduz a assimetria e estabiliza a variância da série.
- Transformação Box-Cox: Permite uma transformação mais flexível da série.

DECOMPOSIÇÃO DE SÉRIES TEMPORAIS:

- Decomposição aditiva: Divide a série em seus componentes sazonal, cíclico e residual.
- Decomposição multiplicativa: Divide a série em seus componentes sazonal, cíclico e de tendência.

5.3 CRIAÇÃO DE LAGS:

- Lags simples: Inclui valores passados da série como *features*.
- Lags estacionários: Ajusta os lags para remover a autocorrelação da série.

5.4 ESTATÍSTICAS DE JANELA:

- Média móvel: Suaviza a série e identifica tendências.
- Desvio padrão móvel: Identifica períodos de alta e baixa volatilidade.
- Outras estatísticas: Kurtosis, skewness, etc.

EXTRAÇÃO DE FEATURES ESPECÍFICAS DO DOMÍNIO (DEMANDA):

- Fatores climáticos
- Fatores econômicos

Estes são algumas extrações possíveis e tratamentos de variáveis, dentre uma gama maior de possibilidades. Zou et al. (2023) Em estudo sobre previsão no mercado acionário, discutiu-se diferentes sub-tarefas de previsão do mercado de ações, dentre elas, a previsão do movimento de ações, a previsão de preços de ações, gerenciamento de portfólio e estratégias de negociação; neste artigo é mencionado o uso da engenharia de variáveis, sendo intrínsecas, como preços, os valores passados das séries estudadas, índices financeiros e extrínsecas, como o uso de dados textuais, variáveis macroeconômicas. Sendo transformadas e usadas de várias maneiras, dependendo do modelo utilizado neste artigo. Tuli et al. (2022) Segue uma abordagem de engenharia de características no TranAD e é fundamentalmente ligada à sua capacidade de processar séries temporais, considerando tanto características intrínsecas quanto extrínsecas dos dados. O modelo utiliza técnicas avançadas para capturar a dependência temporal e a correlação entre diferentes variáveis de entrada, sem necessariamente introduzir um processo manual de engenharia de características. Barandas et al. (2020) Temos um trabalho mais voltado a software e bibliotecas, especificamente em Python — denotando a importância e utilização destas técnicas e como uma ferramenta automática melhora o desempenho dos estudos; a ferramenta em questão calcula mais de 60 recursos diferentes abrangendo domínios temporais, estatísticos e espectrais; ademais, a ferramenta trabalha em três principais domínios : temporal, estatístico e espectral.

A presente dissertação aplicou as seguintes transformações de variáveis para as tarefas de extração das métricas de erro e previsão:

TABELA 5 – Engenharia de variáveis para previsão de y_t

Variável de Engenharia	Descrição
Lag de $\log(y_t)$ com lag 6	Variável $\log(y_{t-6})$
Média Móvel 3 de Lag de $\log(y_t)$	Média móvel de $\log(y_t)$ com janela de 3 períodos
Média Móvel 12 de Lag de $\log(y_t)$	Média móvel de $\log(y_t)$ com janela de 12 períodos
Fourier Seno	Padrões sazonais senoidais
Fourier Cosseno	Padrões sazonais cossenoidais
Dummies Sazonais	Variáveis indicadoras de sazonalidade

A transformação log não é empregada no caso Glnnet Poisson.

Sendo a tabela uma descrição da metodologia empregada nas séries temporais das três características: suave, errática e intermitente. O operador de retardo em seis períodos advém também do horizonte de previsão o qual é seis períodos à frente; contudo, a primeira variável da tabela não é utilizada apenas para as médias móveis. International Monetary Fund (2018) realizou uma pesquisa com previsões em dados econômicos, de 1990 até 2016, indicando que previsões mais curtas para este tipo de dado podem ser mais precisas uma vez que a mudança do cenário econômico não é tão provável, contudo, para cenários mais longos, há uma percepção de superestimação nas previsões, atribuída à natureza imprevisível da ciência econômica, ainda sobre o horizonte de previsão. Nikolopoulos (2020) em um estudo sobre demandas intermitentes — as mais difíceis abordadas neste estudo —, indicam por um lado a necessidade da indústria, em termos de inventário e os problemas logísticos, ou seja, a adaptação do horizonte de previsão dadas as necessidades e barreiras do mundo real, concluindo, em seu escopo, por horizontes menores, sendo passíveis de vieses menores e também erros; em específico, neste estudo, as previsões de curto prazo são de duas semanas e longo prazo, de até dois anos.

Neste ponto, a escolha da previsão com janela de seis meses foi decidida pela necessidade do processo da cadeia produtiva e que permitisse avaliar uma janela de tamanho médio, quando se tratando da deterioração de métricas de erro empregadas e suas comparações.

A seguir, uma tabela descritiva, como exemplo, de duas séries temporais por cada um dos três tipos de característica da série temporal — demanda.

Começando pelas intermitentes :

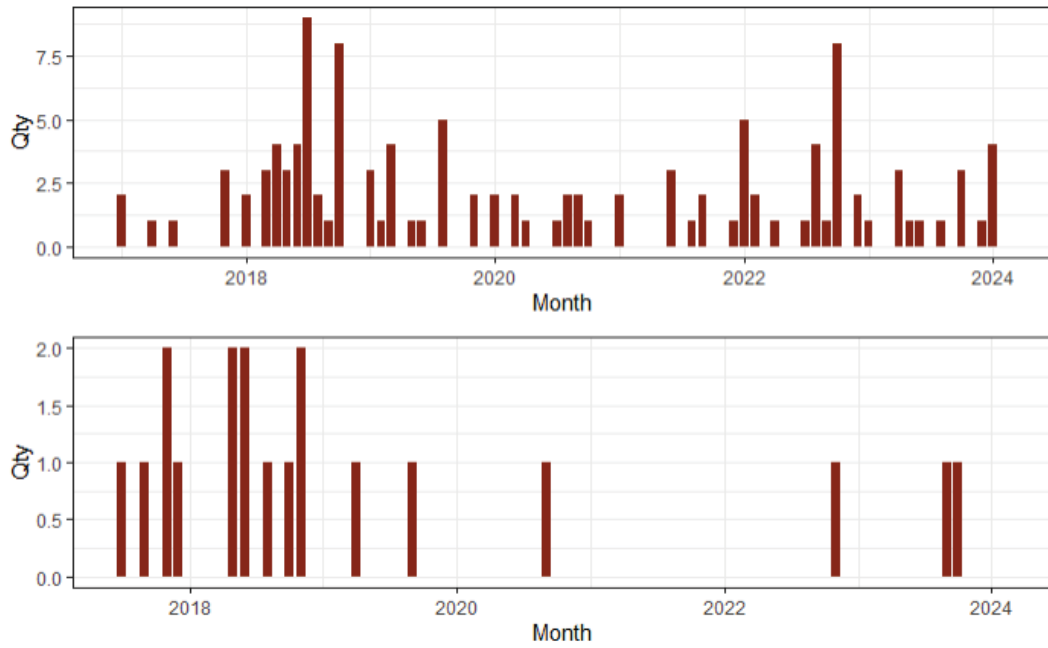


FIGURA 6 – Demanda - Intermitente
Fonte: Elaboração própria

Erráticas:

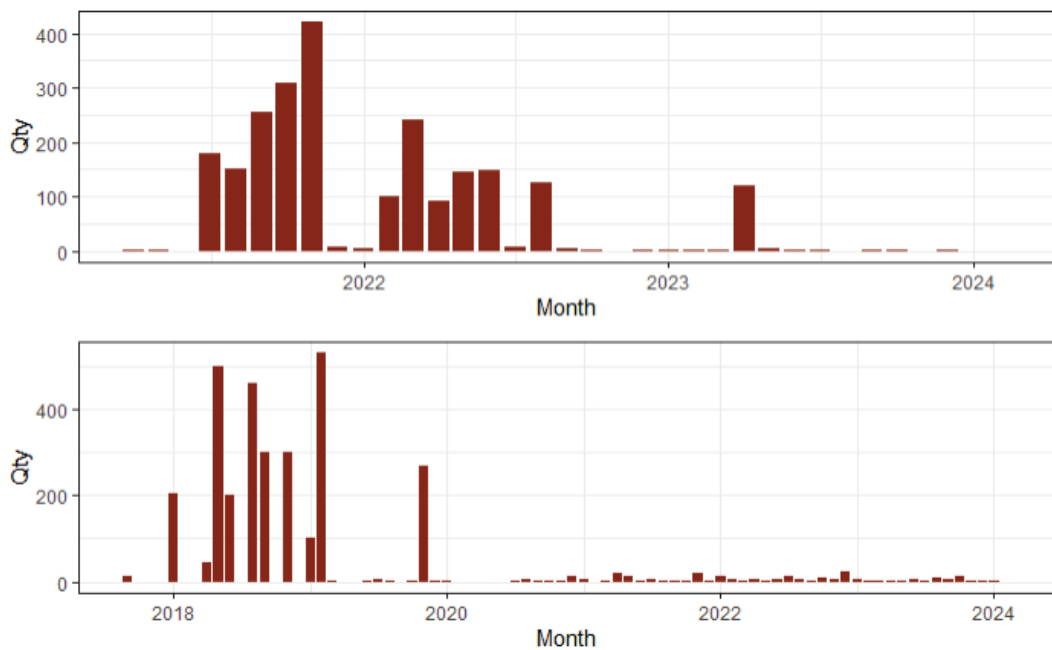


FIGURA 7 – Demanda - Errática
Fonte: Elaboração própria

Suaves:

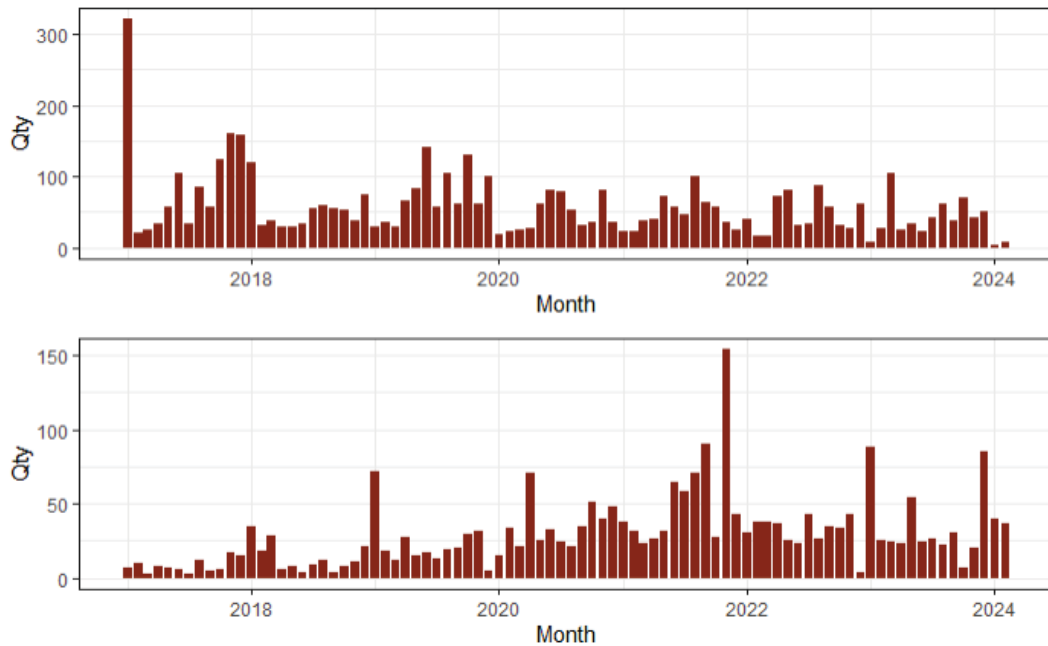


FIGURA 8 – Demanda - Suave

Fonte: Elaboração própria

Tabela resumo:

Peça	Tipo de Demanda	Média	D,padrão	Mediana	Coef,Var	Perc,zero
1	Errática	40,73	111,83	3,0	2,75	0,23
2	Errática	66,57	106,91	2,0	1,61	0,17
3	Intermitente	1,38	1,89	1,0	1,37	0,44
4	Intermitente	0,24	0,53	0,0	2,21	0,81
5	Suave	29,01	24,03	24,5	0,83	0,00
6	Suave	56,95	44,25	45,0	0,78	0,00

TABELA 6 – Descritivo das Peças

Resumo sobre características da série temporal por tipo de demanda

Por fim, existe uma diferença clara advinda da segregação dos tipos de demanda. Enquanto as intermitentes e erráticas, escolhidas randomicamente na amostra, têm algum percentual de zeros em suas séries temporais, as suaves apresentam zero ocorrências; além disso, apresentam maior variabilidade, notadamente pelos seus coeficientes de variação. Olhando mais a fundo, também é perceptível a diferença entre intermitência e errática, quanto ao percentual de zeros, sendo a primeira mais incidente; ademais, à primeira vista, a suave parece a mais amigável para ser modelada, uma vez que não constam zeros e seus desvios padrão são menores que a média.

5.5 INTERMITENTES

Séries temporais com a característica de intermitência apresentam períodos com zeros, muitas vezes seguidos, como visto na tabela 6, sendo o exemplo de apenas duas das 2850 séries temporais intermitentes aqui avaliadas.

TABELA 7 – Rank Médio - Intermitentes

Técnica	Modelo	Rank_smape	Rank_mae	Rank_rmse	Rank_mase
Mimo	Glmnet - Poisson	4,64	9,19	8,69	8,36
Mimo	SVR	5,59	8,61	7,77	7,84
Híbrido	Arima - SVR	7,10	8,87	7,94	8,54
Mimo	Random Forest	7,39	7,96	7,64	7,75
Mimo	Glmnet	7,41	7,22	6,63	7,22
Mimo	Mars	7,82	7,90	7,13	7,81
Híbrido	Arima - Xgboost	7,83	8,86	9,38	8,70
Iterativo	Arima - benchmark	7,97	7,51	6,90	7,62
Híbrido	Arima - GLM	8,05	7,27	6,66	8,03
Híbrido	Arima - RF	8,17	7,46	7,19	8,07
Híbrido	Arima - Mars	8,21	7,49	6,99	8,22
Mimo	XGboost - tweedie	8,82	6,58	7,22	6,88
Mimo	XGboost	9,16	8,89	10,23	8,57
Híbrido	NNAR	9,99	9,99	11,18	9,46
Mimo	NNET	11,40	5,77	8,06	6,49

A tabela 7 demonstra o rank médio por métrica de erro, e está ordenada pelo menor Smape — métrica que reduz e tenta controlar as diferenças de escala, a parte simétrica, refere-se ao denominador da métrica —, temos que para estas séries temporais com uma maior representatividade de zeros, o modelo elastic net com distribuição poisson, teve o melhor rank, seguido pelo SVR e o híbrido Arima SVR. Contudo, na tabela 8, temos que mesmo o rank médio do poisson ser o melhor, não necessariamente este foi o melhor na contagem das séries temporais,

TABELA 8 – Modelo Campeão - Intermitentes

Modelo	Rank_smape	Percentual
Arima - SVR	600	0,209
Glmnet - Poisson	480	0,167
SVR	366	0,127
Arima - Xgboost	352	0,122
Arima - RF	201	0,07
NNAR	172	0,060
XGboost	172	0,060
Arima - Mars	106	0,037
Random Forest	100	0,035
Arima - GLM	97	0,034

Como visto na tabela 8, os três mais representativos por contagem de séries temporais são: Arima SVR, um modelo híbrido, Glmnet com distribuição Poisson e SVR — Support Vector Regression; com suas respectivas contagens, sendo o primeiro modelo, campeão em 589 das 2875 séries temporais intermitentes avaliadas. O mesmo

segue para os dois outros modelos campeões. Ademais, a seguir, as séries temporais separadas por estacionariedade e autocorrelação:

TABELA 9 – Estrutura Temporal - Intermitentes

Estacionária	Autocorrelacionada	Contagem
Não	Sim	319
Não	Não	505
Sim	Sim	358
Sim	Não	1668

Por fim, como esse tipo de série temporal apresenta grande proporção de zeros, em muitos casos, a seguir a representação da distribuição de zeros em todas as 2850 séries temporais :

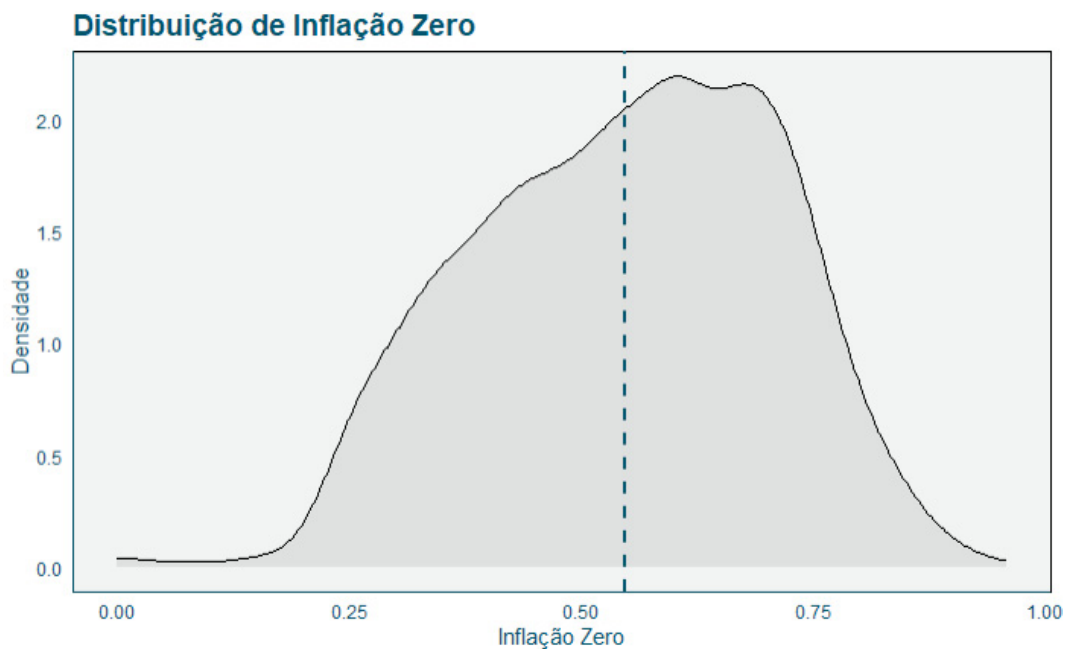


FIGURA 9 – Distribuição de zeros - Intermitente

Fonte: Elaboração própria

Como denotado pela linha tracejada, a média de zeros reside em cerca de 54%, ou seja, em séries temporais mensais, que se iniciam no primeiro janeiro de 2017 e se encerram em dezembro de 2022, em média, com 54% dos meses apresentando ausência de demanda.

5.5.1 Arima - SVR

Sendo uma proposta híbrida, composta pela modelagem Arima na parte linear e o modelo SVR para os resíduos, a parte não linear. A seguir, serão apresentados os resultados para os hiperparâmetros do modelo de aprendizagem de máquina, e

pela segregação da quantidade de zeros em cada série temporal, se é uma série estacionária ou não, e se é autocorrelacionada ou não.

Começando pela densidade dos hiperparâmetros de Custo e Rbfsigma, do modelo SVR, com kernel gaussiano, Ambos os eixos estão em log base 10. A seguir, análise gráfica da densidade conjunta dos parâmetros, para os casos de séries temporais, estacionárias e não estacionárias:

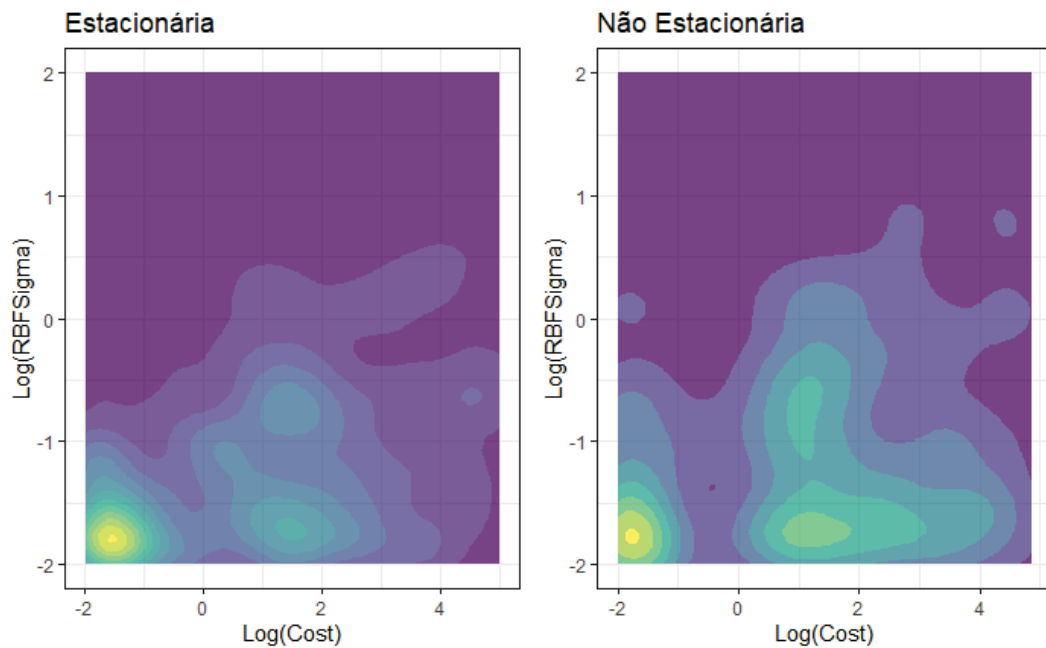


FIGURA 10 – Densidade Conjunta - Arima SVR - Estacionariedade - Intermitente

A seguir, densidade conjunta dos parâmetros Cost e RbfSigma, para os casos de séries temporais, Não autocorrelacionadas e Autocorrelacionadas:

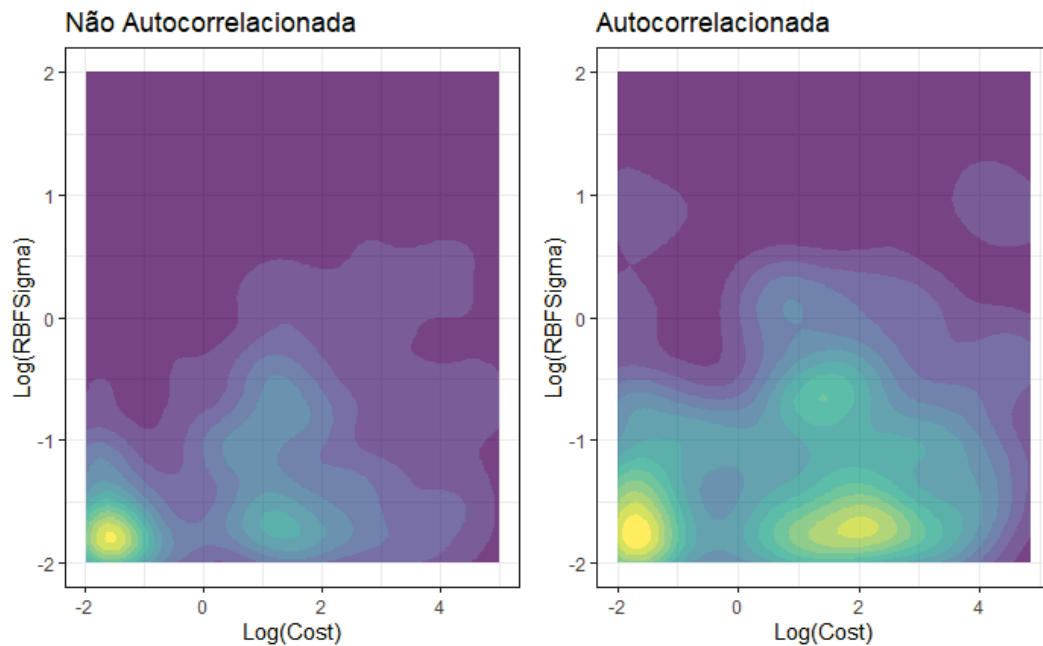


FIGURA 11 – Densidade Conjunta - Arima SVR - Autocorrelação - Intermitente

Em ambos os casos, aparenta uma concentração perto dos valores -2 , ou seja $10^{(-2)} = 0,01$, cerca de $0,01$, em suma, valores pequenos para os hiperparâmetros do modelo. Contudo, é perceptível diferenças entre os casos estacionários e não estacionários, quando da incidência do parâmetro Cost, próximos ao valor 2 e a expressividade das curvas de nível, as quais são um tanto diferentes - com áreas de incidência maiores em autocorrelação e não estacionariedade. Ademais, comparando os casos de não autocorrelação e autocorrelação, também, demonstra uma certa bimodalidade entre $\{-2, -2\}$ e $\{-2, 2\}$, ou seja, o parâmetro de custo em valores mais expressivos, ocorrendo quando a série temporal tem comportamento autocorrelacionados ou não estacionários. Por fim, em termos de variabilidade, os gráficos demonstram essa característica com maior força nos casos autocorrelacionados e não estacionários, com as áreas de densidade sendo maiores e mais espaçadas que seus pares estacionários e não autocorrelacionados,

A seguir, será analisada a variabilidade dos hiperparâmetros quando distribuídos pela inflação de zeros das séries temporais, uma vez que, em se tratando de séries intermitentes, como visto pelo gráfico 11, tem-se que a representatividade de zeros é grande. Para tal, foi utilizada a separação por quantis, para segregação do percentual de zeros em faixas.

5.5.1.1 Parâmetros SVR - Híbrido

Os gráficos contendo boxplot e densidade dos parâmetros são encontrados no apêndice A. A síntese a seguir será apresentada no formato tabela, com ambos parâmetros do modelo SVR e kernel Rbf, sendo Cost e Rbf, Inflação de zeros, desvio

padrão (SD) e média do logarítmo base 10, e por fim a quantidade de séries temporais em cada segregação (Qtd).

TABELA 10 – Parâmetros SVR - Autocorrelação - Intermitente

Inflação	Autocorrelação?	Sd - Rbf	Media - Rbf	Sd - Cost	Média - Cost	Qtd
0-0,5	Sim	0,89	-1,01	2,01	0,83	58
0-0,5	Não	1,67	-0,55	2,06	0,69	98
0,51-0,59	Sim	0,75	-1,12	2,12	0,70	33
0,51-0,59	Não	1,46	-0,89	2,10	0,35	111
0,60-0,69	Sim	0,82	-1,17	1,98	1,35	28
0,60-0,69	Não	1,05	-1,09	1,95	0,73	115
0,70-0,94	Sim	0,87	-0,93	2,06	1,12	19
0,70-0,94	Não	0,87	-0,93	1,92	1,40	131

A tabela 10 faz a síntese de escolha dos parâmetros Cost e Rbfsigma que reduziram o erro de treinamento, indicando diferenças mais fortes de média entre os casos de menor inflação de zeros quando das médias dos hiperparâmetros (0 - 0,59), sendo os casos de autocorrelação para cost com valores maiores, por exemplo $10^{0,83}$ e para sigma o inverso em $10^{-1,01}$; mais comentários sobre os resultados no próximo parágrafo.

A seguir, o kernel RbfSigma:

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (5.1)$$

Este kernel é conhecido por sua capacidade de trabalhar com dados de características não lineares,

- Rbf: A média do desvio padrão deste parâmetro é sobremaneira menor nas séries com autocorrelação, indicando que a presença de autocorrelação permite uma escolha, talvez, mais estável e menos dispersa — se tratando de menores inflações de zeros. Sugerindo que o modelo se ajusta melhor e com menos variabilidade quando há autocorrelação.
- Cost: Os desvios padrão são bastante próximas entre os grupos, porém a média no grupo sem autocorrelação tende a ser menor, sugerindo que o modelo dexaria mais erros ocorrerem, uma característica de maior generalização.

Em suma, a presença da autocorrelação parece trazer uma maior estabilidade para o grid do parâmetro RBF, o que pode resultar em um modelo que se adapta melhor às características dos dados. Em séries com a característica de ausência de

autocorrelação, a escolha parece ser mais difícil, tendo maior variabilidade em RBF. Ademais, nos casos em que a média de RBF é mais baixa, a média do parâmetro custo é maior.

Seguindo para a característica de estacionariedade, e sua dupla natureza. Bem como no caso de autocorrelação, é notado que para o caso de ausência de autocorrelação a variabilidade é menor, quando da análise gráfica, em comparação ao caso autocorrelacionado.

TABELA 11 – Parâmetros - Estacionariedade - Intermitente

Inflação	Estacionária?	Sd - Rbf	Media - Rbf	Sd - Cost	Média - Cost	Qtd
0-0,5	Não	0,92	-0,98	2,11	1,28	44
0-0,5	Sim	1,59	-0,62	1,98	0,53	112
0,51-0,59	Não	1,71	-0,71	2,04	0,18	28
0,51-0,59	Sim	1,22	-1,00	2,12	0,49	116
0,60-0,69	Não	0,89	-0,99	2,13	0,99	40
0,60-0,69	Sim	1,05	-1,15	1,90	0,80	103
0,70-0,94	Não	0,82	-0,98	1,95	1,40	50
0,70-0,94	Sim	0,90	-0,90	1,93	1,34	100

A tabela 11 demonstra que em casos não estacionários, temos menores variações do parâmetro sigma em Rbf, em 3 dos quatro cenários de inflação de zeros. Já nos casos estacionários, de natureza mais previsível, principalmente para menor faixa de inflação de zeros, permitiu uma distinção mais clara do parâmetro Cost, indicando que para este caso estacionário, temos $10^{0,53}$, indicando uma maior tolerância a erros; ademais, os desvios padrão deste parâmetro são bastante homogêneos e grandes, refletindo o gráfico de densidade e também nos casos do parâmetro de kernel.

- Rbf: Séries não estacionárias geralmente apresentam um desvio-padrão menor para o parâmetro RbfSigma, sugerindo menor variabilidade, exceto no intervalo de 0,5-0,59 onde a série não estacionária tem um desvio-padrão maior.
- Cost: As séries não estacionárias parecem ter uma maior variabilidade no parâmetro Cost, sugerindo que a adaptação do modelo a essas séries pode ser mais complexa.

5.5.1.2 Comentários Finais - ArimaSVR

Ambas as características das séries temporais, autocorrelação e estacionariedade, demonstram um certo grau de escolha de parâmetros para otimizar a captura destes padrões nos dados; aqui observamos algumas diferenças mais acentuadas quando da presença de autocorrelação e o comportamento do parâmetro sigma RBF

e também no parâmetro Cost. Enquanto a autocorrelação pede um modelo mais penalizativo e menos sensível, vide as médias de Cost, a estacionariedade permite um modelo mais sensível com penalidades menores. Essa dicotomia observada salta aos olhos quando desta análise, fazendo uso destes dados temporais com a técnica MIMO de predição.

5.5.2 Glmnet - Poisson

Aqui temos a proposta MIMO, composta pela modelagem por meio do modelo Glmnet com distribuição de probabilidade Poisson, fazendo uso da técnica de previsão : MIMO. Sem diferenciação ao item anterior, a análise será apresentada seguindo a mesma lógica.

Começando pela densidade dos hiperparâmetros de mixture e penalty. Ambos os eixos estão na escala natural, sem alteração, uma vez que esses parâmetros variam entre 0 e 1.

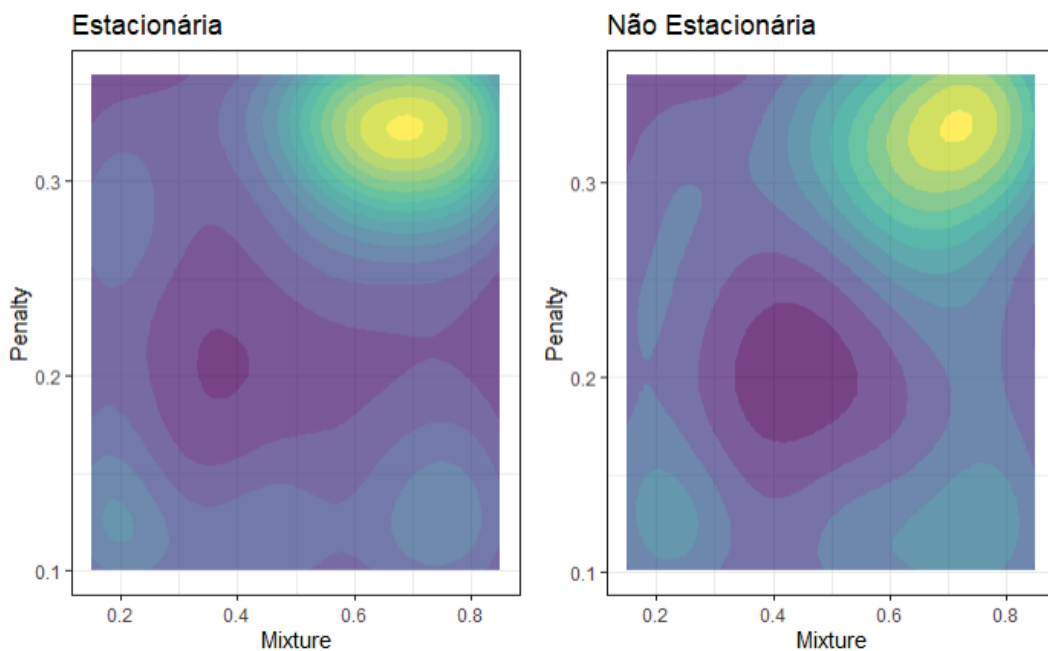


FIGURA 12 – Densidade Conjunta - Glmnet Poisson - Estacionariedade - Intermitente
Fonte: Elaboração própria

Em ambos os casos, denotado pelo gráfico 12, há uma concentração do parâmetro mixture entre 0,6 e 0,8, se aproximando mais da penalização L1 (lasso) — sendo a área superior direita de maior densidade em ambos os casos. A diferença mais pujante entre a estacionária e a não estacionária reside no fato de a estacionária ter maior variabilidade e uma densidade em torno de 0,1 do parâmetro penalty e 0,2 de mixture, mais expressiva que no caso não estacionário; contudo, este último tem uma área de ocorrência maior que o anterior quando do mixture entre 0,6 e 0,8 e a penalidade entre 0,1 e 0,2.

Para o gráfico 13, percebe-se que as séries autocorrelacionadas têm densidade mais forte em torno de 0,8 do parâmetro mixture e uma penalidade mais alta — a necessidade de uma penalização mais alta para modelar a autocorrelação entre os dados. Contudo, parece ter uma maior variabilidade da combinação dos parâmetros quando comparada à não autocorrelacionada, suas áreas de densidade são mais fortes em regiões distantes da área de maior densidade. Já a não autocorrelacionada também tem sua maior densidade em níveis de penalização acima de 0,3, entretanto, parece ser a única área de grande relevância da combinação do parâmetro, e tendo sua área de maior densidade visivelmente maior que a do caso antagônico.

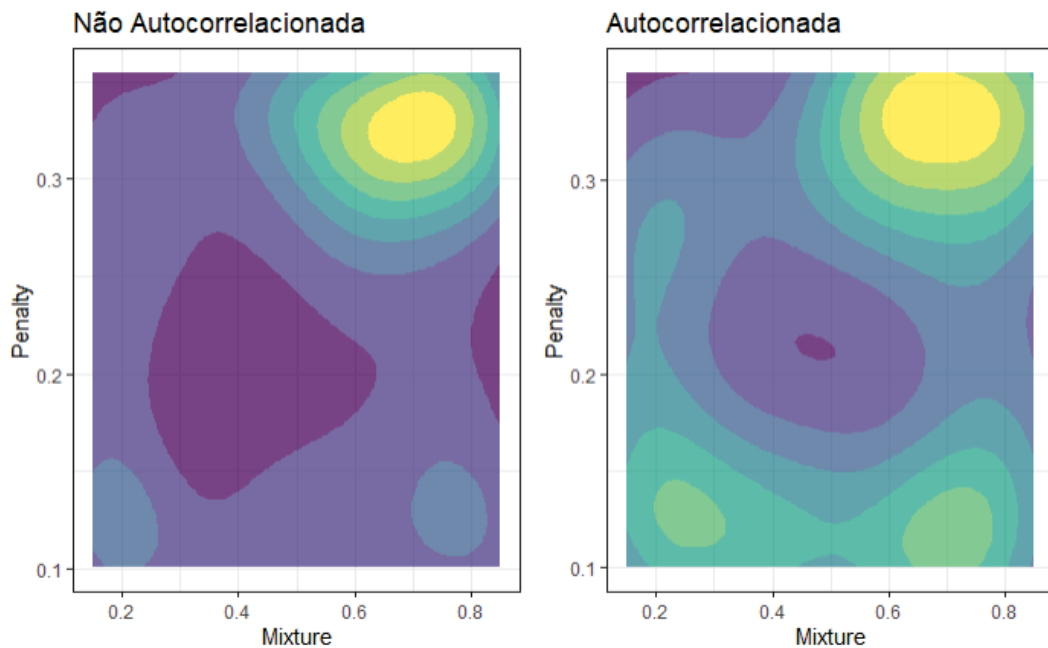


FIGURA 13 – Densidade Conjunta - Glmnet Poisson - Autocorrelação - Intermitente
Fonte: Elaboração própria

5.5.2.1 Parâmetros Glmnet - MIMO

Os parâmetros do modelo Glmnet com distribuição Poisson, conforme a tabela 12, a qual está marginalmente diferente das anteriores. Optou-se por este formato pois as diferenças entre os parâmetros Mixture e Penalty foram bastante marginais e homogêneas em termos de média e desvio padrão.

TABELA 12 – Resumo - Glmnet Poisson - Intermitente

Inflação	1	2	SD - Mix	Média - Mix	SD - Penalty	Média - Penalty	Qtd
0,02-0,4	não	sim	0,20	0,67	0,11	0,24	20
0,02-0,4	não	não	0,23	0,57	0,09	0,25	21
0,02-0,4	sim	sim	0,26	0,56	0,10	0,23	15
0,02-0,4	sim	não	0,20	0,62	0,08	0,28	68
0,41-0,5	não	sim	0,24	0,58	0,09	0,19	14
0,41-0,5	não	não	0,20	0,57	0,09	0,26	27
0,41-0,5	sim	sim	0,21	0,42	0,10	0,22	14
0,41-0,5	sim	não	0,22	0,56	0,09	0,25	68
0,51-0,60	não	sim	0,27	0,48	0,10	0,23	11
0,51-0,60	não	não	0,22	0,52	0,10	0,23	17
0,51-0,60	sim	sim	0,20	0,60	0,10	0,24	15
0,51-0,60	sim	não	0,21	0,53	0,09	0,24	71
0,61-0,88	não	sim	0,20	0,41	0,09	0,22	8
0,61-0,88	não	não	0,19	0,61	0,07	0,30	20
0,61-0,88	sim	sim	0,21	0,48	0,11	0,21	12
0,61-0,88	sim	não	0,24	0,53	0,09	0,25	76

1 = "Estacionária ?"; 2 = "Autocorrelação ?"

Aqui vemos que diferentes combinações de estacionariedade e autocorrelação levam a diferentes escolhas, em média, do parâmetro mixture e de penalização L1 e L2. Contudo, em muitos casos, essas diferenças são bastante marginais. Quando se trata da variabilidade da incidência de ambos parâmetros, quase não há diferença visível.

- Mixture: A média do parâmetro apresenta diferenças maiores quando da baixa inflação de zeros e ausência de estacionariedade e também alta inflação e ausência de estacionariedade que é o caso mais distinto com 0,41 e 0,61, indicando que uma alta inflação de zeros e ausência de estacionariedade com presença de autocorrelação a mistura tende mais para modelos esparsos e ausência de autocorrelação para modelos para modelos com coeficiente pequenos, porém diferentes de zero.
- Penalty: o parâmetro em questão, determinante da magnitude da penalização de regularização, apresenta, também, as maiores diferenças nos casos de inflações mais altas de zero e ausência de estacionariedade. Aqui também, a variabilidade da incidência do parâmetro não apresenta diferenças além das marginais.

5.5.2.2 Comentários Finais - GLMNET Poisson

Aqui vemos que, apesar do uso de diversas variáveis, que advém da técnica MIMO, a penalização nos parâmetros não chega a ser tão severa, e o balanceamento do modelo é parcimonioso, tendo um valor máximo de 0,67 e uma penalidade máxima

em séries de maior inflação em 0,3, ainda fazendo uso das variáveis mais próximas de um componente temporal, como as médias móveis de y e aumentadas de Fourier. Aqui enfatizo uma limitação: a quantidade de séries temporais por faixa. Talvez, com mais séries temporais, possamos observar diferenças mais interessantes entre os parâmetros.

5.5.3 Support Vector Regression

Aqui temos a segunda proposta MIMO, sendo o modelo que ficou em terceiro lugar pela métrica de erro escolhida. O modelo em questão segue com o kernel RbfSigma, igual ao primeiro modelo desta seção; contudo, aqui o SVR foi utilizado diretamente na série temporal e não nos resíduos de um modelo ARIMA de forma anterior.

Começando pela densidade dos hiperparâmetros Cost e RbfSigma. Ambos os eixos estão na escala log base 10.

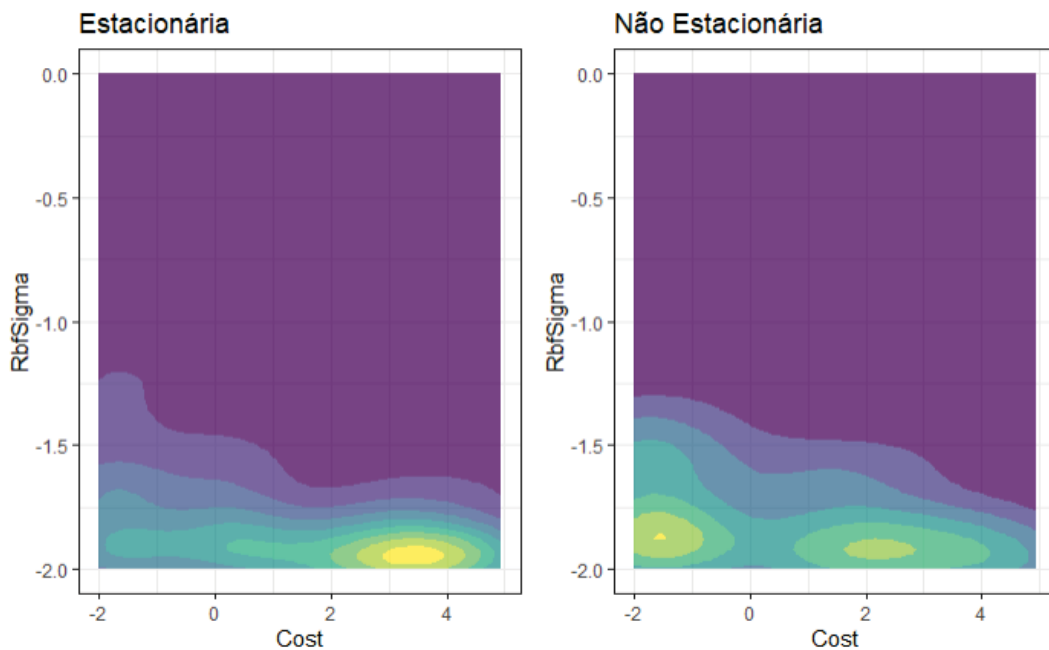


FIGURA 14 – Densidade conjunta - SVR - Estacionariedade - Intermitente

Se tratando do caso estacionário ou não, há uma concentração em valores pequenos do RbfSigma, uma diferença interessante nestes casos reside na bimodalidade aparente do caso não estacionário, em torno do custo -2 e +2, já no caso estacionário, demonstra uma densidade mais forte em torno de 3,5 – 4 no parâmetro custo. Em suma, o caso não estacionário demonstra uma variabilidade maior.

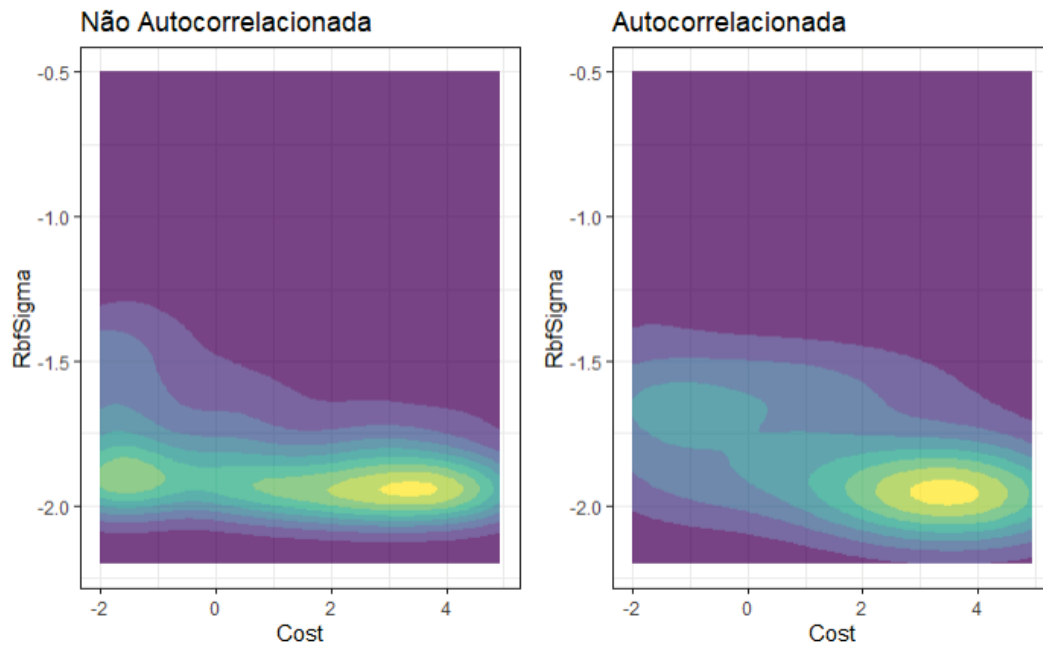


FIGURA 15 – Densidade conjunta - SVR - Autocorrelação - Intermitente

No caso de autocorrelação e não autocorrelação, começando pela não autocorrelação, apresenta duas áreas de maior densidade, representando uma bimodalidade, o que implica em dois conjuntos distintos onde os hiperparâmetros têm melhor eficácia. No caso autocorrelacionado, há uma única região grande amarela, sugerindo ser unimodal e mais larga, indicando uma ampla gama de hiperparâmetros possíveis com tendência a valores de custo mais altos. Em se tratando de variabilidade, o gráfico autocorrelacionado demonstra uma gama maior de valores a serem considerados, e o gráfico não autocorrelacionado, tendo dois picos bem definidos, demonstra que alguns valores são preferíveis a outros.

5.5.3.1 Parâmetros SVR - MIMO

Analisando tabela 13, é notado que algumas diferenças tanto em termos de média quanto da variabilidade pelo desvio padrão, em ambos os parâmetros.

TABELA 13 – Parâmetros SVR - MIMO - Autocorrelação

Inflação	Autocorrelação?	Sd - Rbf	Media - Rbf	Sd - Cost	Média - Cost	Qtd
0,24-0,5	sim	2,25	0,66	2,14	0,80	24
0,24-0,5	não	2,05	0,45	2,05	1,25	65
0,51-0,60	sim	0,91	-1,34	2,54	0,50	22
0,51-0,60	não	1,32	-1,35	2,08	0,71	69
0,61-0,69	sim	1,57	-1,38	2,65	1,28	13
0,61-0,69	não	1,41	-1,38	2,27	0,82	75
0,70-0,86	sim	2,12	-0,91	2,46	1,25	14
0,70-0,86	não	0,57	-1,71	2,61	0,63	77

- Rbf: A média deste parâmetro se confunde nos casos centrais de inflação de zeros, contudo seus desvios são diferentes; sendo os casos de autocorrelação presente com maiores desvios padrão. Já no caso sem autocorrelação, tendo menores variações, pode indicar séries mais estáveis sem necessidade de modelar a dependência entre os dados.
- Cost: Aqui os desvios padrão são muito próximos, contudo, as médias apresentam as diferenças mais pujantes. A média quando da presença de autocorrelação tende a ser menor em casos de mais baixa inflação de zeros, indicando uma menor penalidade por erro e no caso sem autocorrelação, valores maiores nestes casos, indicando maior penalidade em dados sem dependências tão claras.

Especificamente, a faixa 0,24 – 0,5 nos permite observar uma média maior para RBF e menor para o parâmetro Cost, e vice-versa, indicando que na presença de autocorrelação, existe uma incidência mais flexível em RBF e mais punitiva no parâmetro Cost.

A seguir, a mesma comparação, agora para ausência ou presença de estacionariedade.

TABELA 14 – Parâmetros SVR - MIMO - Estacionariedade

Inflação	Estacionária?	Sd - Rbf	Media - Rbf	Sd - Cost	Média - Cost	Qty
0,24-0,5	não	2,18	0,66	2,26	0,85	28
0,24-0,5	sim	2,06	0,44	1,98	1,25	61
0,51-0,60	não	0,73	-1,50	2,28	0,71	27
0,51-0,60	sim	1,38	-1,29	2,17	0,64	64
0,61-0,69	não	1,36	-1,53	2,32	1,05	22
0,61-0,69	sim	1,46	-1,33	2,34	0,83	66
0,70-0,86	não	1,01	-1,63	2,25	-0,26	30
0,70-0,86	sim	1,01	-1,56	2,62	1,22	61

- Rbf: aqui temos valores mais distintos na faixa de baixa inflação de zeros em termos de média, muito similar ao caso de autocorrelação. Os demais casos apresentam níveis similares nas médias segregadas.
- Cost: Aqui os desvios padrão são muito próximos, contudo, as médias apresentam as diferenças mais pujantes. A média quando da ausência de estacionariedade tende a ser menor em casos de mais baixa inflação de zeros e muito alta, indicando uma menor penalidade por erro e no caso não estacionário, valores maiores nestes casos, indicando maior penalidade em dados de média a invariância inconstantes.

Em séries com muitos zeros, as médias do parâmetro Sigma do kernel ficam muito próximas, em valores baixos, como $10^{-1,50} = 0,03$, indicando um decaimento

mais suave da função de kernel, sugerindo um modelo menos sensível às variações pontuais, uma característica das séries com inflação de zeros.

5.5.3.2 Comentários Finais - SVR

Nas séries temporais estacionárias ou não, é observada uma homogeneidade em termos de desvio padrão do parâmetro, em quase todas as faixas de inflação de zeros. As séries com presença de autocorrelação exigem uma abordagem mais dinâmica, indicando desvios diferentes no parâmetro de kernel, ajustando os parâmetros para acomodar a dependência temporal entre observações mais próximas; assim, o modelo tem a capacidade de prever com mais eficácia. A técnica MIMO tem a intenção de ajudar a melhorar a robustez do modelo ao lidar com a imprevisibilidade e a variação nestas séries temporais com inflação de zeros, e fornece um ambiente que captura as correlações cruzadas e as possíveis dependências mútuas.

5.6 SÉRIES TEMPORAIS - SUAVES

A análise das séries temporais suaves segue da mesma forma que na seção anterior, com o ranking médio das métricas de erro, seguidas da tabela com a contagem pelo rank do SMAPE, indicando os três modelos campeões.

TABELA 15 – Ranking dos Modelos

Técnica	Modelo	Rank_smape	Rank_mae	Rank_rmse	Rank_mase
MIMO	Glmnet	7,68	8,05	8,07	8,15
MIMO	Random Forest	8,29	8,63	8,53	8,73
MIMO	XGboost - Tweedie	8,40	8,47	8,54	8,57
MIMO	SVR	8,41	8,33	8,27	8,43
Iterativa	benchmark	8,59	8,52	8,42	8,61
MIMO	Cubist	8,76	9,15	9,50	9,22
MIMO	Glmnet - Poisson	8,88	10,58	11,00	10,57
MIMO	Mars	8,90	9,01	8,98	9,08
MIMO	Nnet	9,07	8,97	8,64	9,06
Híbrido	Exp - Random Forest	9,11	9,00	9,02	9,07
Híbrido	Arima - Glmnet	9,51	8,97	8,75	8,86
Híbrido	Exp - SVR	9,68	9,47	9,26	9,51
Híbrido	Arima - Random Forest	9,77	9,20	9,02	9,00
Híbrido	Arima Xgboost Res	9,93	10,46	10,71	10,45
Híbrido	Arima - SVR	10,04	9,22	8,86	9,03
Híbrido	Arima - MARS	10,14	9,63	9,43	9,45
MIMO	XGboost	12,13	11,82	12,05	11,77
Iterativo	NNAR	13,36	13,21	13,65	13,13

Aqui, temos como campeões três propostas híbridas do ARIMA, configurando mais uma vez o Arima SVR, e depois híbridos com modelos Random Forest e XGboost.

TABELA 16 – Ranking dos Modelos - Contagem

Modelo	Rank Contagem	Percentual
Arima - RF	187	0,097
Arima - Xgboost	176	0,091
Arima - SVR	147	0,076
Glmnet - Poisson	139	0,072
NNAR	124	0,064
Nnet	121	0,063
Cubist	120	0,062
Arima - Glmnet	115	0,06
Arima - MARS	111	0,058

Ademais, a seguir, as séries temporais separadas por estacionariedade e autocorrelação:

TABELA 17 – Estrutura Temporal - Suaves

Estacionária	Autocorrelacionada	Contagem
Não	Sim	535
Não	Não	387
Sim	Sim	433
Sim	Não	627

A seguir, temos a análise dos hiperparâmetros para estes casos. Para os casos intermitentes, utilizou-se a segregação por inflação de zeros; aqui, em séries suaves, de característica de ausência de zeros, optaremos pelos coeficientes de variação, no intuito de verificar as diferenças entre séries mais e menos variáveis e os efeitos na escolha dos parâmetros.

5.6.1 Arima - Random Forest

Sendo uma proposta híbrida, com o modelo Random Forest para regressão e técnica MIMO aplicadas aos resíduos do modelo ARIMA.

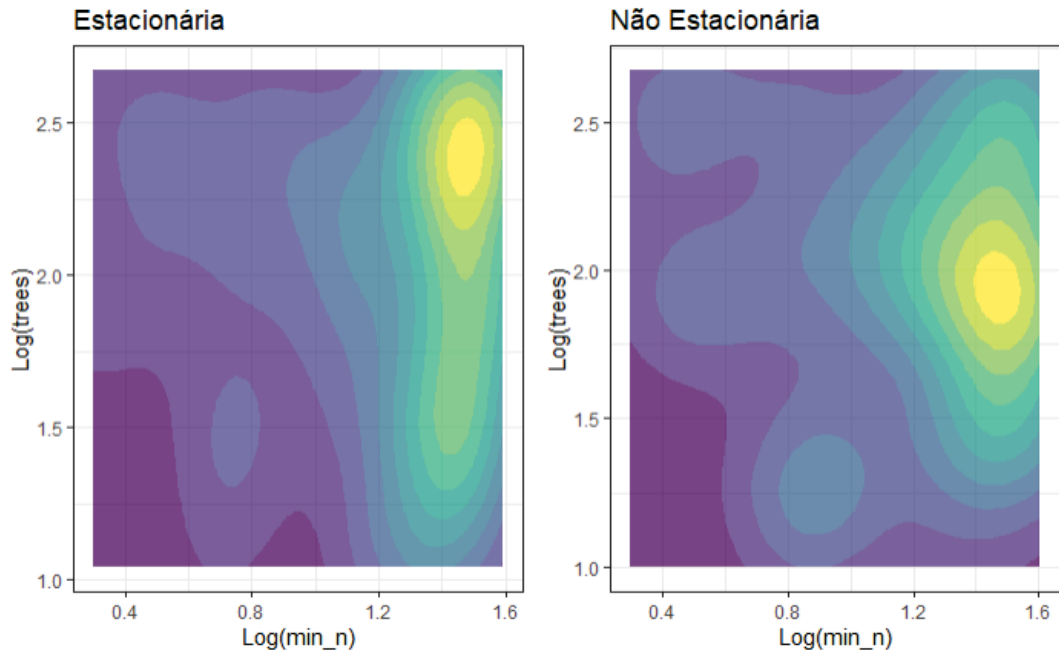


FIGURA 16 – Densidade conjunta - ArimaRF - Estacionariedade - Suaves

Para o caso estacionário e não estacionário percebe-se uma unimodalidade, sendo a diferença principal na quantidade média de árvores, onde o caso estacionário tende a 10^2 e o não estacionário em $10^{2,5}$. Ademais, as curvas de nível do caso estacionário apresentam uma certa força, tendo um efeito de esparramar a curva de nível em direção a $10^{1,5}$, ou seja, uma redução da quantidade de árvores.

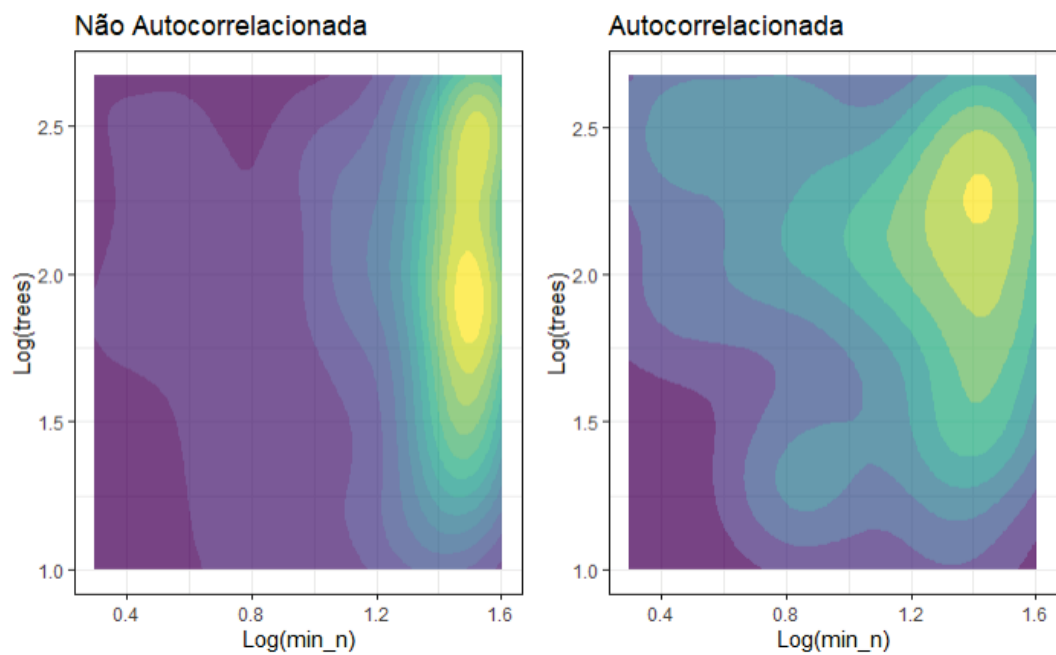


FIGURA 17 – Densidade conjunta - ArimaRF - Autocorrelação - Suaves

Para avaliação de Autocorrelação também é percebida uma unimodalidade, contudo aqui, vemos uma diferença mais acentuada, na qual a dispersão do parâmetro

min n tende a ser sempre próxima de $10^{1,6}$ em séries com ausência de autocorrelação, já no caso autocorrelacionado, onde temos um fator possivelmente inercial nas séries, percebe-se que em determinadas séries temporais houve escolhas próximas de $10^{0,4}$ e **trees** = $10^{2,5}$, indicando um menor uso de observações e também árvores mais complexas,

5.6.1.1 Parâmetros RF - Híbrido

Para efeitos de formatação, substituirei a palavra média por M e a palavra desvio padrão por SD, nas duas tabelas a seguir.

TABELA 18 – Parâmetros RF - Autocorrelação - Suaves

CV	Autocorrelação ?	SD - Árvores	M - Árvores	SD - min n	M - min n	Qtd
0,04-0,14	Sim	134	143	12	16	29
0,04-0,14	Não	115	104	10	26	16
0,15-0,28	Sim	117	139	11	16	24
0,15-0,28	Não	115	133	11	26	22
0,29-0,45	Sim	155	222	12	19	22
0,29-0,45	Não	140	167	14	20	23
0,46-0,67	Sim	134	157	7	26	11
0,46-0,67	Não	152	160	11	26	35

- Min n: A média chega a ter uma diferença de 10 observações entre casos de autocorrelação e ausência de, nos dois casos de coeficiente de variação baixos. Ainda, quão maior a variação a média deste parâmetro tende a se aproximar da séries não autocorrelacionada. O desvio padrão do indicador é misto entre classes.
- Árvores: Aqui, apresenta variabilidade maior, em presença de autocorrelação em comparação ao caso de ausência em quase todos os casos, e suas médias tendem a ser maiores no caso de autocorrelação, também com exceção dos casos mais variáveis, onde a diferença é marginal.

Aqui, é perceptível que com o aumento do coeficiente de variação, alguns casos apresentam aumento de média dos parâmetros, indicando que esta maior variabilidade pode requerer modelos com mais capacidade de generalização. Quando da segregação mais alta do coeficiente de variação, temos os valores mais altos de Min_n . Ademais, conforme os gráficos 40 e 42, indicando que a dispersão do parâmetro Min_n é maior para o caso não estacionário, para todas as faixas de segregação. Para o parâmetro de árvores, a dispersão também é maior em casos não autocorrelacionados e as medianas tendem a 10^2 .

TABELA 19 – Parâmetros RF - Estacionariedade - Suaves

CV	Estacionária ?	SD - Árvores	M - Árvores	SD - min n	M - min n	Qtd
0,04-0,14	Não	141	154	13	19	19
0,04-0,14	Sim	116	111	11	19	26
0,15-0,28	Não	121	137	13	19	28
0,15-0,28	Sim	109	135	11	23	18
0,29-0,45	Não	155	190	12	20	25
0,29-0,45	Sim	144	200	14	18	20
0,46-0,67	Não	103	100	9	28	18
0,46-0,67	Sim	159	198	10	24	28

- Min n: As médias são muito próximas entre séries estacionárias e não estacionárias, bem como a variabilidade do parâmetro. A principal diferença reside nas séries de maior CV e médio CV onde a diferença chega a ser de 4 observações.
- Árvores: Aqui, temos um comportamento com um limiar de mudança, as séries estacionárias tem menor média até a segunda segmentação do CV, e após isso as médias superam as séries não estacionárias, A variabilidade da quantidade de árvores nas séries não estacionárias é maior com exceção das séries temporais de maior variabilidade.

Estes resultados indicam uma abordagem diferenciada na formulação do modelo quando aplicado aos resíduos, sugerindo que séries estacionárias tendem a configurações mais robustas, utilizando mais árvores quanto mais variabilidade. Contrastando, as séries não estacionárias tendem a ter formulações mais flexíveis em séries com bastante variabilidade e mais dados, pois necessitam de respostas mais rápidas aos dados. Comparando-se os gráficos 41 e 43 que a dispersão do parâmetro Min_n é grande em ambas as condições de estacionariedade ou não, contudo, para a faixa de maior coeficiente de variação, temos notadamente uma concentração em $10^{1.5}$ e suas medianas também, indicando uma necessidade por mais dados por nó.

5.6.1.2 Comentários Finais - ArimaRF

Nos resultados analisados, tanto séries estacionárias ou não apresentam uma variação significativa nos parâmetros, e a maior variabilidade das séries demonstra um aumento na média do parâmetro $Trees$ e Min_n ; demonstrando a necessidade por modelos mais gerais em casos de grande variabilidade. Séries estacionárias e autocorrelacionadas tendem a apresentar uma maior estabilidade no modelo, com parâmetros que indicam uma maior robustez e capacidade de generalização. Séries não estacionárias e com menor autocorrelação exigem modelos mais flexíveis e rápidos na adaptação, refletindo a necessidade de ajustes frequentes devido à natureza imprevisível dos dados.

5.6.2 Arima - XGBoost

Sendo uma proposta híbrida, com o modelo Xgboost para regressão e técnica MIMO aplicada aos resíduos do modelo ARIMA.

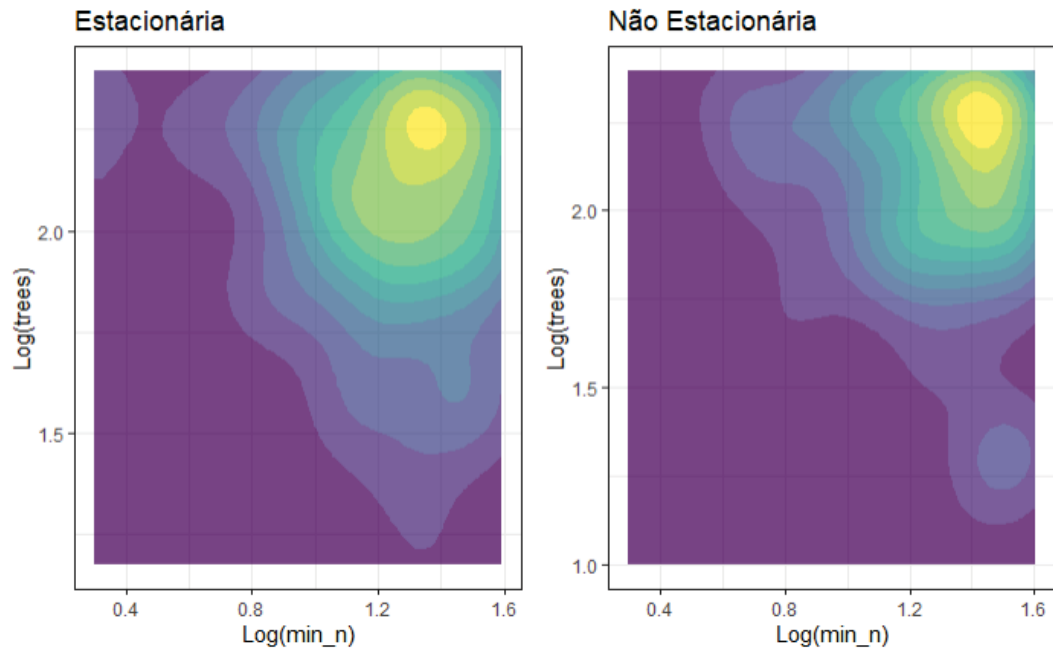


FIGURA 18 – Densidade conjunta - ArimaXGboost - Estacionariedade - Suaves

As séries não estacionárias tendem a fazer uso de um número ligeiramente maior de árvores quando comparado ao caso estacionário, e o parâmetro Min_n parece consistente entre ambas as características, estacionária ou não. As séries não estacionárias têm um esparramento menor da densidade para valores menores da quantidade de árvores, possivelmente pela sua característica de maior variabilidade de complexidade.

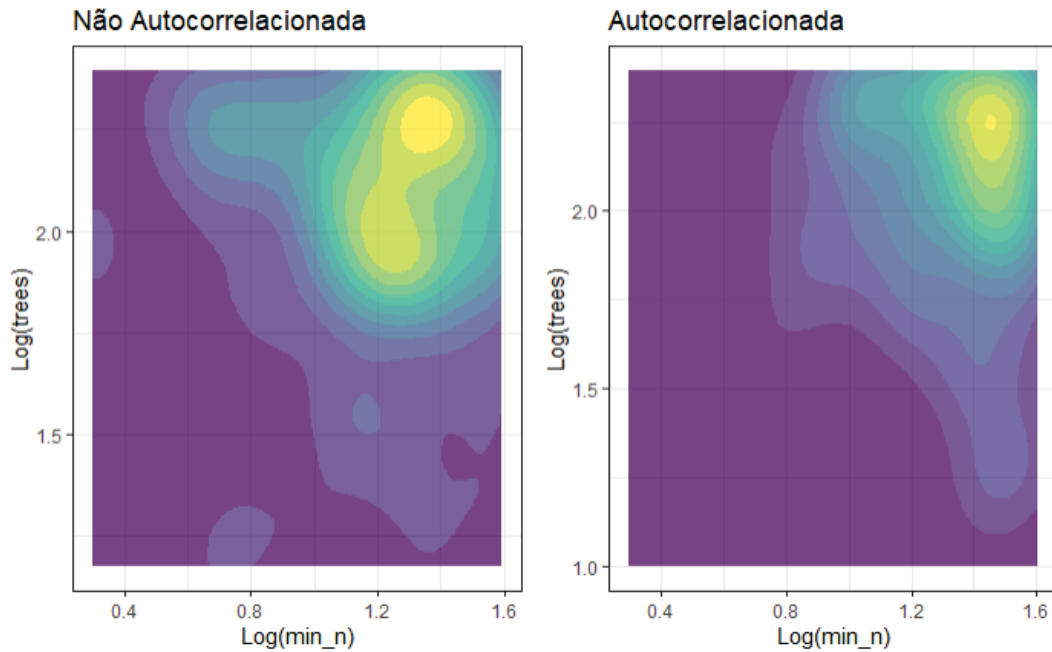


FIGURA 19 – Densidade conjunta - ArimaXGboost - Autocorrelação - Suaves

As séries não autocorrelacionadas parecem apresentar uma bimodalidade quanto à quantidade de árvores e em menor grau no parâmetro Min_n , ademais o esparramento para valores menores de Min_n parece maior que no caso de autocorrelação. O caso da autocorrelação parece necessitar consistentemente de mais árvores e valores maiores de Min_n . Isso indica que a presença de autocorrelação pode exigir ajustes no número de árvores para capturar a dependência temporal.

5.6.2.1 Parâmetros XBboost - Híbrido

TABELA 20 – Parâmetros XGboost - Autocorrelação - Suaves

CV	Autocorrelação?	SD - Árvores	M - Árvores	SD - min_n	M - min_n	Qtd
0,04-0,14	Sim	76	150	9	26	30
0,04-0,14	Não	56	159	8	20	13
0,14-0,27	Sim	74	144	11	23	25
0,14-0,27	Não	74	177	9	22	18
0,27-0,44	Sim	64	102	9	22	14
0,27-0,44	Não	67	110	11	13	26
0,44-0,7	Sim	71	123	13	25	14
0,44-0,7	Não	59	106	11	19	28

- Min n: As séries com autocorrelação tendem a ter uma média maior de Min_n , indicando que necessitam de mais estabilidade no modelo para capturar a dependência temporal, contudo a variabilidade entre as séries para o número mínimo de amostras por nó é similar.

- Árvores: Aqui, temos médias maiores para o caso de não autocorrelação com exceção da última faixa de coeficiente de variação, corroborando o gráfico de dispersão anterior. A variabilidade também tende a ser menor quanto ao caso autocorrelacionado, Isso indica que séries autocorrelacionadas podem exigir um número ligeiramente maior de árvores para capturar a dependência temporal.

A partir dos gráficos 44 e 46 é percebido que as distribuições de Min_n são mais estreitas e centradas, no caso da autocorrelação; já no caso da ausência de autocorrelação, são mais dispersas e com medianas mais próximas do terceiro quartil. Já para o parâmetro de árvores, a dispersão também é sobremaneira maior em casos não estacionários.

TABELA 21 – Parâmetros XGboost - Estacionaridade - Suaves

CV	Estacionária?	SD - Árvores	M - Árvores	SD - min_n	M - min_n	Qtd
0,04-0,14	Não	65	122	9	23	27
0,04-0,14	Sim	73	190	9	21	16
0,14-0,27	Não	74	152	10	20	28
0,14-0,27	Sim	63	114	10	24	15
0,27-0,44	Não	63	85	11	20	23
0,27-0,44	Sim	68	123	9	13	20
0,44-0,7	Não	71	152	13	24	16
0,44-0,7	Sim	57	106	11	18	26

- Min n: Este parâmetro demonstra uma homogeneidade entre as series de baixa variabilidade e uma diferença entre as duas últimas faixas, para uma média maior; contudo, com desvios padrão muito similares. Corroborando o gráfico 18, op qual demonstra uma dispersão muito similar entre os casos.
- Árvores: Aqui, temos uma média maior para o grau mais baixo e mais alto de CV, sendo que o grau mais baixo a diferença entre as médias é a maior dentre todas. Indicando que em séries mais controladas e estacionárias a modelagem dos resíduos tende a necessitar de mais árvores, bem como no caso menos controlado com o maior CV.

5.6.2.2 Comentários Finais - ArimaXGboost

O modelo XGboost aplicado aos resíduos claramente potencializa o modelo, dado os resultados indicados pela contagem e Smape. Ademais, é percebido de forma clara a diferença de variabilidade entre o Random Forest e o XGboost, no qual o Random Forest chega a ter desvios padrão duas vezes maiores que no caso do modelo XGboost, quando da escolha do parâmetro. Há um comportamento similar de médias de árvores e Min_n entre os dois modelos quando aplicados aos resíduos. A necessidade de mais árvores tanto no XGBoost como no modelo Random Forest pode ser atribuída

ao fato de que o ARIMA, ao não capturar bem todos os padrões em séries com alta variabilidade ou complexidade não linear, deixa resíduos com uma estrutura complexa. Por fim, temos uma inversão no comportamento comparando-se Min_n entre estes dois modelos e a característica de autocorrelação; existe uma inversão: o Random Forest tem médias menores na presença de autocorrelação. Já o modelo XGboost tem médias maiores para quantidade de dados em cada nó e menores em ausência de autocorrelação.

5.6.3 Arima - SVR

Sendo uma proposta híbrida, com o modelo SVR para regressão e técnica MIMO aplicadas aos resíduos do modelo ARIMA.

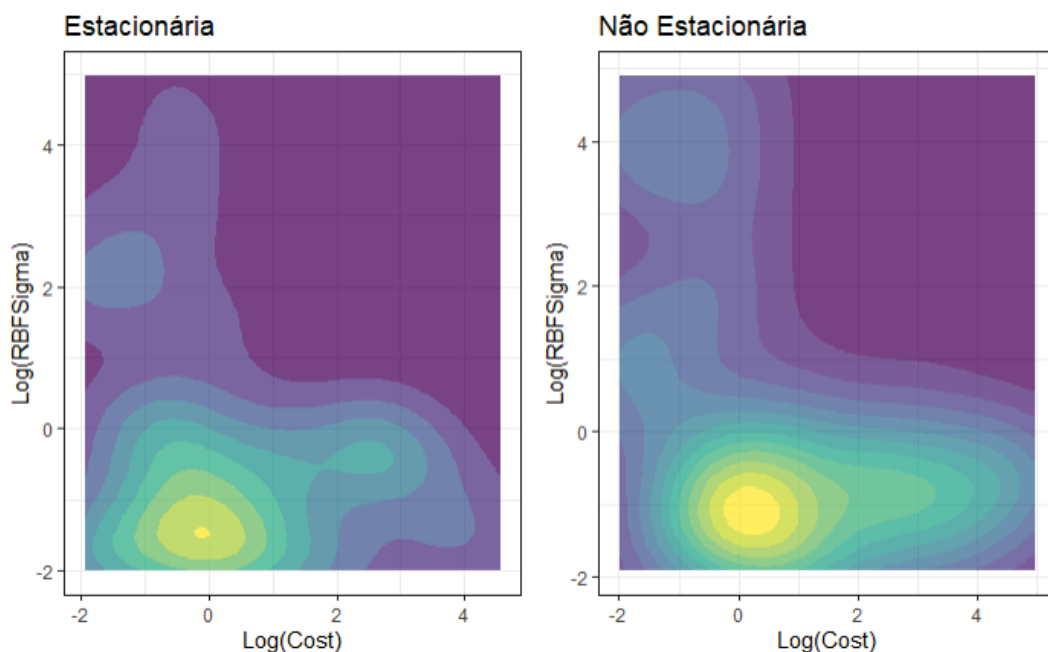


FIGURA 20 – Densidade conjunta - ArimaSVR - Estacionariedade - Suaves

Para o caso de estacionariedade temos uma densidade concentrada em torno de valores próximos a 10^0 na variável Cost, e para os valores de sigma do kernel em torno de $10^{-1,5}$, demonstrando uma preferência por valores pequenos em ambos os parâmetros. O caso de não estacionariedade tem sua similaridade quando da escolha dos parâmetros, contudo, com uma maior dispersão para o parâmetro de Cost, sugerindo valores mais altos de Cost, e também uma maior variabilidade para o sigma.

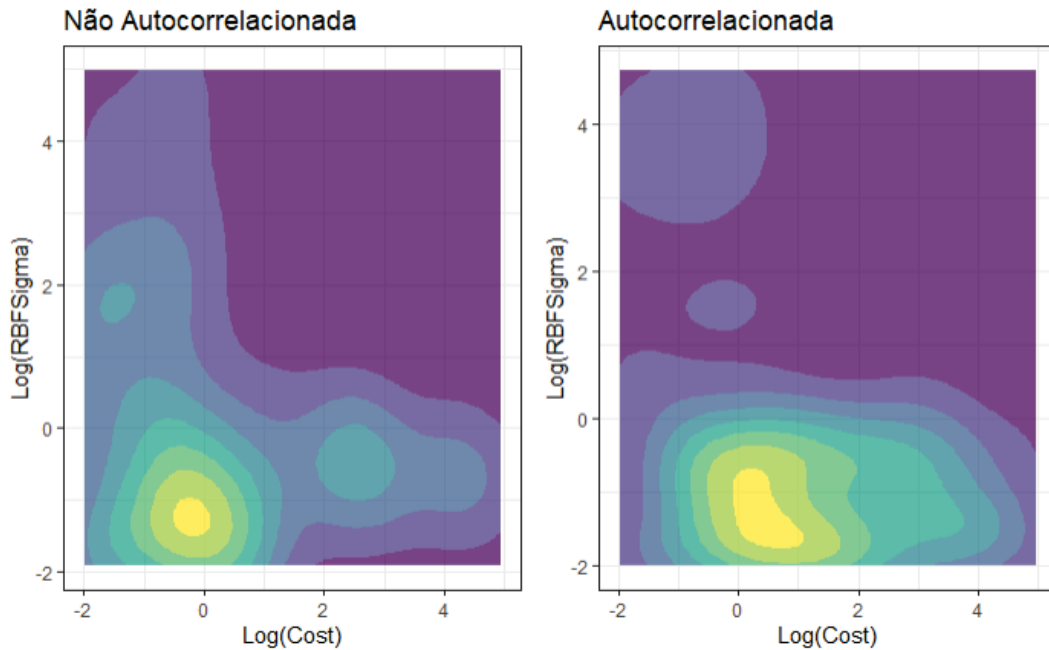


FIGURA 21 – Densidade conjunta - ArimaSVR - Autocorrelação - Suaves

Para o caso de autocorrelação ou ausência de autocorrelação. Começando pela ausência de autocorrelação, temos uma densidade concentrada em valores de Cost próximos a 0, similar ao caso das séries estacionárias, indicando que valores de custos em torno de 10^0 são preferíveis e demonstrando menor variabilidade do parâmetro que as séries autocorrelacionadas, a qual tem um comportamento similar de dispersão quando comparado ao caso não estacionário.

5.6.3.1 Parâmetros SVR - Híbrido

TABELA 22 – Parâmetros ArimaSVR - Autocorrelação - Suave

CV	Autocorrelação?	Sd - Rbf	Média - Rbf	Sd - Cost	Média - Cost	Qtd
0,05-0,16	Sim	2,05	-1,30	1,60	0,62	25
0,05-0,16	Não	1,13	-0,78	2,05	-0,05	12
0,17-0,25	Sim	1,56	-0,69	1,45	0,48	19
0,17-0,25	Não	1,89	-0,62	1,58	-0,43	18
0,26-0,40	Sim	1,90	-0,79	1,98	2,03	19
0,26-0,40	Não	1,73	-0,84	1,55	-0,62	17
0,41-0,70	Sim	1,94	-0,80	1,99	0,06	12
0,41-0,70	Não	1,96	-0,05	2,12	-0,34	24

- Rbf: A média do desvio padrão deste parâmetro é sobremaneira maior nas séries com autocorrelação, indicando que a presença de autocorrelação permite uma escolha, talvez, menos estável e mais dispersa — . As médias do parâmetro também são menores, principalmente nas séries com menor e maior variabilidade.

- Cost: Neste parâmetro é demonstrado uma diferença entre as médias do parâmetro, sendo que as séries autocorrelacionadas tem médias maiores, indicando uma maior penalização pelo erro e para séries sem autocorrelação um maior afrouxamento para ocorrência de erros.

A seguir, a análise no que concerne às séries estacionárias e não estacionárias.

TABELA 23 – Parâmetros ArimaSVR - Estacionariedade - Suave

CV	Estacionária?	Sd - Rbf	Média - Rbf	Sd - Cost	Média - Cost	Qtd
0,05-0,16	Não	2,18	-0,63	1,70	0,15	20
0,05-0,16	Sim	0,89	-1,29	1,77	1,01	17
0,17-0,25	Não	1,83	-0,29	1,25	0,09	11
0,17-0,25	Sim	1,65	-0,70	1,69	-0,18	26
0,26-0,40	Não	1,65	-0,74	2,11	2,26	20
0,26-0,40	Sim	2,03	-1,40	1,50	-0,13	16
0,41-0,70	Não	2,06	-0,36	2,28	-0,19	23
0,41-0,70	Sim	1,89	-0,02	1,57	-0,48	13

- Rbf: A média do parâmetro tende a ser mais negativa — indicando valores menores — para os casos de estacionariedade em quase todos os casos menor no de maior variabilidade da série temporal e a variabilidade do parâmetro tende a ser menor neste tipo de série temporal.
- Cost: Neste parâmetro temos uma diferença grande na escolha quando da presença ou ausência de estacionariedade. Para séries estacionárias a média tende a ser menor, com exceção das séries menos voláteis e a maior diferença reside na última segregação de volatilidade, na qual a séries não estacionárias tem uma média de $10^{2,26}$. Também, a variabilidade do parâmetro tem diferenças, corroborando o primeiro gráfico desta seção.

5.6.3.2 Comentários Finais - ArimaSVR

Séries estacionárias e autocorrelacionadas tendem a ter valores menores de Sigma, sendo uma suavização mais forte, em séries que têm componentes mais bem definidos, ajudando a capturar esse padrão com uma maior sensibilidade. Nos casos não estacionários e não autocorrelacionados, temos valores maiores para o parâmetro, trabalhando com séries temporais que podem ter mais variabilidade de padrões não definidos, de forma a capturar padrões gerais. Para séries não autocorrelacionadas e estacionárias, os valores de Cost são geralmente menores; indicando que o modelo penaliza menos os erros, pois a ausência de dependência temporal torna os padrões mais fáceis de capturar.

5.7 SÉRIES TEMPORAIS - ERRÁTICAS

A análise das séries temporais erráticas segue da mesma forma que na seção anterior, com o ranking médio das métricas de erro, seguidas da tabela com a contagem pelo rank do SMAPE, indicando os três modelos campeões.

TABELA 24 – Ranking dos Modelos

Técnica	Modelo	Rank_smape	Rank_mae	Rank_rmse	Rank_mase
MIMO	Glmnet	7,68	8,05	8,07	8,15
MIMO	Random Forest	8,29	8,63	8,53	8,73
MIMO	XGboot - Tweedie	8,40	8,47	8,54	8,57
MIMO	SVR	8,41	8,33	8,27	8,43
Iterativa	Arima - Benchmark	8,59	8,52	8,42	8,61
MIMO	Cubist	8,76	9,15	9,50	9,22
MIMO	Glmnet - Poisson	8,88	10,58	11,00	10,57
MIMO	MARS	8,90	9,01	8,98	9,08
MIMO	NMLP	9,07	8,97	8,64	9,06
Híbrido	Exp - Random Forest	9,11	9,00	9,02	9,07
Híbrido	Arima - glmnet	9,51	8,97	8,75	8,86
Híbrido	Exp - Svm	9,68	9,47	9,26	9,51
Híbrido	Arima - Random Forest	9,77	9,20	9,02	9,00
Híbrido	Arima Xgboost Res	9,93	10,46	10,71	10,45
Híbrido	Arima - SVR	10,04	9,22	8,86	9,03
Híbrido	Arima - Mars	10,14	9,63	9,43	9,45
MIMO	XGboost	12,13	11,82	12,05	11,77
Iterativo	NNAR	13,36	13,21	13,65	13,13

Aqui, temos como campeões três propostas híbridas de Arima. Configurando mais uma vez o Glmnet Poisson, e depois híbridos com modelos XGboost e SVR.

TABELA 25 – Rank - Contagem

Modelo	Rank_smape	Percentual
GLMNET - Poisson	216	0,111
Arima Xgboost Res	185	0,095
Arima - Svr	171	0,088
Arima - RF	150	0,077
CUBIST	129	0,066
Arima Mars	110	0,056

A seguir a tabela com a estrutura temporal das séries erráticas :

TABELA 26 – Estrutura Temporal - Suaves

Estacionária	Autocorrelacionada	Contagem
Não	Sim	594
Não	Não	403
Sim	Sim	380
Sim	Não	611

A seguir, temos a análise dos hiperparâmetros para estes casos. Para os casos intermitentes, utilizou a segregação por inflação de zeros; aqui, em séries erráticas, de característica de ausências de zeros, optaremos pelos coeficientes de variação, no intuito de verificar as diferenças entre séries mais e menos variáveis e os efeitos na escolha dos parâmetros.

5.7.1 Glmnet - Poisson

Aqui temos a proposta MIMO, composta pela modelagem através do modelo Glmnet com distribuição de probabilidade Poisson, agora para séries temporais erráticas.

Começando pela densidade dos hiperparâmetros de mixture e penalty. Ambos os eixos estão na escala natural, sem alteração, uma vez que estes parâmetros variam entre 0 e 1.

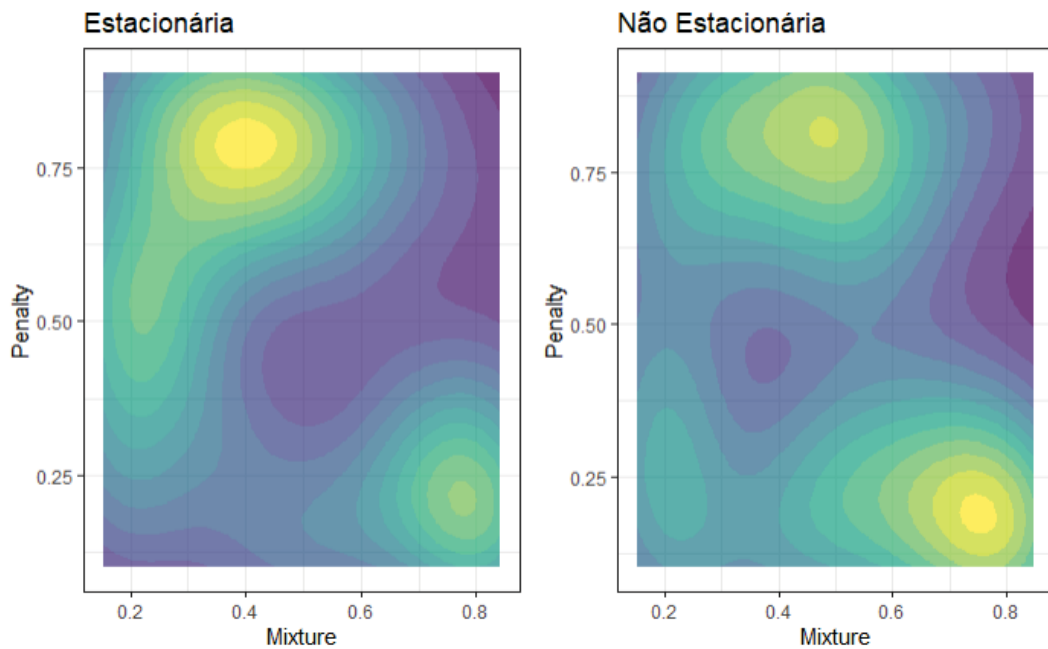


FIGURA 22 – Densidade conjunta - Glmnet Poisson - Estacionariedade - Erráticas

Para o caso sobre se as séries são estacionárias ou não, o parâmetro penalty parece estar concentrado em torno de 0,5 e 0,75, sugerindo no caso das séries

estacionárias, que o modelo tem alta penalização para o termo de regularização, enquanto no caso não estacionário existe uma concentração em 0,75, contudo também em 0,25 indicando uma quase bimodalidade, indicando uma maior variabilidade na incidência do parâmetro. O parâmetro de mixture, no caso estacionário, demonstra um pico em torno de 0,4, sugerindo uma mistura mais enfática para a penalização ridge e no caso não estacionário indicando uma maior variabilidade dando ênfase em ambos Lasso e Ridge.

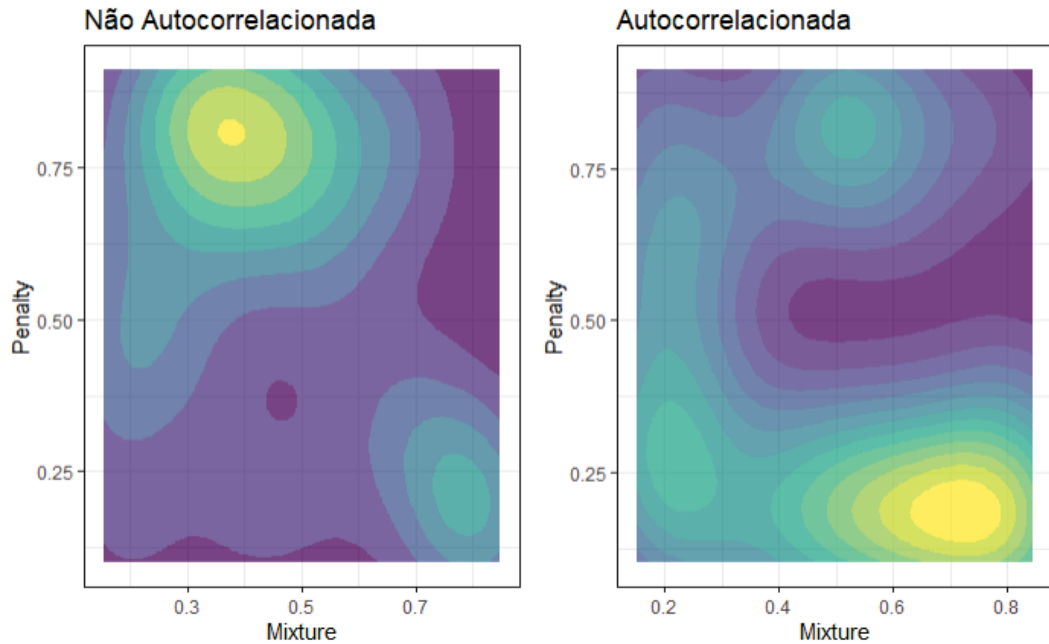


FIGURA 23 – Densidade conjunta - Glmnet Poisson - Autocorrelação - Erráticas

Para o caso sobre se as séries são estacionárias ou não, o parâmetro penalty parece estar concentrado em torno de 0,5 e 0,75, sugerindo no caso das séries estacionárias, que o modelo tem alta penalização para o termo de regularização, enquanto no caso não estacionário existe uma concentração em 0,75, contudo também em 0,25 indicando uma quase bimodalidade, indicando uma maior variabilidade na incidência do parâmetro. O parâmetro de mixture, no caso estacionário, demonstra um pico em torno de 0,4, sugerindo uma mistura mais enfática para a penalização ridge e no caso não estacionário indicando uma maior variabilidade dando ênfase em ambos Lasso e Ridge.

5.7.1.1 Parâmetros Glnet Poisson - MIMO

TABELA 27 – Parâmetros Glnet Poisson - Autocorrelação - Erráticas

CV	Autocorrelação?	SD - Mix	Média - Mix	SD - Penalty	Média - Penalty	Qtd
0,11-0,40	Sim	0,25	0,46	0,27	0,35	29
0,11-0,40	Não	0,22	0,42	0,27	0,55	25
0,41-0,53	Sim	0,21	0,53	0,32	0,48	21
0,41-0,53	Não	0,19	0,43	0,20	0,69	32
0,54-0,62	Sim	0,24	0,53	0,29	0,43	27
0,54-0,62	Não	0,19	0,51	0,29	0,55	26
0,63-0,96	Sim	0,23	0,47	0,23	0,36	22
0,63-0,96	Não	0,21	0,49	0,28	0,52	32

- Mixture: Há uma maior variabilidade na combinação de penalizações L1 e L2 no caso de autocorrelação, contudo, marginalmente maior. A média do parâmetro, é geralmente maior em séries com autocorrelação, com exceção da faixa de maior variabilidade; indicando uma ligeira ênfase no parâmetro L2 (ridge).
- Penalty: Aqui o desvio padrão se mostra maior em séries autocorrelacionadas, apenas na segunda faixa de variabilidade, na qual o desvio padrão alcança quase 53% da média. Quanto a média, é menor em séries autocorrelacionadas, indicando uma menor penalização para erros nestas séries, permitindo uma maior flexibilidade.

As medianas não estão expostas na tabela anterior, contudo, pelos gráficos 52 e 54, as séries autocorrelacionadas tendem a apresentar menor variabilidade no parâmetro de penalty e uma grande diferença de mediana; sugerindo medianas bem menores para o caso não autocorrelacionados, e as concentrações da incidência em valores mais próximos de 0,25. Já para o parâmetro mixture, em cada faixa do coeficiente de variação não parece existir uma grande diferença.

TABELA 28 – Parâmetros Glnet Poisson - Estacionariedade - Erráticas

CV	Estacionária?	SD - Mix	Média - Mix	SD - Penalty	Média - Penalty	Qtd
0,11-0,40	Não	0,20	0,43	0,29	0,45	23
0,11-0,40	Sim	0,25	0,45	0,28	0,44	31
0,41-0,53	Não	0,20	0,58	0,32	0,58	22
0,41-0,53	Sim	0,17	0,39	0,24	0,63	31
0,54-0,62	Não	0,21	0,51	0,30	0,46	31
0,54-0,62	Sim	0,21	0,54	0,28	0,54	22
0,63-0,96	Não	0,21	0,47	0,29	0,44	27
0,63-0,96	Sim	0,22	0,49	0,25	0,47	27

- Mixture: O parâmetro em se tratando de estacionariedade ou não, tem diferenças marginais, com diferença maior entre a faixa de $0,40 - 0,53$ de variabilidade, em relação a média do próprio parâmetro, sendo o caso não estacionário indicando uma maior ênfase para a penalização ridge. Ademais, há uma certa homogeneidade na dispersão do parâmetro; como esperado pelo gráfico de densidade.
- Penalty: As médias do parâmetro tem alguma diferença significativa nas camadas do meio, entre $0,4 - 0,62$, sendo que os casos de estacionariedade tendem a uma penalização maior nas séries estacionárias, e as não estacionárias com uma característica de maior flexibilidade ao ajuste. A variabilidade do parâmetro é quase homogênea, vide os gráficos de dispersão.

As medianas não estão expostas na tabela anterior, contudo, pelos gráficos 53 e 54, as séries estacionárias têm medianas sobremaneira menores que as séries não estacionárias em termos do parâmetro mixture e de maior concentração e menor variabilidade, bem como o parâmetro penalty. Contudo, as medianas do último parâmetro diferem em termos do coeficiente de variação sem direção única e aparente. Em suma, a estacionariedade sugere reduzir a variabilidade nestes dois parâmetros.

5.7.1.2 Comentários Finais - Glmnet Poisson

Então, é notada a diferença entre a escolha dos parâmetros em séries erráticas e intermitentes. As intermitentes demonstravam uma concentração mais homogênea dos parâmetros de penalização em cerca de $0,25$ para ambos os casos de estacionariedade ou não e autocorrelação ou não. Nas séries erráticas, percebe-se uma penalização maior, sugerindo que em séries erráticas o modelo tende a restringir de forma mais forte os coeficientes, evitando o sobreajuste. Já o valor do parâmetro mixture parece bem similar entre ambas as aplicações de séries erráticas ou intermitentes.

5.7.2 Arima - XGBoost

Sendo uma proposta híbrida, com o modelo Xgboost para regressão e técnica MIMO aplicada aos resíduos do modelo Arima, nas séries erráticas.

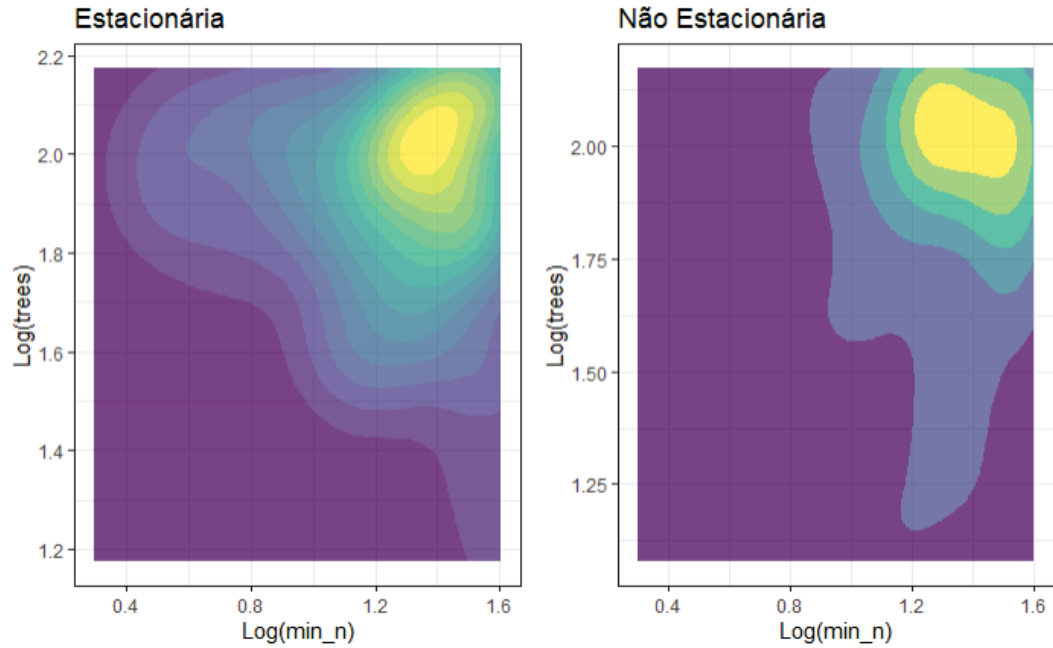


FIGURA 24 – Densidade conjunta - XGboost - Estacionariedade - Erráticas

O gráfico de densidade para os casos estacionários e não estacionários tem algumas diferenças interessantes. No caso estacionário, temos a existência de mais camadas e uma menor incidência de variabilidade dos parâmetros; contudo, de uma maior amplitude em Min_n quando comparado ao caso não estacionário. Seguindo para o caso de não estacionariedade, há uma concentração de área maior da combinação dos parâmetros, tendo uma variabilidade maior nos parâmetros de árvores, sugerindo uma maior variabilidade total pela área amarelada maior.

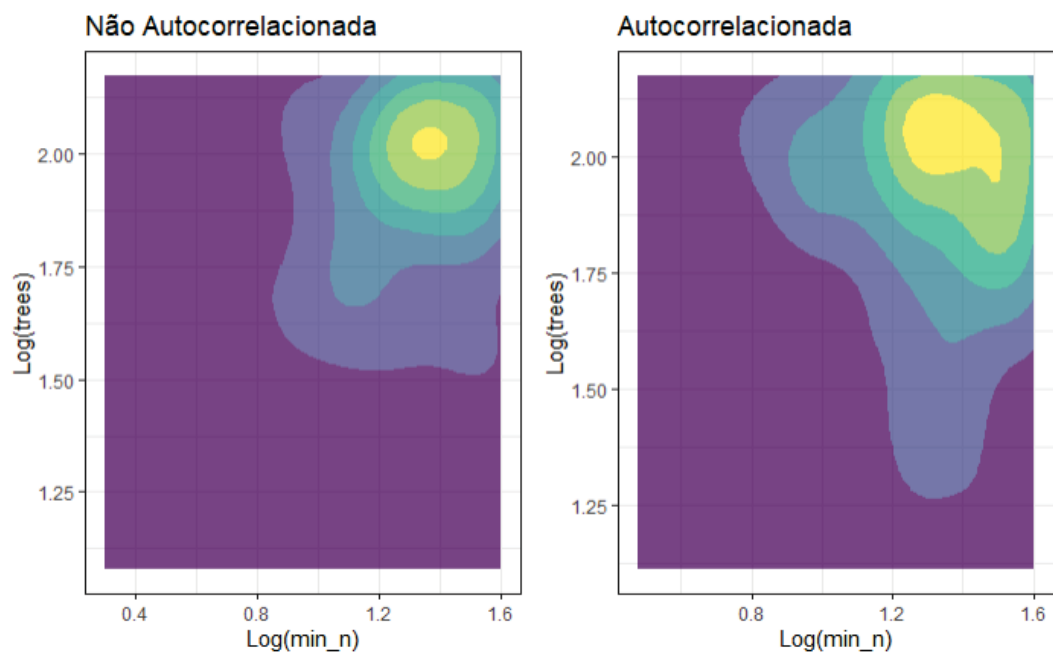


FIGURA 25 – Densidade conjunta - XGboost - Autocorrelação - Erráticas

Para a condição de estacionariedade ou não, é perceptível uma densidade melhor formada e delimitada dos casos não autocorrelacionados; já para os casos autocorrelacionados demonstra uma área maior de maior densidade, sugerindo uma maior variabilidade da combinação dos parâmetros. A densidade mais concentrada sugere que a presença de autocorrelação exige ajustes mais precisos nos parâmetros para capturar a dependência temporal.

5.7.2.1 Parâmetros XGboost - Híbrido

TABELA 29 – Parâmetros XGboost - Autocorrelação - Erráticas

CV	Autocorrelação?	SD - Árvores	M - Árvores	SD - min_n	M - min_n	Qtd
0,12-0,33	Sim	34	83	9	21	31
0,12-0,33	Não	33	81	11	27	15
0,34-0,45	Sim	41	104	10	23	24
0,34-0,45	Não	43	79	11	21	21
0,46-0,78	Sim	43	92	11	23	38
0,46-0,78	Não	38	100	11	19	53

- Min n: A variabilidade é marginalmente diferente na camada de menor coeficiente de variação, e também a média tende a ser sobremaneira maior nesta camada para as séries não autocorrelacionadas — com maior variabilidade. As demais camadas as diferenças de variabilidade são inexistentes e a média é levemente maior para os casos autocorrelacionados, indicando mais observações por nó nestes casos.
- Árvores: Aqui, as médias não tem diferença significativa na primeira faixa, já na segunda faixa de variação, o caso autocorrelacionado necessita sobremaneira mais árvores e quanto a terceira faixa as médias se aproximam, contudo, o caso de não autocorrelação requer mais árvores.

A presença ou não de autocorrelação quando da modelagem dos resíduos afeta a escolha dos parâmetros, principalmente em camadas de mais variabilidade das séries, onde o modelo predecessor ARIMA tem maior dificuldade de adaptação aos dados.

As medianas não estão expostas na tabela anterior, contudo, pelos gráficos 56 e 58, nas três faixas de variabilidade para as séries autocorrelacionadas sugerem um valor fixo em torno de $10^{1,2}$ e variabilidade menor em comparação à não autocorrelação, em se tratando o parâmetro Min_n . Para o caso das Árvores, a mediana e o comportamento da dispersão são bastante similares, seja em presença de autocorrelação ou não.

TABELA 30 – Parâmetros XGboost - Estacionariedade - Erráticas

flagcv	Estacionária?	SD - Árvores	M - Árvores	SD - min_n	M - min_n	Qtd
0,12-0,33	Não	34	92	8	20	26
0,12-0,33	Sim	32	78	12	26	20
0,34-0,45	Não	46	83	9	24	26
0,34-0,45	Sim	36	87	11	20	19
0,46-0,78	Não	42	96	11	20	56
0,46-0,78	Sim	37	102	11	20	35

- Min n: A média deste parâmetro tem uma distinção mais aparente nas duas primeiras faixas de variação, e comportamentos inversos, na primeira a média da observações por nó é maior no caso estacionário e na próxima faixa o inverso e também sua variabilidade em termos de desvio padrão. Quando a última faixa o parâmetro tem a mesma média e desvio padrão, casos mais difíceis de se prever de meu erraticidade.
- Árvores: Para séries estacionárias, a média do número de árvores varia de 78 a 102, enquanto para séries não estacionárias varia de 83 a 96. A variabilidade (SD) do número de árvores tende a ser maior para séries não estacionárias, indicando um campo maior de incidência pela dificuldade de uma série errática e não estacionária.

As medianas não estão expostas na tabela anterior, contudo, pelos gráficos 57 e 59, nas três faixas de variabilidade para as séries estacionárias e não estacionárias sugerem um valor próximo de $10^{1,2}$ e variabilidade similar, para o parâmetro Min_n , indicando uma quantidade de observações por folha ou nó, muito similar. Em relação ao parâmetro de Árvores, a mediana das séries estacionárias fica em torno de $10^{1,8}$, já as não estacionárias em torno de 10^2 .

5.7.2.2 Comentários Finais - ArimaXGboost

Séries com e sem autocorrelação não apresentam uma influência nos parâmetros, vide os gráficos de dispersão e as tabelas resumo. A presença de autocorrelação nas séries temporais não apresenta uma influência significativa nos parâmetros do XGBoost, como o número de árvores e o número mínimo de amostras por folha. Séries estacionárias tendem a requerer um número menor de árvores em comparação com séries não estacionárias. No entanto, a variabilidade dos parâmetros é semelhante entre os grupos.

5.7.3 Arima - SVR

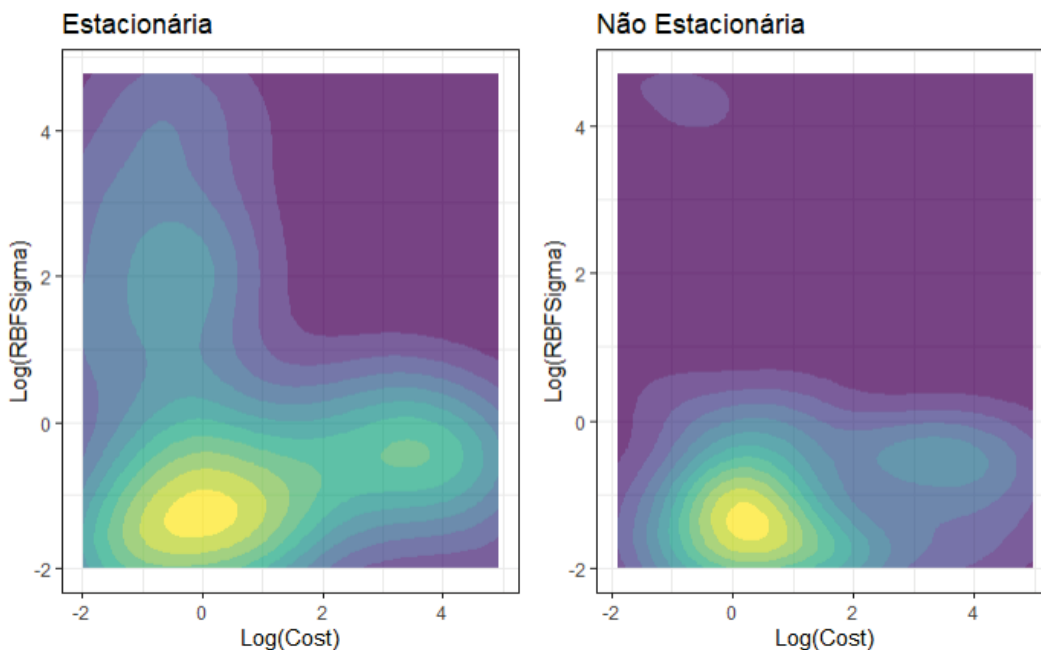


FIGURA 26 – Densidade conjunta - SVR - Estacionariedade - Erráticas

Há uma concentração de densidade no quadrante inferior direito, sugerindo que uma grande gama de combinações, contudo, em torno de zero (10^0) para o parâmetro cost e -1,5 para o parâmetro de kernel sigma, em se tratando das séries estacionárias. No caso não estacionário, a densidade está mais espalhada, mas também concentrada no quadrante inferior esquerdo. Ademais, indicam-se menores valores para sigma em casos não estacionários, sugerindo uma menor suavização da função de decisão, em séries de teor mais imprevisível.

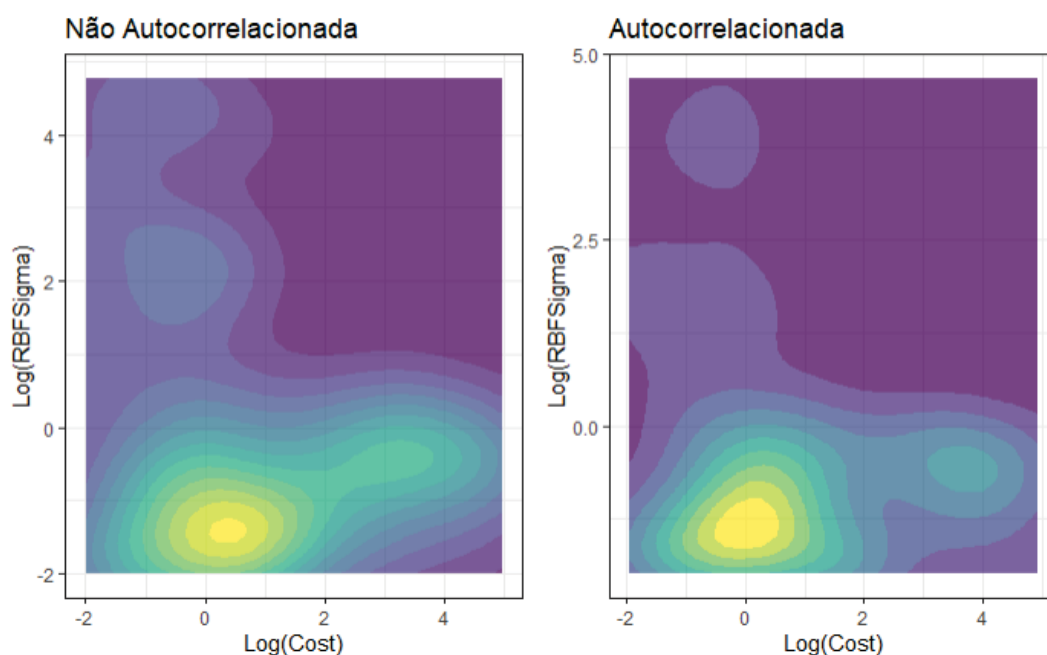


FIGURA 27 – Densidade conjunta - SVR - Autocorrelação - Erráticas

Há uma concentração de densidade no quadrante inferior direito, sugerindo que uma grande gama de combinações, contudo, em torno de zero (10^0) para o parâmetro cost e -1,5 para o parâmetro de kernel sigma, em se tratando das séries estacionárias. No caso não estacionário a densidade está mais espalhada, mas também concentrada no quadrante inferior esquerdo. Temos indicativos de menores valores para sigma em casos não estacionários, sugerindo uma menor suavização da função de decisão, em séries de teor mais imprevisível.

5.7.3.1 Parâmetros SVR - Híbrido

TABELA 31 – Parâmetros SVR - Autocorrelação - Errática

CV	Autocorrelação?	Desvio Cost	Média Cost	SD RBF	Média RBF	Qtd
0,07-0,38	Sim	1,47	0,04	1,77	-0,58	24
0,07-0,38	Não	1,94	-0,21	1,89	-0,57	19
0,39-0,49	Sim	2,27	-0,33	1,62	-0,69	21
0,39-0,49	Não	1,88	-0,15	0,83	-0,93	21
0,50-0,64	Sim	1,78	0,82	0,67	-1,38	18
0,50-0,64	Não	2,09	0,68	2,19	-0,28	24
0,65-0,83	Sim	1,70	0,52	1,33	-0,76	17
0,65-0,83	Não	1,92	0,37	1,99	-0,83	25

- Cost: Em séries com autocorrelação presente a média do parâmetro varia, contudo, tende a ser positiva ou ligeiramente negativa e maior que nos casos de ausência de autocorrelação. Em todas as faixas percebe-se uma diferença

entre médias, já os desvios variam entre as faixas de coeficiente de variação. Os valores sugerem que em séries com autocorrelação, permitindo menos erros de suavização da função de decisão.

- RBF: Este parâmetro é negativo em todos os casos, tendo maior variabilidade em séries sem autocorrelação, com exceção da segunda faixa de segregação. Aqui as médias se misturam, não parece existir um comportamento claro de segregação entre a série autocorrelacionada ou não, por faixa de coeficiente de variação.

Analisando os gráficos — encontrados no apêndice — 60 e 62, há indicativo de que casos autocorrelacionados têm distribuições mais estreitas e centradas que o caso de não autocorrelação quanto ao parâmetro RBF, principalmente nas camadas de mais variação, sendo que a terceira camada demonstra a maior diferença da mediana. Para o parâmetro Cost, para séries autocorrelacionadas, a penalização de erros é mais consistente e não necessita de grande variabilidade nos valores; já nos casos de ausência de autocorrelação, a distribuição é mais espaçada, contudo, com medianas próximas.

TABELA 32 – Parâmetros de Cost e RBF - Estacionariedade - Séries Temporais

CV	Estacionária?	Desvio Cost	Média Cost	Desvio RBF	Média RBF	Qtd
0,07-0,38	Não	1,58	0,29	1,52	-1,13	15
0,07-0,38	Sim	1,75	-0,26	1,87	-0,42	28
0,39-0,49	Não	1,83	-0,20	0,92	-1,05	18
0,39-0,49	Sim	2,26	-0,03	1,50	-0,54	24
0,50-0,64	Não	1,60	0,67	0,88	-1,19	22
0,50-0,64	Sim	2,28	0,78	2,35	-0,13	20
0,65-0,83	Não	1,92	1,49	2,03	-0,76	17
0,65-0,83	Sim	1,70	0,11	1,57	-0,83	25

- Cost: As médias demonstram boa separabilidade entre o caso estacionário e não estacionário, sendo médias mais positivas para o caso não estacionário, sugerindo uma busca mais ávida em minimizar os erros de previsão. Em suma, a variabilidade em casos de estacionariedade é maior, deste parâmetro.
- RBF: A média deste parâmetro é de forma consistente negativa, e os casos de ausência de estacionariedade apresentam menor variabilidade da escolha do parâmetro, com exceção da última faixa de segregação. De forma geral os dois tipos de séries temporais contam com o kernel em uma forma mais estreita.

Analisando os gráficos — encontrados no apêndice — 61 e 63, temos uma distribuição mais ampla para os casos de não estacionariedade, principalmente no caso

de maior coeficiente de variação; a mediana, neste caso, também difere sobremaneira. Percebe-se uma necessidade de maior flexibilidade na busca de parâmetros para a ausência de estacionariedade. Para o parâmetro RBF sigma, são notadas distribuições de menor dispersão em relação ao parâmetro cost, e o caso estacionário mais estreito que o caso não estacionário.

5.7.3.2 Comentários Finais - Arima SVR

Então é perceptível uma diferença na escolha dos parâmetros entre as séries temporais que são estacionárias ou não e autocorrelacionadas ou não, e também, principalmente quando da comparação com séries temporais intermitentes ou suaves. A incidência dos parâmetros e suas densidades mudam sobremaneira entre tipos de demanda. Nota-se que séries mais variáveis em termos de não apresentarem autocorrelação ou não serem estacionárias, tendem a ter valores mais dispersos para os parâmetros Cost e Sigma e séries com autocorrelação parecem ter valores mais centrados no intuito de capturar dependências temporais.

5.8 RESULTADOS CONSOLIDADOS - MODELOS

A seguir, uma sintetização em formato tabular dos modelos, uma vez que percebe-se uma recorrência dos campeões. Começando pelos modelos Support Vector Regression e o híbrido Arima-SVR.

TABELA 33 – Resumo dos parâmetros SVR para séries intermitentes, erráticas e suaves.

Característica	Intermitentes	Erráticas	Suaves
Parâmetro Cost	Médias muito maiores que os outros casos	Tendência a médias maiores que suaves, com maior dispersão	Cost mais baixo
Parâmetro Sigma (Kernel)	Sigma menor em média	Sigma mais estreito para capturar mudanças bruscas	Sigma em média maior que os demais, contudo ainda estreito.

TABELA 34 – Resumo dos efeitos da estacionariedade e autocorrelação nos parâmetros SVR.

Característica	Estacionário com Autocorrelação	Estacionário sem Autocorrelação	Não Estacionário com Autocorrelação	Não Estacionário sem Autocorrelação
Parâmetro Cost	Cost médio mais baixo, com menor dispersão	Cost médio baixo, com dispersão moderada	Cost médio um pouco maior, com dispersão maior	Cost médio mais alto, com maior dispersão
Parâmetro Sigma	Sigma médio maior, capturando variações suaves e tendências	Sigma médio moderado, balanceando variações globais e locais	Sigma médio menor, focando em padrões locais devido à autocorrelação	Sigma médio menor, capturando variações rápidas e imprevisíveis

A seguir, o resumo pelo ElasticNet, bem como seus parâmetros de mistura e penalização nos tipos de demanda intermitentes e erráticas, denotando o poder do modelo e a utilização da técnica de predição MIMO, em casos em que as séries temporais são sobremaneira mais voláteis que as suaves e inclusive com ocorrências de lacunas na demanda, como é o caso da intermitente.

TABELA 35 – Resumo dos efeitos em parâmetros do Glnet Poisson para séries intermitentes e erráticas.

Característica	Séries Intermitentes	Séries Erráticas
Mistura (α)	Valores médios mais altos (próximos de 0,6 a 0,7). Enfatiza a penalização Lasso (esparsidade).	Valores médios mais baixos (próximos de 0,4 a 0,5). Equilíbrio entre Lasso e Ridge para lidar com a variabilidade.
Penalização (λ)	Valores médios mais baixos (próximos de 0,2 a 0,3). Menor penalização para capturar eventos raros.	Valores médios mais altos (próximos de 0,4 a 0,7). Maior penalização para evitar overfitting em dados instáveis.
Estacionariedade	Pouca influência nos parâmetros. Parâmetros estáveis independentemente da estacionariedade.	Séries não estacionárias têm o α maior e λ menor. Ajustes nos parâmetros para lidar com variabilidade.
Autocorrelação	Pouca influência nos parâmetros. Leve redução em λ em séries autocorrelacionadas.	Séries autocorrelacionadas têm o α maior e λ menor. Modelo ajusta-se para capturar dependências temporais.

Por fim, os campeões que têm como base a formulação de árvore, os híbridos

Arima-Random Forest e Arima-Xgboost. Sendo as características do Arima RF nas séries erráticas, o Arima XGB nas séries suaves e erráticas, respectivamente.

TABELA 36 – Resumo dos efeitos da autocorrelação e estacionariedade nos modelos baseados em árvores.

Modelo e Tipo de Série	Parâmetro	Efeito da Autocorrelação	Efeito da Estacionariedade	Observações
ARIMA RF	Média do Número de Árvores	Maior em séries autocorrelacionadas	Menor em séries estacionárias	Autocorrelação aumenta a complexidade; estacionariedade reduz a necessidade de árvores
	Média do <i>min_n</i>	Menor em séries autocorrelacionadas	Sem diferença significativa	Folhas menores capturam padrões locais em séries autocorrelacionadas
ARIMA XGB	Média do Número de Árvores	Sem padrão claro	Maior em séries estacionárias nas faixas de CV baixas	Estacionariedade pode exigir mais árvores para capturar variações sutis
	Média do <i>min_n</i>	Ligeiramente maior em séries autocorrelacionadas	Menor em séries estacionárias	Folhas menores em séries estacionárias permitem ajuste mais fino
ARIMA XGB	Média do Número de Árvores	Maior em séries autocorrelacionadas	Menor em séries estacionárias	Autocorrelação aumenta a complexidade; estacionariedade reduz a necessidade de árvores
	Média do <i>min_n</i>	Menor em séries autocorrelacionadas	Maior em séries estacionárias	Folhas menores capturam padrões locais em séries autocorrelacionadas; folhas maiores generalizam em séries estacionárias

De maneira geral, modelos não sequenciais como ElasticNet com distribuição Poisson, com a técnica MIMO, funcionam bem para séries intermitentes e erráticas, bem como o modelo SVR. Percebe-se uma potencialização em termos de redução do SMAPE quando utilizada a técnica híbrida de ARIMA com modelos baseados em árvores e o SVR. O modelo híbrido ARIMA-SVR foi consistentemente o mais eficaz, quando se tratando de demanda de baixa variabilidade, em termos de SMAPE. A hibridização com Random Forest também demonstrou desempenho competitivo, sendo uma boa alternativa quando há leve não-linearidade no padrão da demanda. Quanto às demandas intermitentes, caracterizadas por inflação de zeros, os principais achados foram: a técnica MIMO com GLMNET-Poisson, que teve destaque como modelos mais robustos, seguido pelo híbrido XGboost, com desempenho sólido, utilizando-se de técnicas de regularização para tratar a inflação de zeros; já o SVR apresentou resultados mistos, sendo eficaz apenas em séries com menor proporção de zeros. No caso das demandas erráticas, os híbridos de modelos baseados em árvore foram os melhores desempenhos, trazendo a capacidade de capturar não-linearidades; modelos paramétricos, como ARIMA puro, tiveram desempenho inferior devido à incapacidade de lidar com a alta variabilidade.

6 CONCLUSÃO

Nesta dissertação uma gama de técnicas de previsão foi explorada, sendo estatísticas e de aprendizagem de máquina, técnicas híbridas de modelos paramétricos e os modelos de aprendizagem de máquina aplicados aos seus resíduos. Por meio de uma comparação abrangente focada principalmente pela métrica de erro Smape, foi feita a classificação dos modelos por contagem de quantas vezes cada modelo havia mostrado o melhor Smape; em séries temporais de demanda erráticas, intermitentes e suaves em dados de cadeia de suprimentos de máquinas de mineração e caminhões.

Foi reportado resultados indicando que os modelos híbridos têm performance superior em séries temporais suaves, erráticas e intermitentes, configurando entre os três melhores nos três tipos de demanda, e com mais expressividade em séries suaves. Outro ponto que salta aos olhos é a técnica MIMO aplicada em modelos não sequenciais como o SVR ou Glnet Poisson, reportando ótimos resultados e foi aplicada nos três contextos, mostrando potencial na redução da acumulação de erros em previsões de longo prazo. Esta abordagem foi especialmente benéfica em séries temporais complexas, como as erráticas e intermitentes. A utilização de variáveis exógenas e técnicas de *feature engineering* também contribuiu para a melhoria das previsões.

O modelo Glnet Poisson e SVR em séries intermitentes apresentou consistência em todas as métricas de erro, com um indicativo de robustez do uso do kernel RBF e das penalizações L1 e L2. Quanto aos parâmetros destes modelos e sua incidência, notou-se maior variabilidade em séries não estacionárias e autocorrelacionadas, de maneira geral. Especificamente aos parâmetros do modelo SVR, foi observado que séries não estacionárias exibem uma bimodalidade aparente, sugerindo duas áreas de maior densidade para os parâmetros de custo e sigma. Quanto aos parâmetros SVR aplicados aos resíduos, temos uma utilização de valores mais extremos em ambos os parâmetros, quando em comparação ao SVR na técnica MIMO, e a bimodalidade se mantém. Por fim, a técnica MIMO vem com indícios de ser benéfica na modelagem de demanda inflacionada de zero e intermitentes.

Para as séries temporais suaves, a escolha dos parâmetros dos modelos também é influenciada pela presença de características como estacionariedade e autocorrelação. Nestes tipos de séries temporais, os modelos que melhor apresentaram resultados são os híbridos Arima, contudo, Glnet Poisson configurou em quarto lugar. O SVR configura mais uma vez entre os melhores, sendo na técnica híbrida Arima SVR e os modelos baseados em árvore, Random Forest e XGboost. É notada uma diferença mais acentuada entre os parâmetros de árvore do Random Forest e do

XGboost, uma vez que o último utilizou mais árvores em comparação com o modelo Arima - Random Forest, tanto em séries estacionárias quanto não estacionárias, e em séries autocorrelacionadas e não autocorrelacionadas, porém não mais observações por nó. Quanto ao Arima SVR, há uma maior variabilidade em séries não estacionárias, sugerindo que o parâmetro de custo precisa ser mais penalizador quanto maior a variabilidade; já o parâmetro sigma do kernel RBF apresentou valores menores em séries estacionárias e autocorrelacionadas.

Quanto às séries temporais erráticas, que são caracterizadas por alta variabilidade, a escolha dos parâmetros nos modelos Arima-SVR, Arima-XGBoost e Glnet Poisson — os campeões desta seara —. Aqui, corrobora-se a ideia de avaliar os parâmetros dos modelos, uma vez que as diferenças entre a distribuição dos parâmetros dos três modelos são comparáveis com suas incidências em outros tipos de demanda, sejam intermitentes ou suaves. O modelo Glnet Poisson, por exemplo, configurando entre os primeiros em intermitentes e erráticos, demonstrou que a penalização é mais alta em séries erráticas do que em séries intermitentes e mais homogêneas; já o parâmetro de mistura tem uma combinação mais homogênea também em séries intermitentes do que em séries erráticas, as quais demonstraram valores mais extremos em seus gráficos de densidade. O modelo Arima SVR também demonstra diferenças fortes, tendo configurado nas três categorias de demanda.

O estudo tem suas limitações, sendo computacionais, de quantidade de dados, da utilização dos modelos, de correlação cruzada, entre outros. Neste estudo, o foco para otimização dos parâmetros ficou com a técnica Grid Latin Hypercube; contudo, outros métodos podem ser testados e documentados seus resultados. Ademais, outras técnicas de seleção dos dados para modelos não sequenciais, e não apenas o k-fold. Em se tratando do método de predição, o método direto não foi empregado, e outras formas do método MIMO podem ser empregadas e seus resultados reportados. As mesmas técnicas podem ser aplicadas em outros contextos, como dados financeiros de alta frequência, tráfego, varejo e não apenas a cadeia de suprimento aqui elencada. Modelos de deep learning não foram avaliados neste estudo, como RNN ou LSTM, e futuras pesquisas podem abordar este tópico por esta ótica. Outro ponto de grande interesse para melhoria desta dissertação seriam técnicas de explicabilidade em aprendizagem de máquina, para entender de forma mais clara as decisões pelos parâmetros após a validação cruzada.

Esta dissertação visa contribuir com um estudo comparativo e detalhado sobre a distribuição dos hiperparâmetros dos três melhores modelos, categorizados por tipo de série temporal, oferecendo uma análise empírica para cada modelo. Além disso, tem o intuito de demonstrar a eficácia das técnicas híbridas que combinam os modelos ARIMA e Suavização Exponencial com técnicas de aprendizado de máquina aplicadas

aos resíduos dos modelos tradicionais. Adicionalmente, são apresentados os resultados da aplicação de técnicas não sequenciais, como os modelos SVR e GLMNet Poisson, na previsão de dados sequenciais, temporais, ressaltando seus bons resultados em séries erráticas ou intermitentes.

A APÊNDICE A - BOXPLOTS E DENSIDADES

A.1 INTERMITENTES

A.1.1 Arima SVR

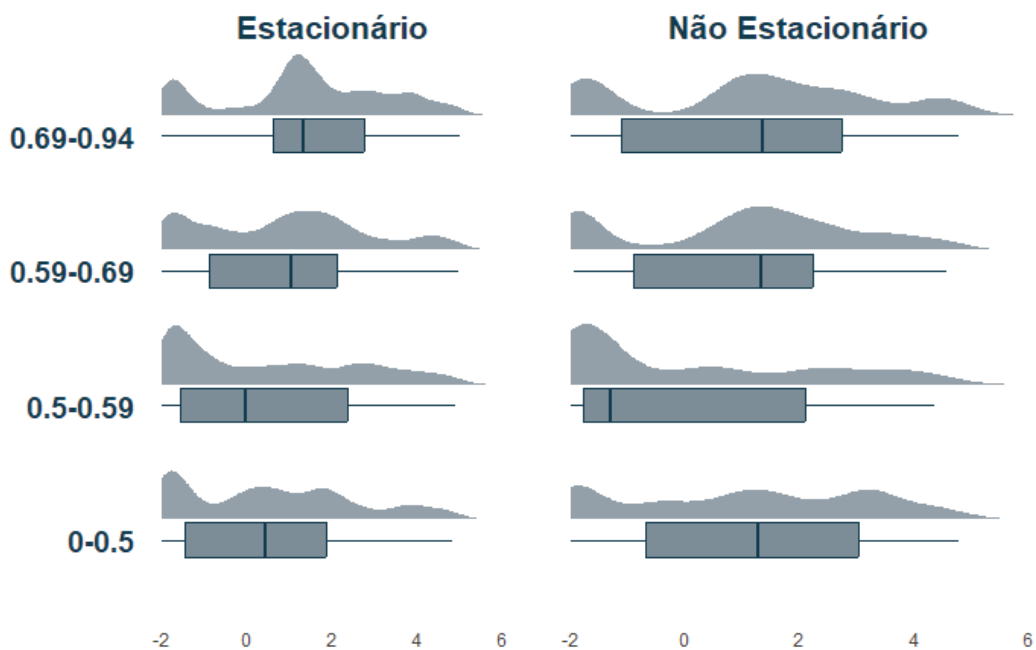


FIGURA 28 – Log(Cost) - Estacionariedade - Intermitente

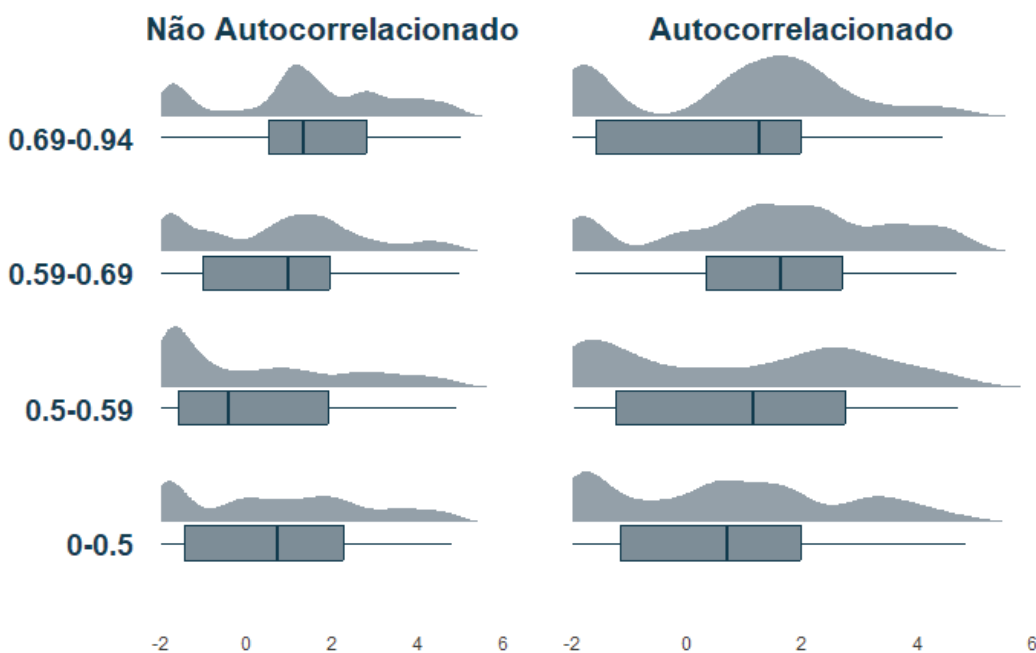


FIGURA 29 – Log(Cost) - Autocorrelação - Intermitente

Fonte: Elaboração própria

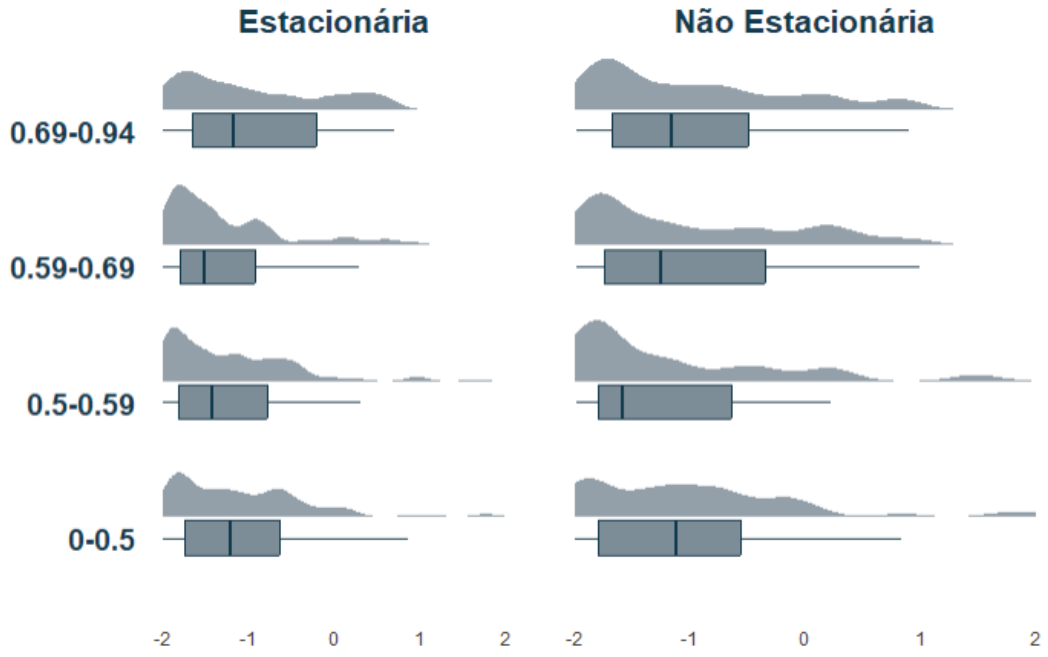


FIGURA 30 – Log(RbfSigma) - Estacionariedade - Intermitente

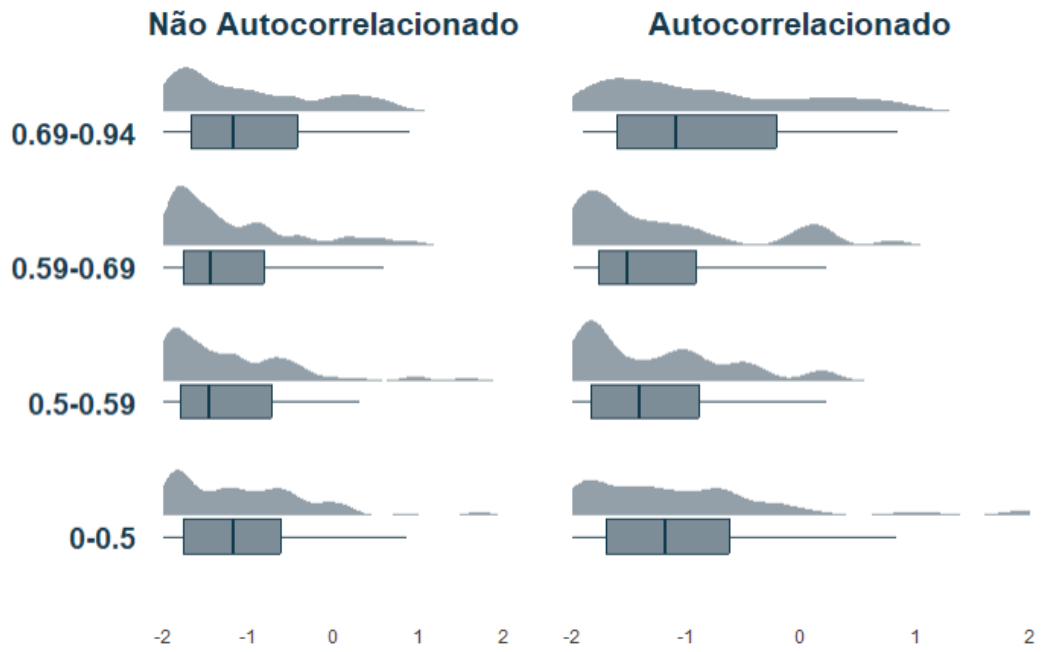


FIGURA 31 – Log(RbfSigma) - Autocorrelação - Intermitente

Fonte: Elaboração própria

A.1.2 Glnet Poisson

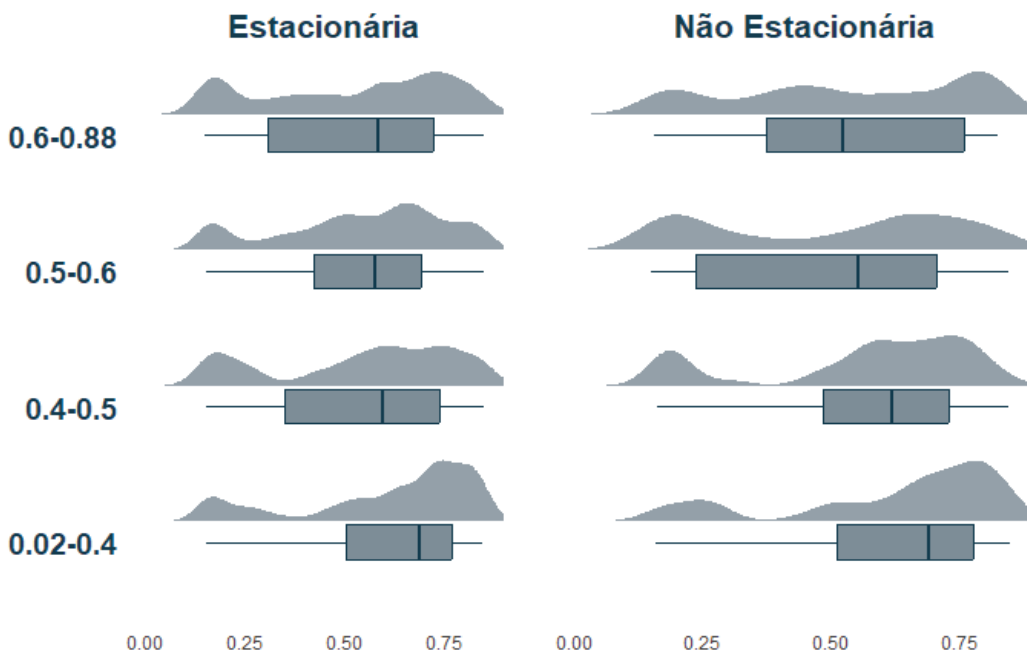


FIGURA 32 – Mixture - Glnet Poisson - Estacionária - Intermitente
 Fonte: Elaboração própria

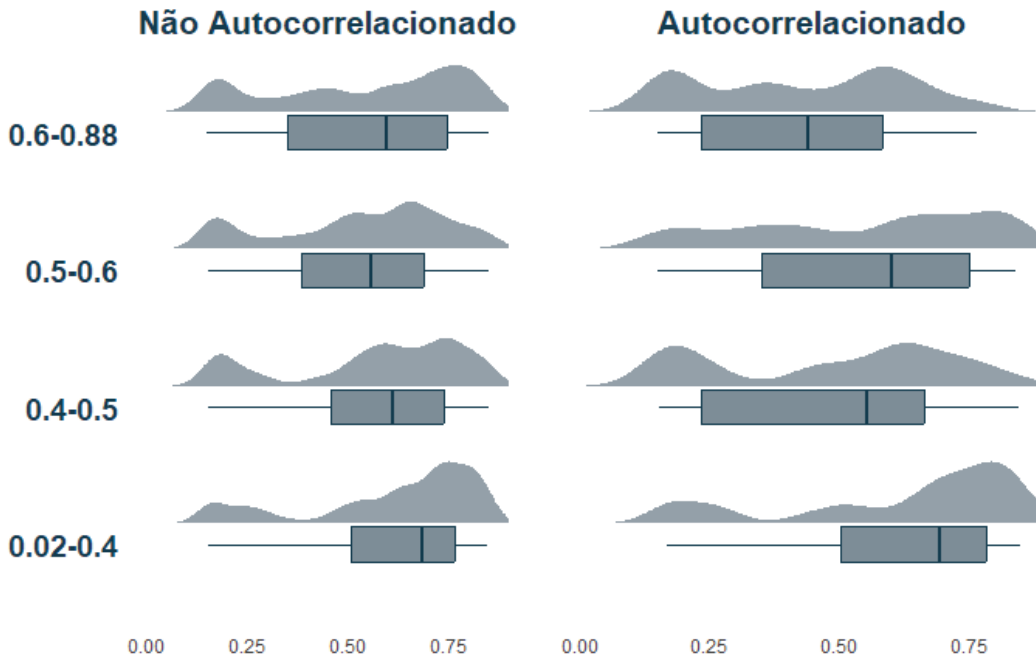


FIGURA 33 – Mixture - Glnet Poisson - Autocorrelação - Intermitente
 Fonte: Elaboração própria

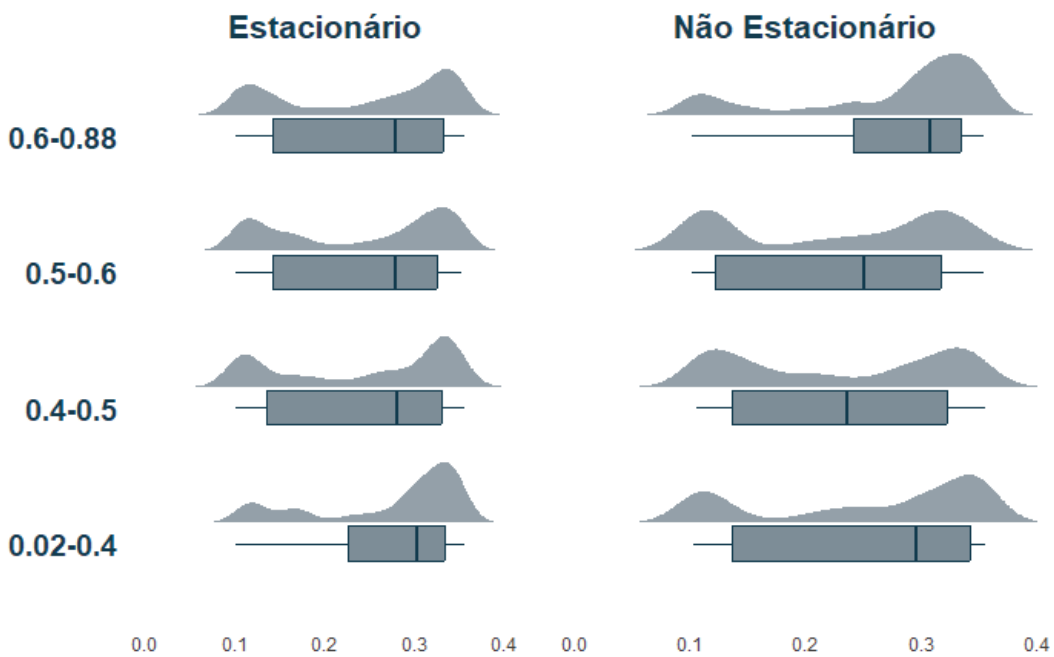


FIGURA 34 – Penalty - Glnet Poisson - Estacionária - Intermitente
 Fonte: Elaboração própria

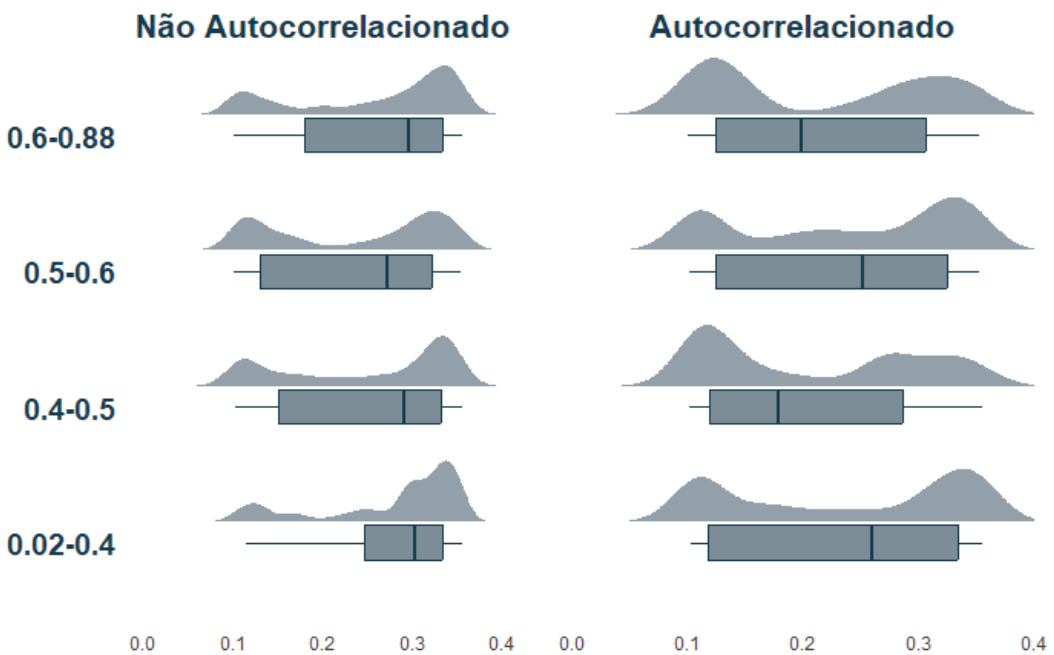


FIGURA 35 – Penalty - Glnet Poisson - Autocorrelação - Intermitente
 Fonte: Elaboração própria

A.1.3 SVR

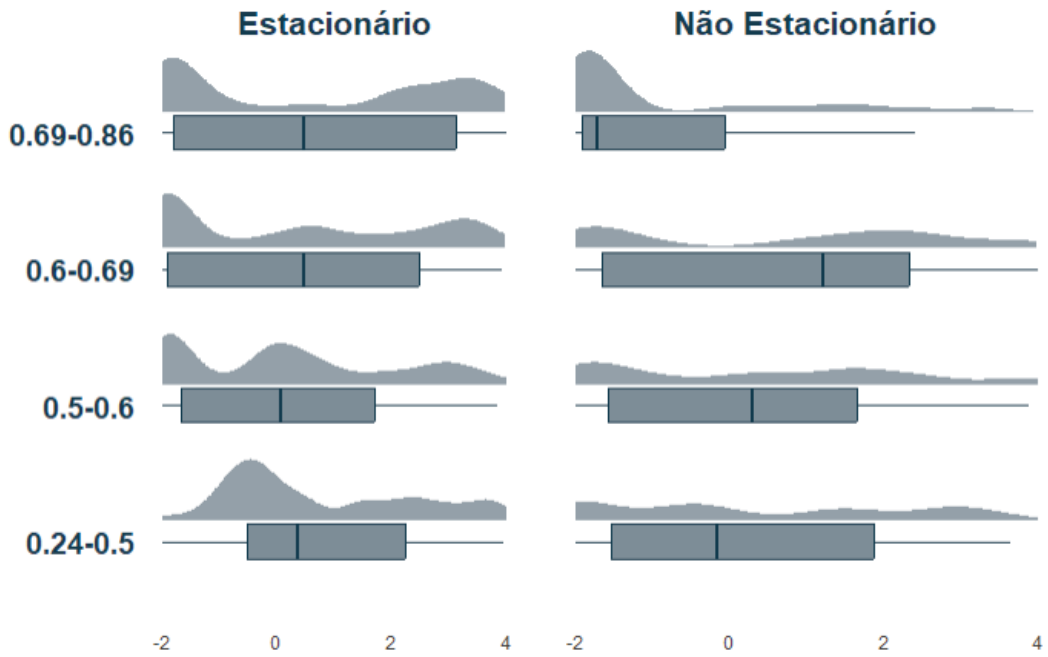


FIGURA 36 – Log(Cost) - SVR - Estacionariedade - Intermitente

Fonte: Elaboração própria

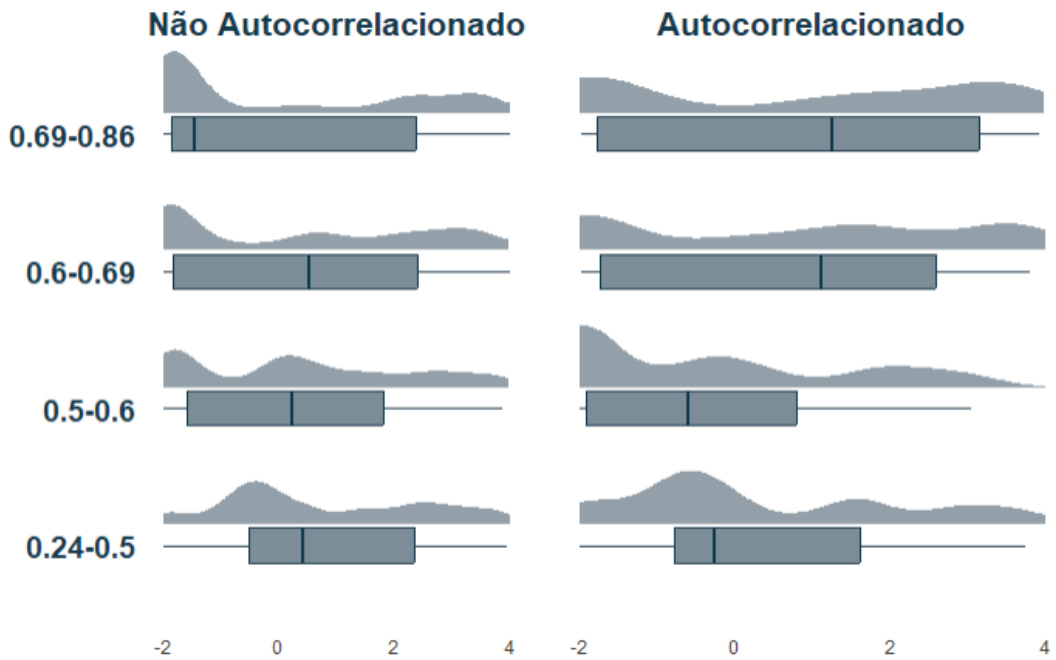


FIGURA 37 – Log(Cost) - SVR - Autocorrelação - Intermitente

Fonte: Elaboração própria

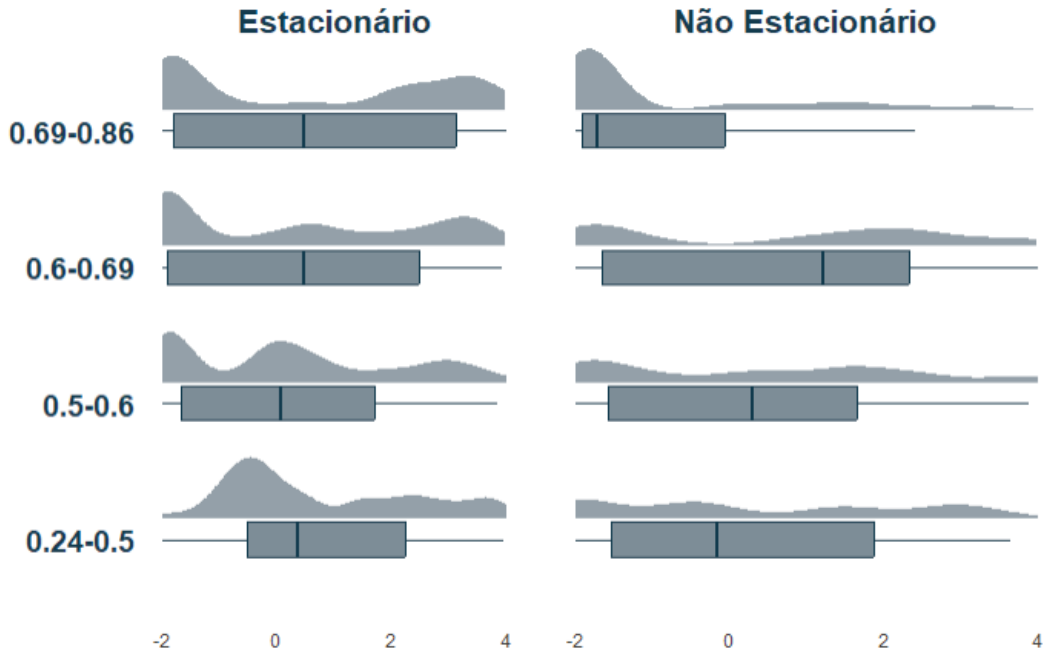


FIGURA 38 – Log(RbfSigma) - SVR - Estacionariedade - Intermitente
 Fonte: Elaboração própria

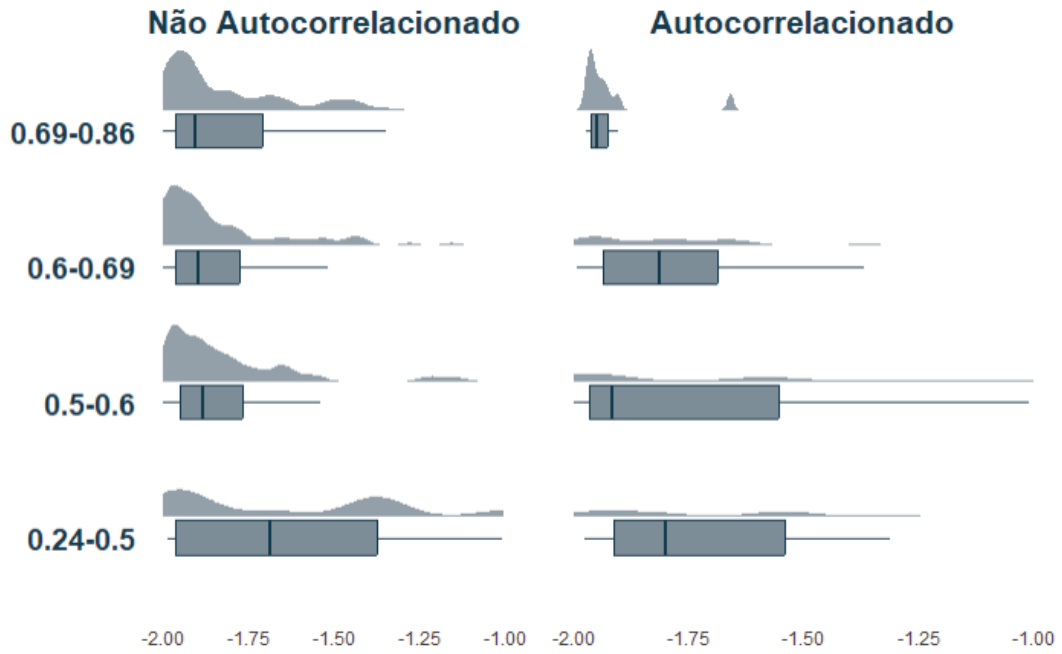


FIGURA 39 – Log(RbfSigma) - SVR - Autocorrelação - Intermitente
 Fonte: Elaboração própria

A.2 SUAVES

A.2.1 Arima Random Forest

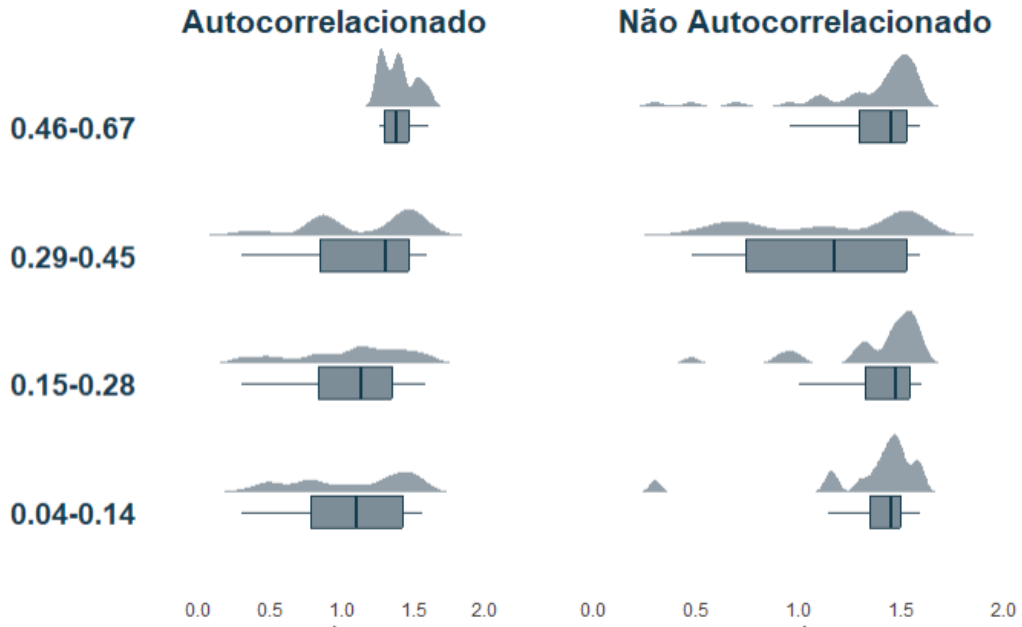


FIGURA 40 – Minn - Arima RF - Autocorrelação - Suave
 Fonte: Elaboração própria

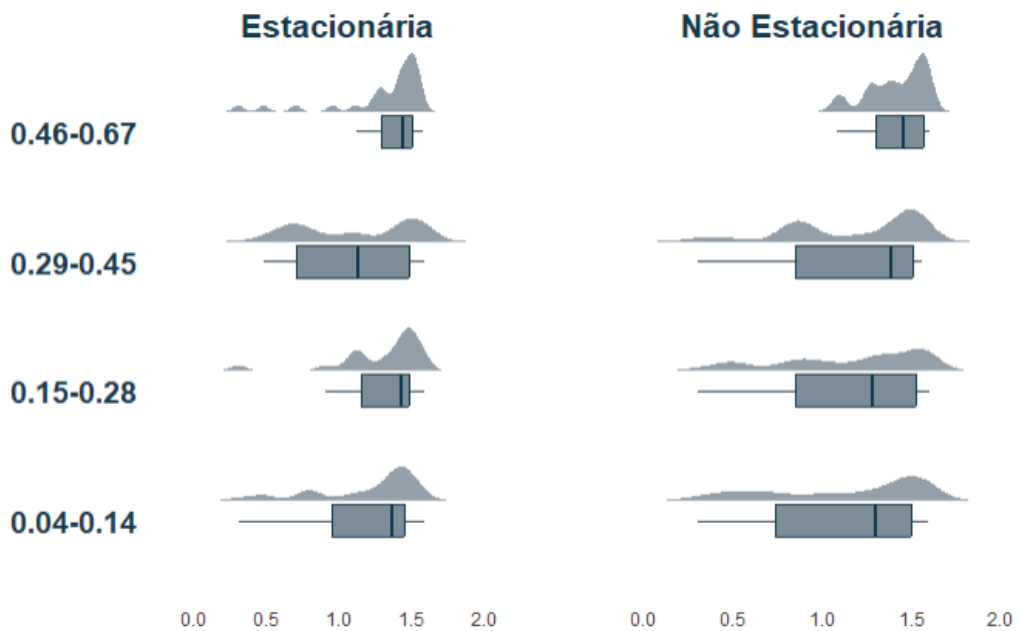


FIGURA 41 – Minn - Arima RF - Estacionariedade - Suave
 Fonte: Elaboração própria

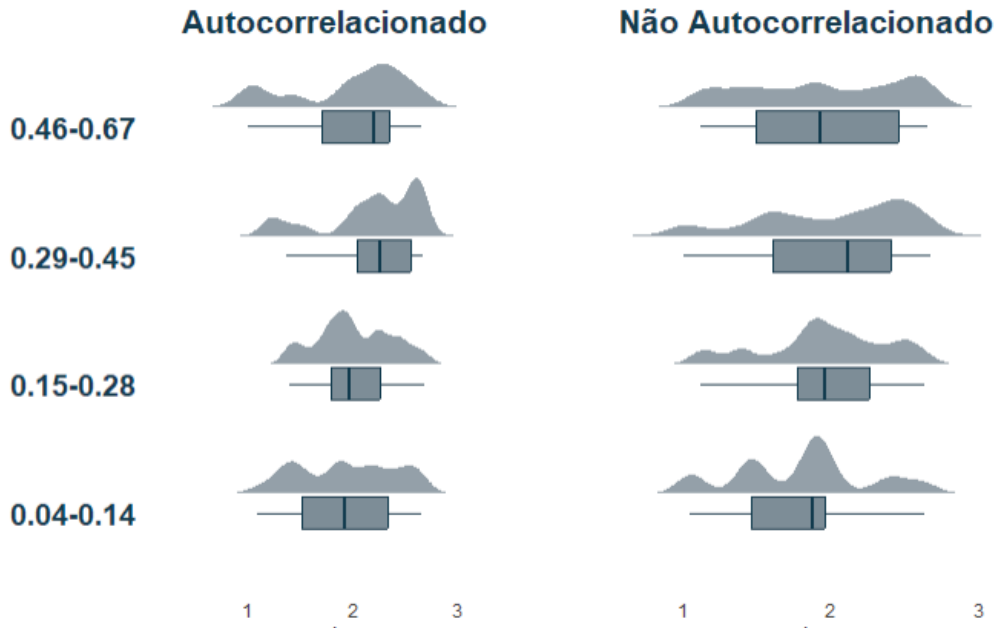


FIGURA 42 – Árvores - Arima RF - Autocorrelação - Suave
 Fonte: Elaboração própria

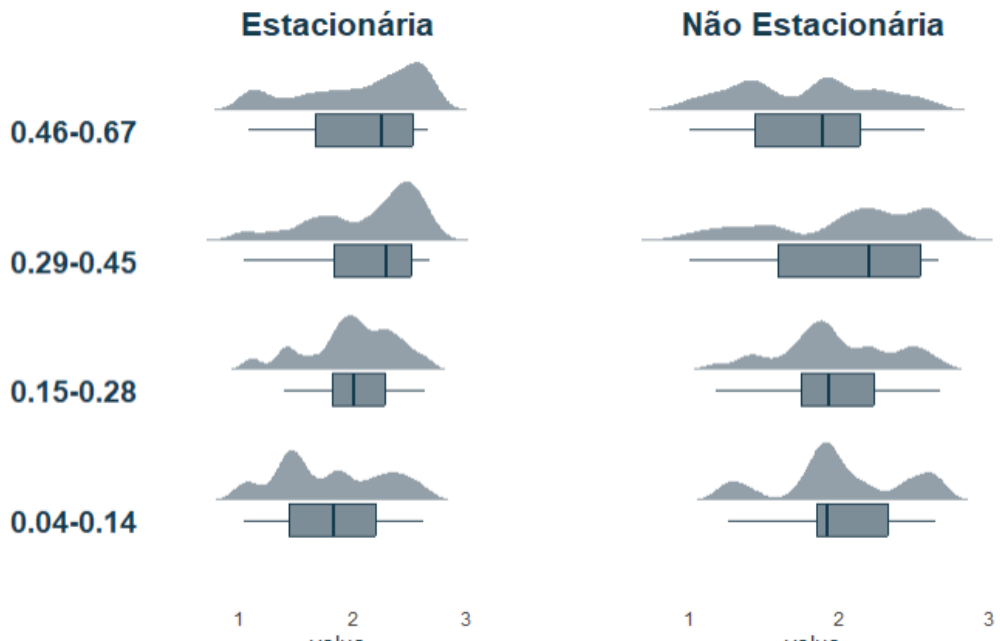


FIGURA 43 – Árvores - Arima RF - Estacionariedade - Suave
 Fonte: Elaboração própria

A.2.2 Arima XGboost

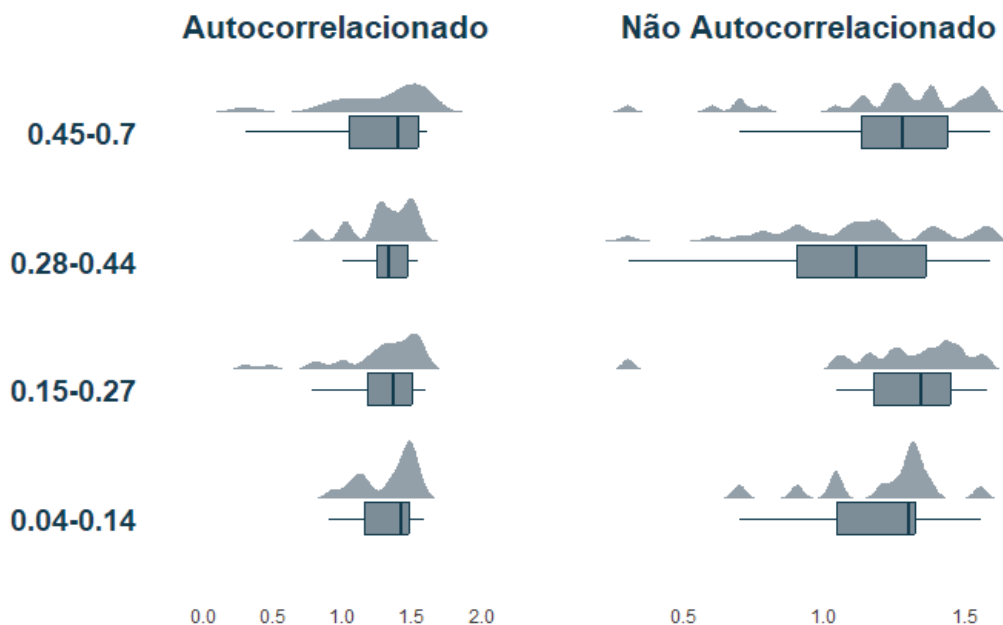


FIGURA 44 – Minn - Arima XGb - Autocorrelação - Suave
 Fonte: Elaboração própria

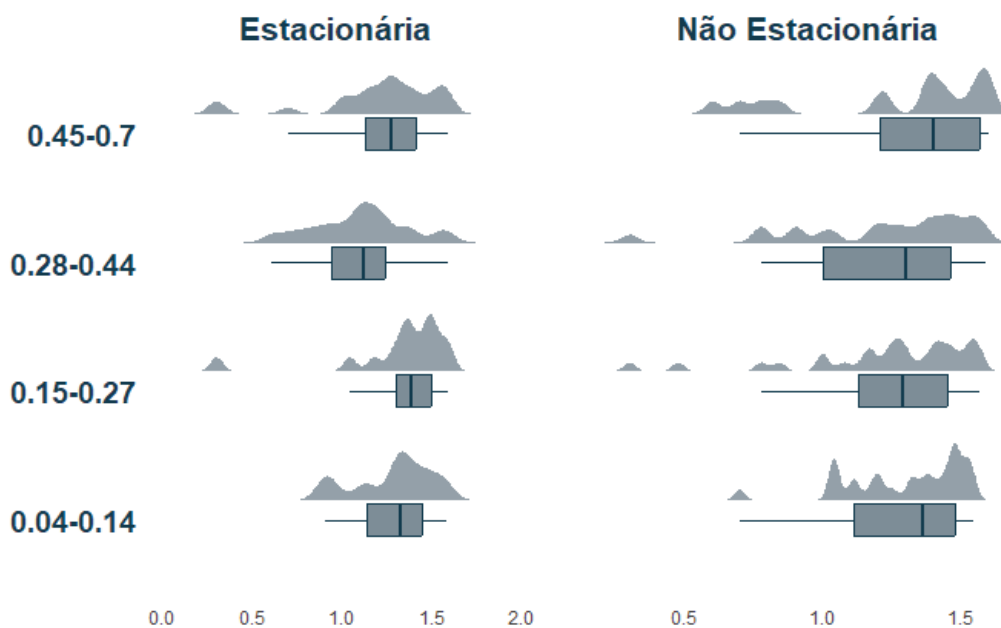


FIGURA 45 – Minn - Arima XGb - Estacionariedade - Suave
 Fonte: Elaboração própria

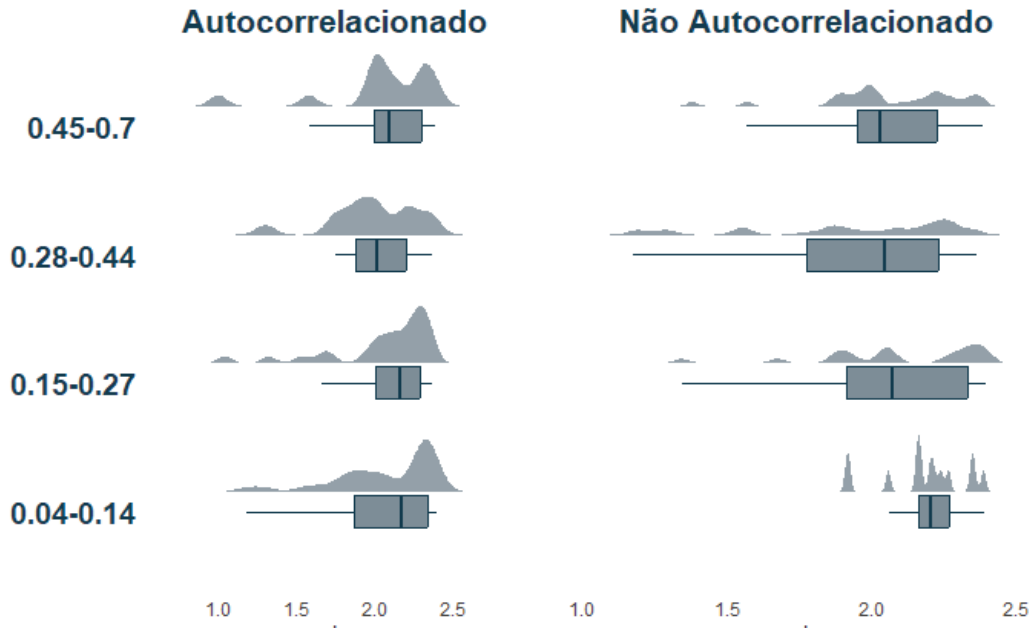


FIGURA 46 – Árvores - Arima XGb - Autocorrelação - Suave
 Fonte: Elaboração própria

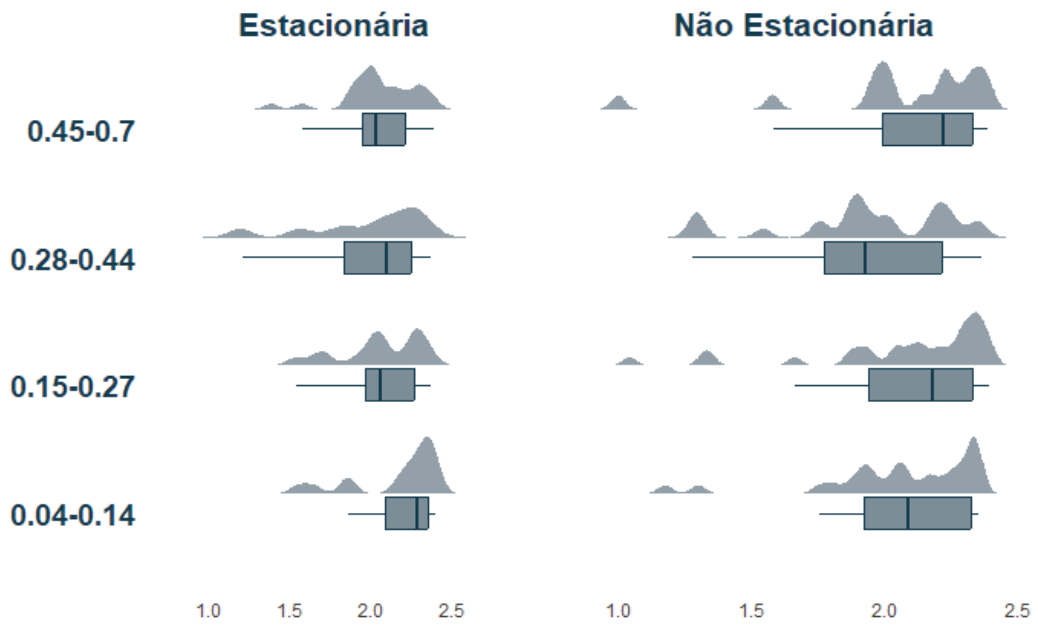


FIGURA 47 – Árvores - Arima XGb - Estacionariedade - Suave
 Fonte: Elaboração própria

A.2.3 Arima SVR

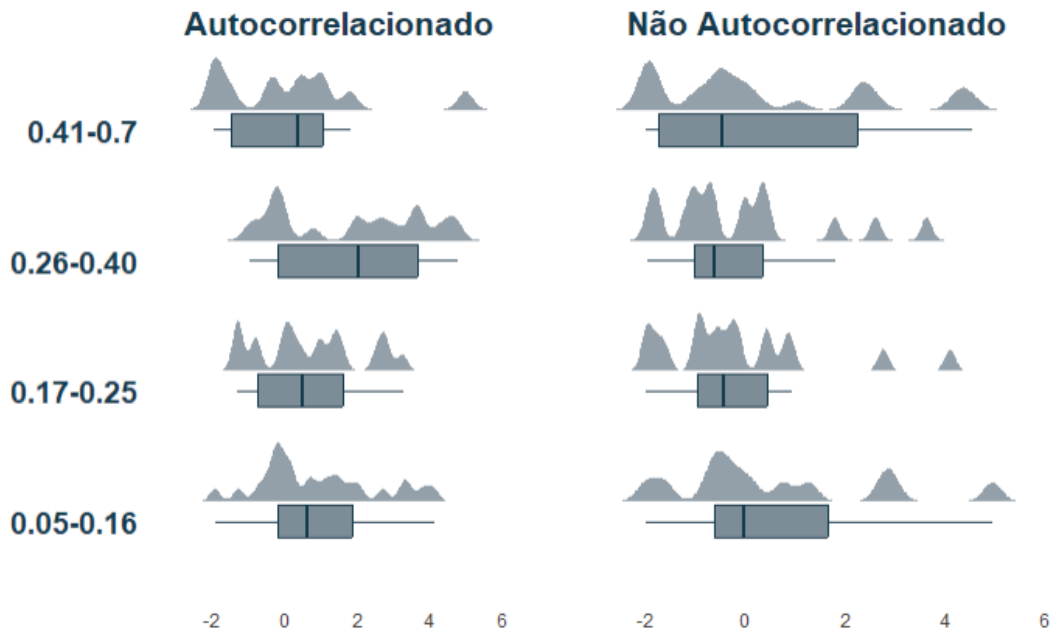


FIGURA 48 – Cost - Arima SVR - Autocorrelação - Suave
 Fonte: Elaboração própria

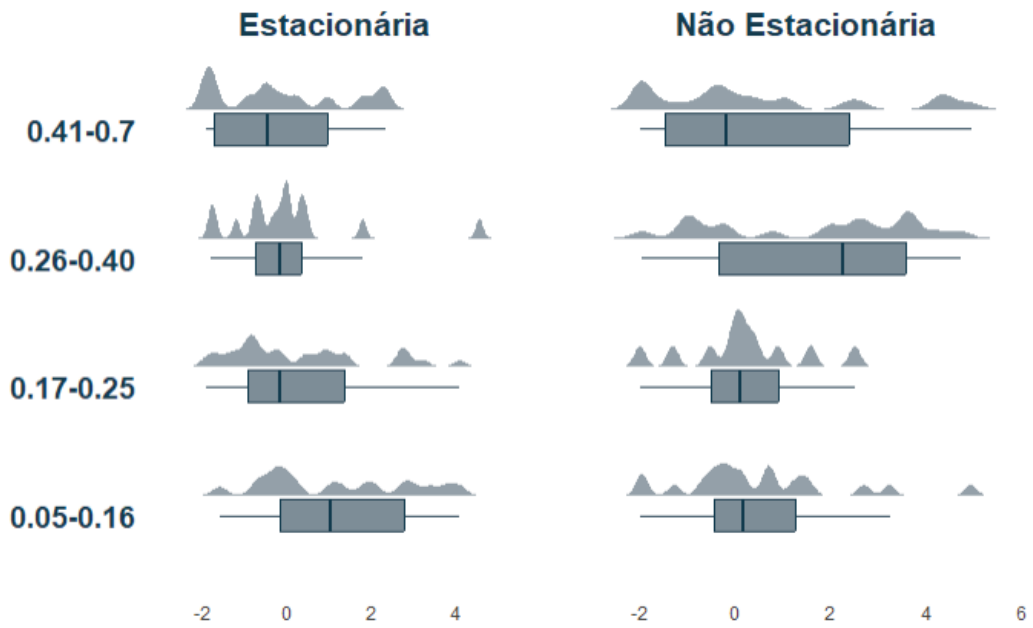


FIGURA 49 – Cost - Arima SVR - Estacionariedade - Suave
 Fonte: Elaboração própria

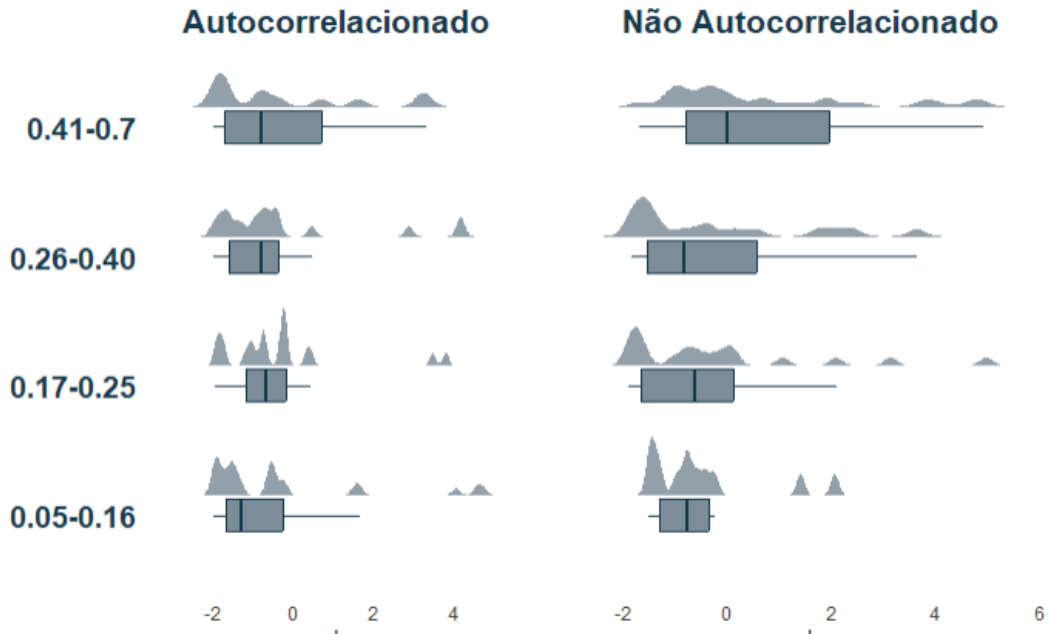


FIGURA 50 – RBF - Arima SVR - Autocorrelação - Suave
 Fonte: Elaboração própria

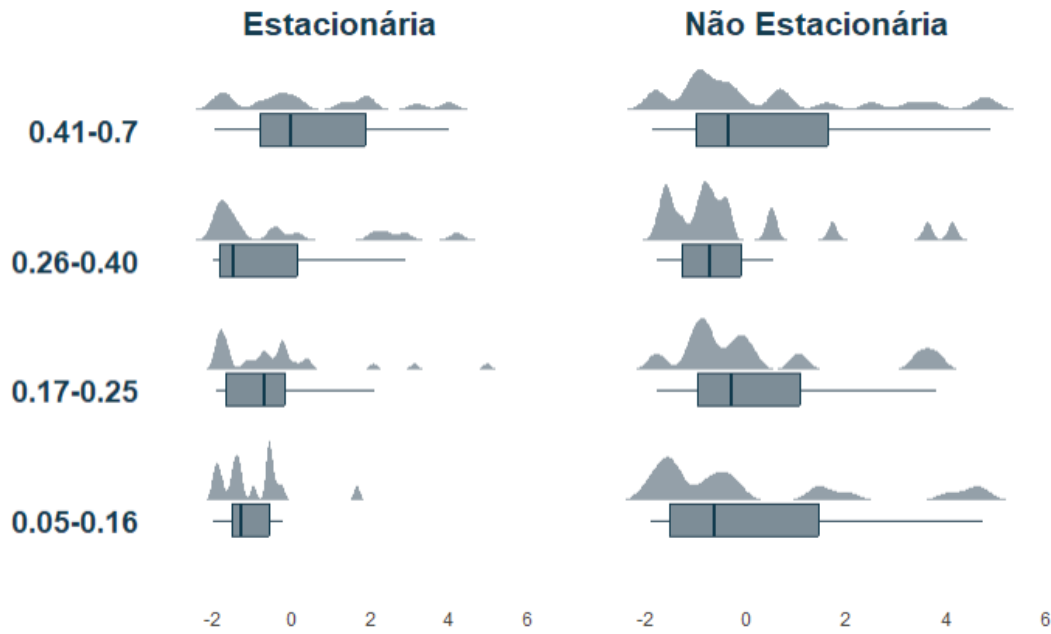


FIGURA 51 – RBF - Arima SVR - Estacionariedade - Suave
 Fonte: Elaboração própria

A.3 ERRÁTICOS

A.3.1 Glnet Poisson

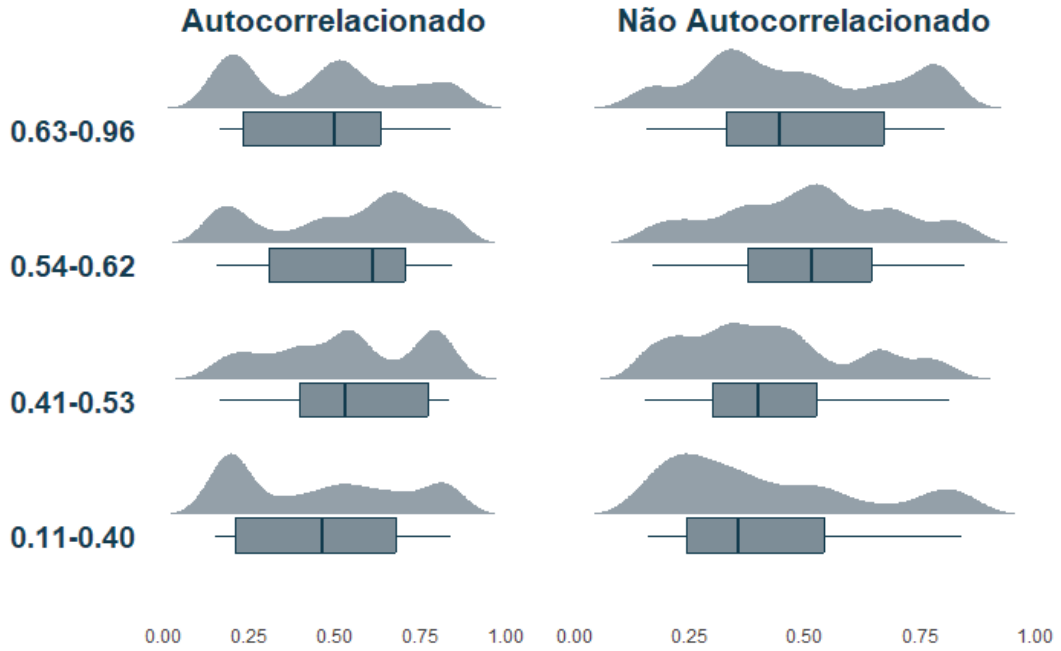


FIGURA 52 – Mixture - Glnet Poisson - Autocorrelação - Errático

Fonte: Elaboração própria

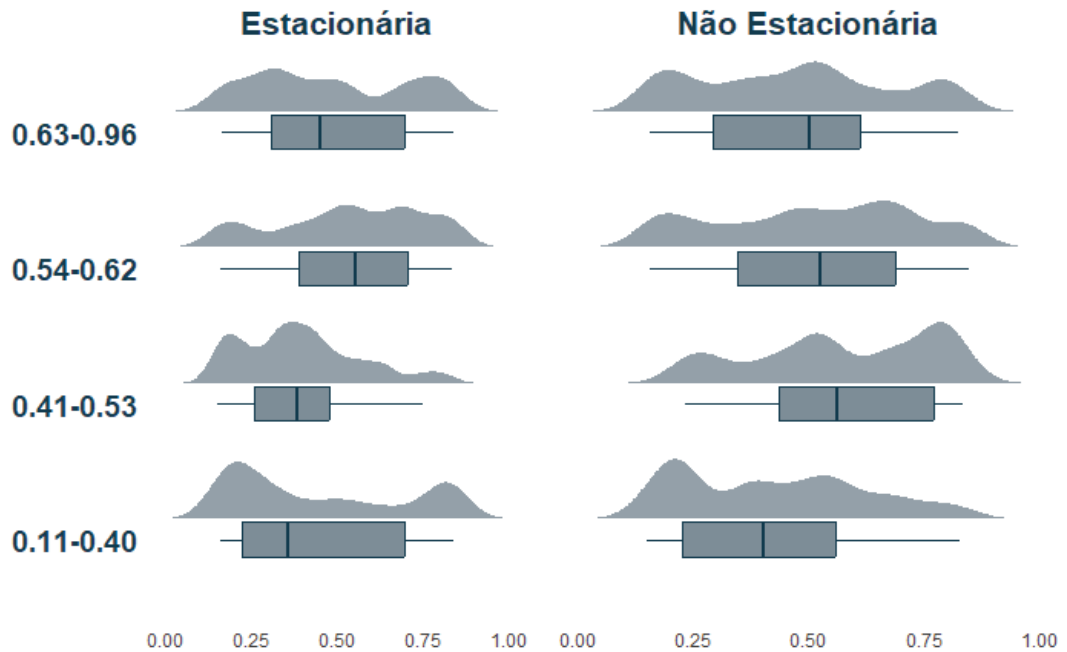


FIGURA 53 – Mixture - Poisson - Estacionariedade - Errático

Fonte: Elaboração própria

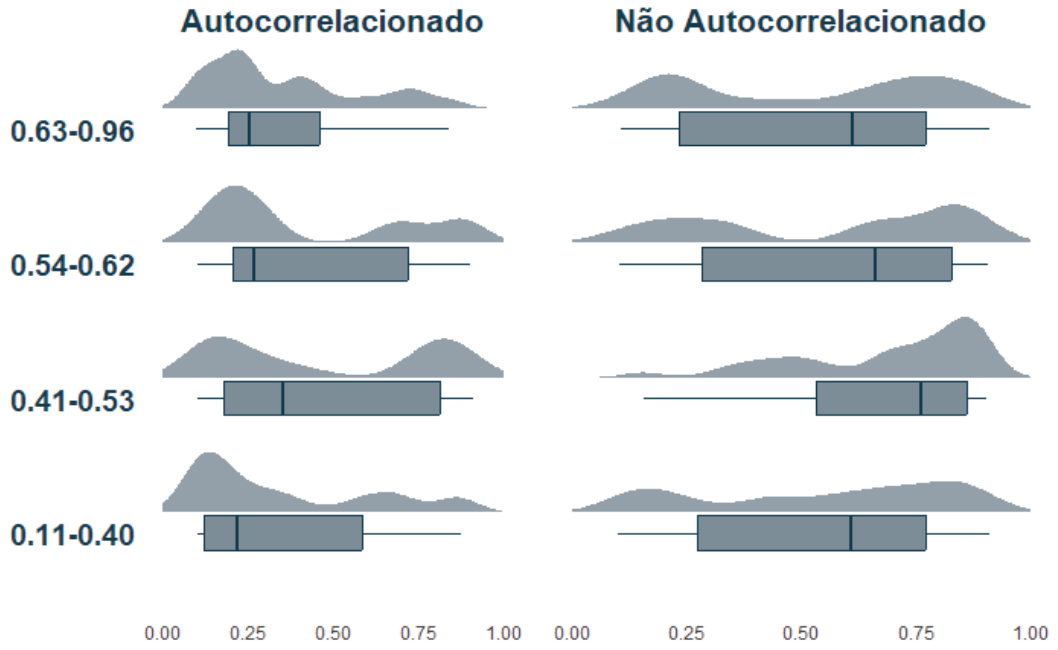


FIGURA 54 – Penalty - Glmnet Poisson - Autocorrelação - Errático
 Fonte: Elaboração própria

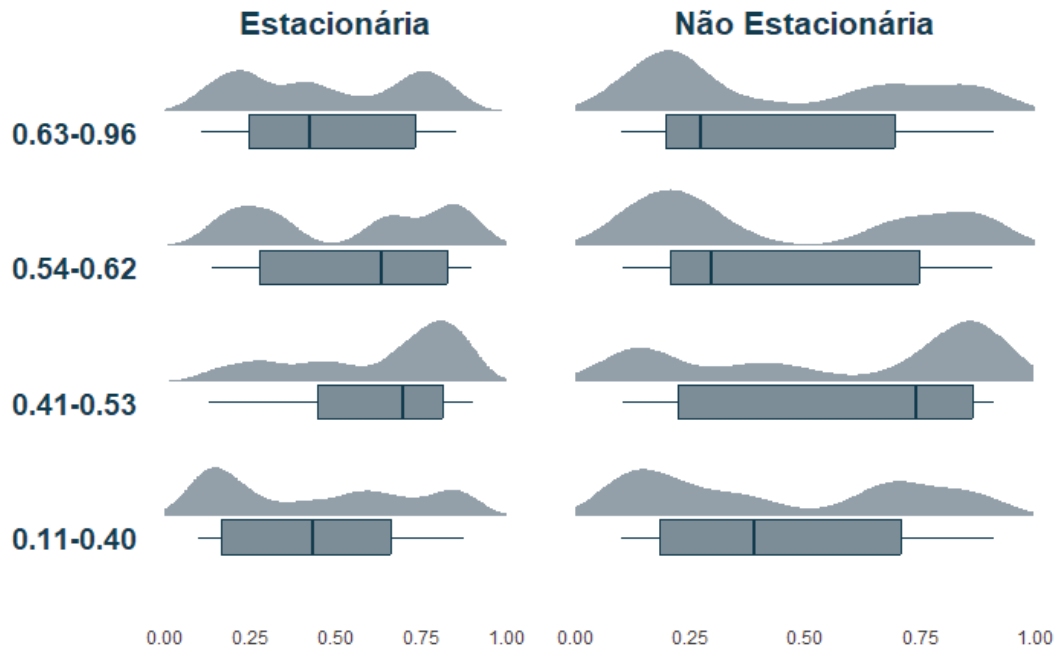


FIGURA 55 – Penalty - Glmnet Poisson - Estacionariedade - Errático
 Fonte: Elaboração própria

A.3.2 Arima XGboost

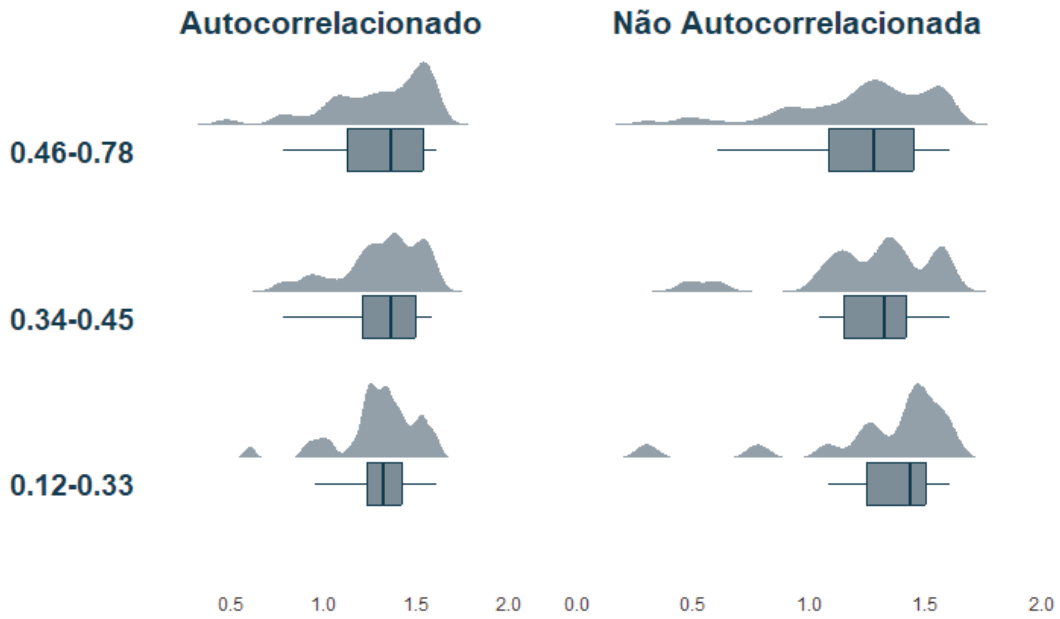


FIGURA 56 – Minn - Xgboost - Autocorrelação - Errático

Fonte: Elaboração própria

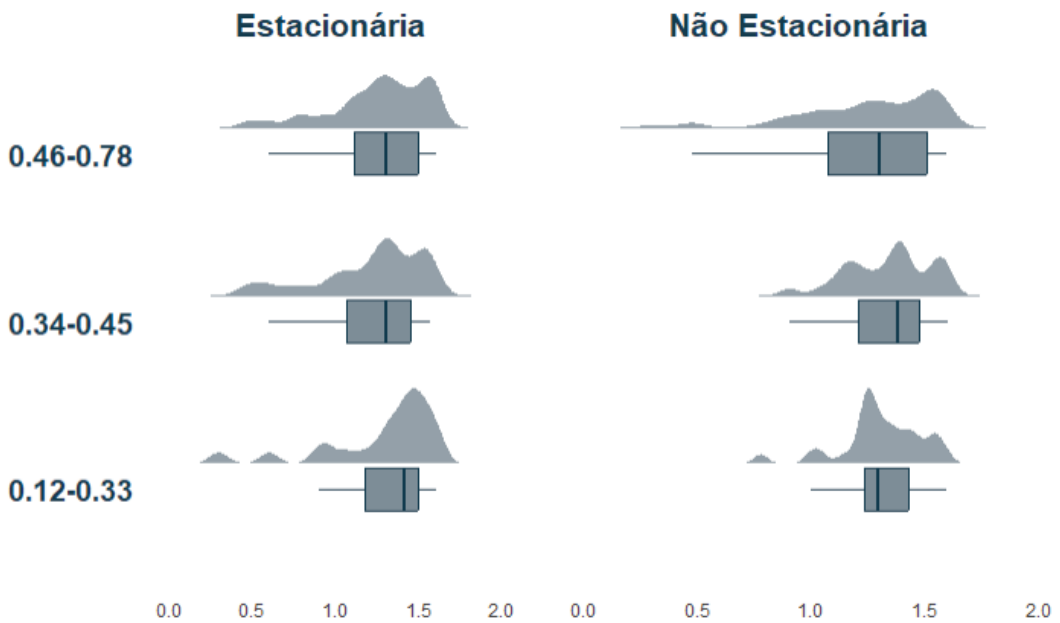


FIGURA 57 – Minn - Xgboost - Estacionariedade - Errático

Fonte: Elaboração própria

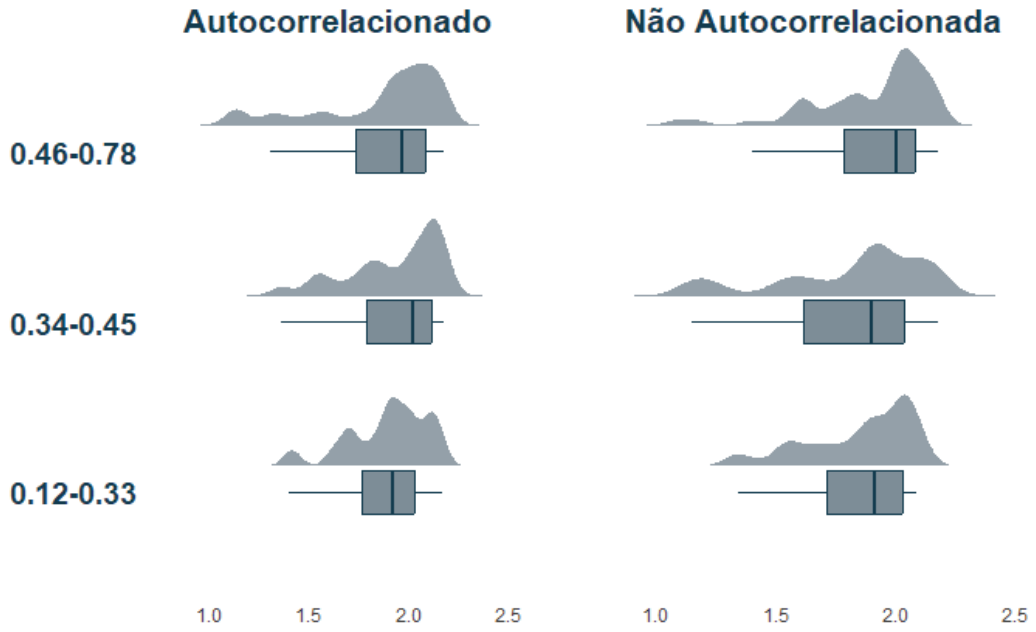


FIGURA 58 – Árvores - Xgboost - Autocorrelação - Errático
 Fonte: Elaboração própria

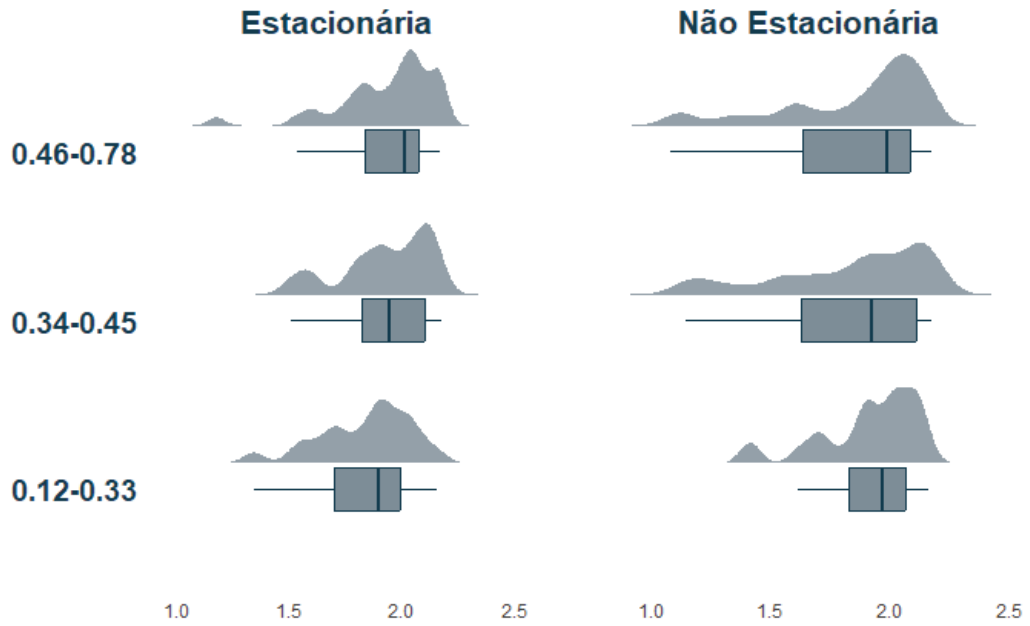


FIGURA 59 – Árvores - Xgboost - Estacionariedade - Errático
 Fonte: Elaboração própria

A.3.3 Arima SVR

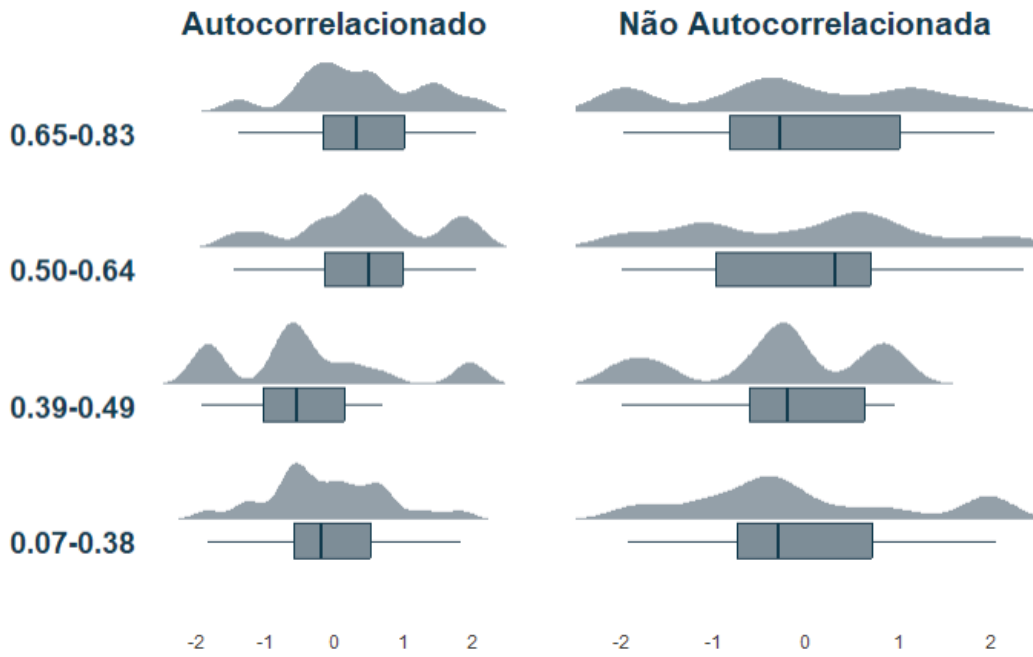


FIGURA 60 – Cost - SVR - Autocorrelação - Errático

Fonte: Elaboração própria

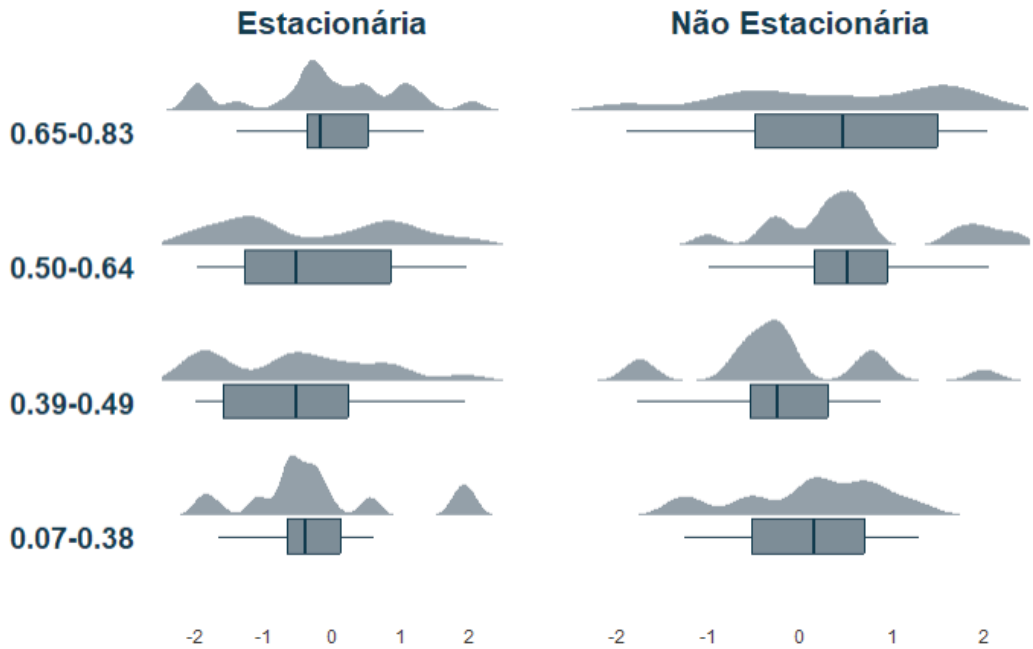


FIGURA 61 – Cost - SVR - Estacionariedade - Errático

Fonte: Elaboração própria

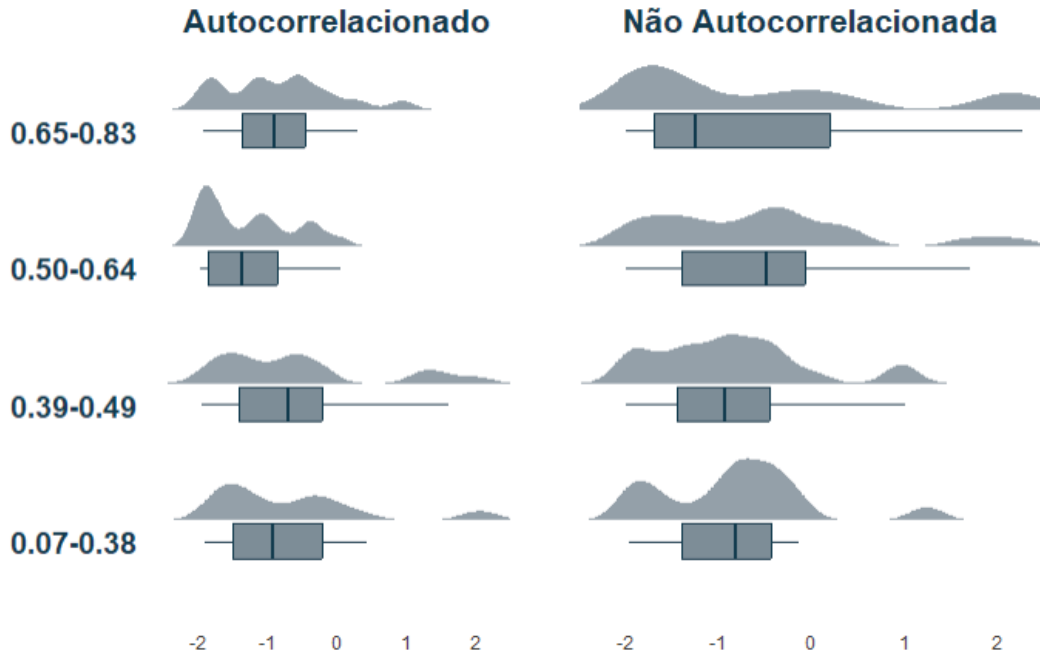


FIGURA 62 – RBFSigma - SVR - Autocorrelação - Errático
 Fonte: Elaboração própria

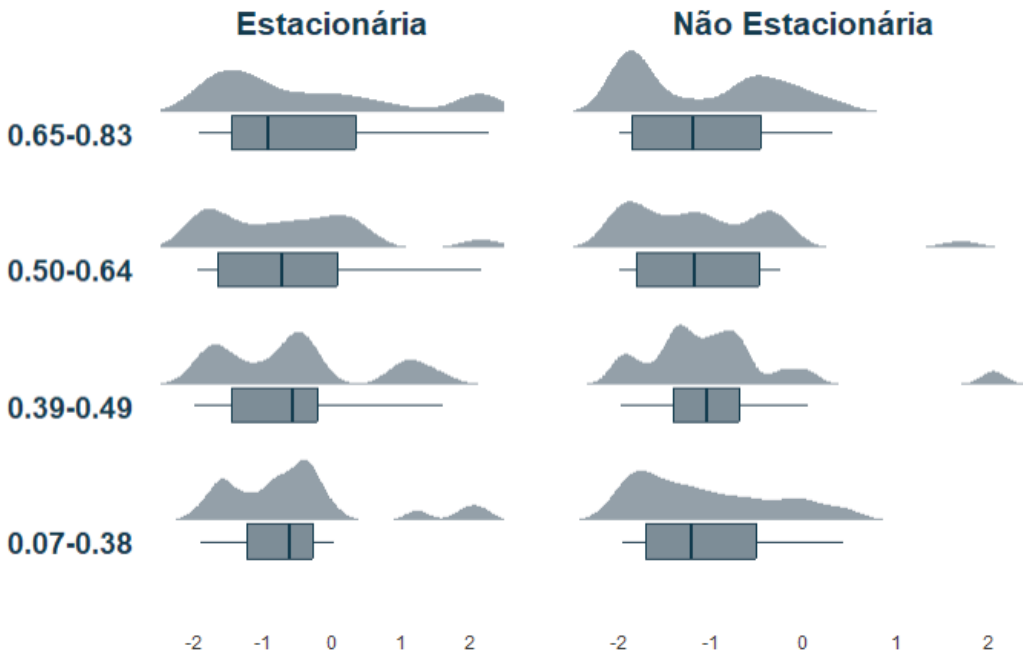


FIGURA 63 – RBFSigma - SVR - Estacionariedade - Errático
 Fonte: Elaboração própria

REFERÊNCIAS

AAMER, A. M.; YANI, L. P. E.; PRIYATNA, I. M. A. Data Analytics in the Supply Chain Management: Review of Machine Learning Applications in Demand Forecasting. en. **OPERATIONS AND SUPPLY CHAIN MANAGEMENT**, n. 14, p. 1–13, 2021. Citado 1 vez na página 30.

ABBASIMEHR, H.; BEHBOODI, A.; BAHIRINI, A. A novel hybrid model to forecast seasonal and chaotic time series. en. **Expert Systems With Applications**, Elsevier, v. 239, p. 122461, jan. 2024. Citado 1 vez na página 51.

ABOLGHASEMI, M.; HYNDMAN, R. J.; TARR, G.; BERGMEIR, C. Machine learning applications in time series hierarchical forecasting, 2019. Disponível em: <https://arxiv.org/abs/1912.00370>. Citado 1 vez na página 44.

ABOLGHASEMI, M.; ROSTAMI-TABAR, B.; SYNTETOS, A. The value of point of sales information in upstream supply chain forecasting: an empirical investigation. **International Journal of Production Research**, Melbourne, Australia, v. 61, n. 7, 2022. Citado 1 vez na página 33.

AHMADI, A.; DACCACHE, A.; SADEGH, M.; SNYDER, R. L. Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. **Computers and Electronics in Agriculture**, Elsevier, v. 215, p. 108424, nov. 2023. DOI: [10.1016/j.compag.2023.108424](https://doi.org/10.1016/j.compag.2023.108424). Citado 1 vez na página 35.

_____. _____. **Computers and Electronics in Agriculture**, Elsevier, v. 215, p. 108424, 2023. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). ISSN 0168-1699. DOI: [10.1016/j.compag.2023.108424](https://doi.org/10.1016/j.compag.2023.108424). Disponível em: <https://doi.org/10.1016/j.compag.2023.108424>. Citado 1 vez na página 70.

BABU, C.; REDDY, B. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. **Applied Soft Computing**, Elsevier, v. 23, p. 27–38, 2014. Citado 1 vez na página 35.

BARANDAS, M.; FOLGADO, D.; FERNANDES, L.; SANTOS, S.; ABREU, M.; BOTA, P.; LIU, H.; SCHULTZ, T.; GAMBOA, H. TSFEL: Time Series Feature Extraction Library.

SoftwareX, Elsevier, v. 11, p. 100456, 2020. Original software publication. DOI: [10.1016/j.softx.2020.100456](https://doi.org/10.1016/j.softx.2020.100456). Disponível em: <https://doi.org/10.1016/j.softx.2020.100456>. Citado 1 vez na página 71.

BAUER, A.; ZÜFLE, M.; GROHMANN, J.; SCHMITT, N.; HERBST, N.; KOUNEV, S. An Automated Forecasting Framework based on Method Recommendation for Seasonal Time Series. In: UNIVERSITY OF WÜRZBURG, GERMANY. PERFORMANCE Learning ICPE '20. Edmonton, AB, Canada: [s.n.], abr. 2020. DOI: [10.1145/3358960.3379123](https://doi.org/10.1145/3358960.3379123). Disponível em: <https://doi.org/10.1145/3358960.3379123>. Citado 1 vez na página 47.

BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, v. 191, p. 192–213, 2012. Citado 1 vez na página 59.

BERGMEIR, C.; HYNDMAN, R. J.; KOO, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. **Computational Statistics Data Analysis**, v. 120, p. 70–83, 2018. Citado 3 vezes nas páginas 52, 59, 61.

BERRAR, D. Overfitting. Edição: W. Dubitzky, O. Wolkenhauer, K.-H. Cho e H. Yokota. Springer, p. 1617–1619, 2013. Citado 1 vez na página 58.

BONTEMPI, G. Long Term Time Series Prediction with Multi-Input Multi-Output Local Learning, p. 145–154, 2008. Citado 3 vezes nas páginas 16, 22, 54.

BONTEMPI, G.; BEN TAIEB, S.; LE BORGNE, Y.-A. Machine Learning Strategies for Time Series Forecasting. **Machine Learning Group**, Bd Triomphe, 1050, Brussels, Belgium, 2013. Disponível em: <http://mlg.ulb.ac.be>. Citado 1 vez na página 36.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. [S.l.]: Holden-Day, 1970. Citado 1 vez na página 70.

BOYLAN, J.; SYNTETOS, A. Spare parts management: Forecasting research and extensions. **Ima Journal of Management Mathematics - IMA J MANAG MATH**, v. 21, p. 227–237, jul. 2010. DOI: [10.1093/imaman/dpp016](https://doi.org/10.1093/imaman/dpp016). Citado 1 vez na página 19.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324). Citado 1 vez na página 44.

BÜYÜKŞAHİN, Ü. Ç.; ERTEKİN, Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition.

Neurocomputing, v. 361, p. 151–164, 2019. Citado 1 vez na página 49.

CERQUEIRA, V.; TORGO, L.; MOZETIČ, I. Evaluating time series forecasting models: an empirical study on performance estimation methods. **Machine Learning**, v. 109, n. 3, p. 1997–2028, 2020. DOI: [10.1007/s10994-020-05910-7](https://doi.org/10.1007/s10994-020-05910-7). Disponível em: <https://doi.org/10.1007/s10994-020-05910-7>. Citado 3 vezes nas páginas 60, 61.

CERQUEIRA, V.; TORGO, L.; SOARES, C. Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters. en. **Notas Técnicas do LIAAD**, p. 1–9, 2019. Citado 1 vez na página 40.

CHAI, T.; DRAXLER, R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, Copernicus GmbH, v. 7, n. 3, p. 1247–1250, 2014. CC Attribution 3.0 License. DOI: [10.5194/gmd-7-1247-2014](https://doi.org/10.5194/gmd-7-1247-2014). Disponível em: <https://www.geosci-model-dev.net/7/1247/2014/>. Citado 1 vez na página 67.

CHEN, P.; LIU, S.; SHI, C.; HOOI, B.; WANG, B.; CHENG, X. NeuCast: Seasonal Neural Forecast of Power Grid Time Series. In: PROCEEDINGS of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018). Stockholm, Sweden: IJCAI, jul. 2018. P. 3315–3321. Citado 1 vez na página 70.

CHEN, Z.; DENG, C.; LI, P.; SHI, Y. The Bullwhip Effect, Demand Uncertainty, and Cost Structure. en. **Contemporary Accounting Research**, Wiley, 2024. Citado 1 vez na página 29.

CHENG, T. C. E.; KAMBLE, S. S.; BELHADI, A.; NDUBISI, N. O.; LAI, K.-h.; KHARAT, M. G. Linkages between Big Data Analytics, Circular Economy, Sustainable Supply Chain Flexibility, and Sustainable Performance in Manufacturing Firms. **International Journal of Production Research**, Volume da revista (se disponível), Número da edição (se disponível), p. 1–13, 2021. Published online: 05 Apr 2021. DOI: [10.1080/00207543.2021.1912254](https://doi.org/10.1080/00207543.2021.1912254). Disponível em: <https://www.tandfonline.com/doi/full/10.1080/00207543.2021.1912254>. Citado 1 vez na página 15.

CHINTAN PATIL, V. P. Supply Chain Cash-Flow Bullwhip Effect: An Empirical Investigation. en. **Journal of Supply Chain Management**, 2024. Citado 1 vez na página 29.

COX, D. Prediction by exponentially weighted moving averages and related methods. **Journal of the Royal Statistical Society. Series B, Statistical Methodology**, v. 23, p. 414–422, 1961. Citado 1 vez na página 54.

CROSTON, J. Forecasting and stock control for intermittent demands. **Journal of the Operational Research Society**, v. 23, p. 289–303, 1972. Citado 2 vezes nas páginas 25, 27.

DOLGUI, A.; PASHKEVICH, A.; PASHKEVICH, M. Bayesian Approach to Modelling of Quasi-Periodic Intermittent Demand. In: IFAC. 16TH Triennial World Congress. Prague, Czech Republic: [s.n.], 2005. Copyright (c) 2005 IFAC. All rights reserved. Citado 1 vez na página 21.

DUDA, R.; HART, P.; STORK, D. Pattern Classification. **John Wiley & Sons**, 2000. Citado 1 vez na página 58.

DUDEK, G. Short-Term Load Forecasting with Random Forest and Multiple Input Multiple Output (MIMO) Methods. **Energies**, MDPI, v. 15, n. 12, p. 7547, 2022. Citado 1 vez na página 45.

EAVES, A.; KINGSMAN, B. Forecasting for the ordering and stock-holding of spare parts. **Journal of the Operational Research Society**, v. 55, p. 431–437, 2004. Citado 1 vez na página 25.

ERJIANG; YU, M.; TIAN, X.; TAO, Y. Dynamic Model Selection Based on Demand Pattern Classification in Retail Sales Forecasting. Edição: Gheorghe Savoiu. **Mathematics**, Multidisciplinary Digital Publishing Institute, Switzerland, v. 10, n. 3179, 2022. Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations. Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). DOI: [10.3390/math10173179](https://doi.org/10.3390/math10173179). Citado 1 vez na página 28.

FERNANDES, B.; CORREIA, P.; SILVA, M.; PEREIRA, P. Equipment Health Monitoring in Nuclear Power Plants Using LSTM Neural Networks. **Applied Soft Computing**, Elsevier, v. 89, p. 106116, jun. 2020. DOI: [10.1016/j.asoc.2020.106116](https://doi.org/10.1016/j.asoc.2020.106116). Citado 1 vez na página 39.

FRANSES, P.; LEGERSTEE, R. A unifying view on multi-step forecasting using an autoregression. **Journal of Economic Surveys**, v. 24, p. 389–401, 2009. Citado 1 vez na página 54.

GRANGER, C. W. J. **Investigating Causal Relations by Econometric Models and Cross-Spectral Methods**. [S.l.]: North-Holland Publishing Company, 1969. Citado 1 vez na página 70.

GUO, L.; FANG, W.; ZHAO, Q.; WANG, X. The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. en. **Computers Industrial Engineering**, Elsevier, v. 161, p. 107598, ago. 2021. Citado 1 vez na página 50.

HASAN, M. R.; KABIR, M. A.; SHUVRO, R. A.; DAS, P. A Comparative Study on Forecasting of Retail Sales, 2021. Citado 1 vez na página 33.

HEWAMALAGE, H.; ACKERMANN, K.; BERGMEIR, C. Forecast evaluation for data scientists: common pitfalls and best practices. **Data Mining and Knowledge Discovery**, v. 37, 788–832, 2023. DOI: [10.1007/s10618-022-00894-5](https://doi.org/10.1007/s10618-022-00894-5). Disponível em: <https://doi.org/10.1007/s10618-022-00894-5>. Citado 1 vez na página 62.

HILL, H. A.; JAIN, P.; OK, C. Y.; SASAKI, K.; CHEN, H.; WANG, M. L.; CHEN, K. Integrative Prognostic Machine Learning Models in Mantle Cell Lymphoma. **Cancer Research**, American Association for Cancer Research, v. 2767-9764, 2023. DOI: [10.1158/2767-9764.CRC-23-0083](https://doi.org/10.1158/2767-9764.CRC-23-0083). Disponível em: <https://doi.org/10.1158/2767-9764.CRC-23-0083>. Citado 1 vez na página 65.

HO, T. Random Decision Forests. In: PROCEEDINGS of the 3rd International Conference on Document Analysis and Recognition. [S.l.: s.n.], 1995. Vol. 1, p. 278–282. Citado 1 vez na página 40.

HOND, T.; FAN, S. Probabilistic electric load forecasting: A tutorial review. **International Journal of Forecasting**, v. 32, n. 3, p. 914–938, jul. 2016. DOI: [10.1016/j.ijforecast.2015.11.011](https://doi.org/10.1016/j.ijforecast.2015.11.011). Disponível em: <https://doi.org/10.1016/j.ijforecast.2015.11.011>. Citado 1 vez na página 70.

HOPP, W. J.; SPEARMAN, M. L. **Factory Physics: Foundations of Manufacturing Management**. 3rd. New York, NY: McGraw-Hill Education, 2008. Citado 1 vez na página 28.

HU, X.; ZHANG, X.; ZHANG, D. A Novel Approach of Intermittent Demand Prediction in Industrial Domain. **Journal of Physics: Conference Series**, IOP Publishing, Shanghai, China, v. 1756, p. 012009, 2021. DOI: [10.1088/1742-6596/1756/1/012009](https://doi.org/10.1088/1742-6596/1756/1/012009). Citado 1 vez na página 28.

HYNDMAN, R.; KOEHLER, A.; ORD, J.; SNYDER, R. **Forecasting with Exponential Smoothing: The State Space Approach**. [S.l.]: Springer-Verlag, 2008. Citado 1 vez na página 35.

HYNDMAN, R. J. **Errors on percentage errors**. [S.l.: s.n.], 2014. Available at <https://robjhyndman.com/hyndsight/smape/>. Accessed on 02 July 2023. Citado 1 vez na página 68.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, Elsevier, v. 22, n. 3, p. 679–688, 2006. DOI: [10.1016/j.ijforecast.2006.03.001](https://doi.org/10.1016/j.ijforecast.2006.03.001). Citado 3 vezes nas páginas 40, 67, 68.

INTERNATIONAL MONETARY FUND. **World Economic Outlook: Navigating Global Divergences**. Washington, DC: [s.n.], 2018. November 10, 2018. Citado 1 vez na página 72.

JOHNSTON, F.; BOYLAN, J. Forecasting for items with intermittent demand. **Journal of the Operational Research Society**, v. 47, p. 113–121, 1996. Citado 1 vez na página 25.

JUNAID, M.; ALI, S.; EID, F.; EL-SAPPAGH, S.; ABUHMED, T. Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. **Computer Methods and Programs in Biomedicine**, v. 234, p. 107495, 2023. DOI: [10.1016/j.cmpb.2023.107495](https://doi.org/10.1016/j.cmpb.2023.107495). Disponível em: <https://www.elsevier.com/locate/cmpb>. Citado 1 vez na página 60.

KANNAN, B. A.; KODI, G.; PADILLA, O.; GRAY, D.; SMITH, B. C. Forecasting Spare Parts Sporadic Demand Using Traditional Methods and Machine Learning - a Comparative Study. **SMU Data Science Review**, v. 3, n. 2, article 9, 2020. Part of the

Data Science Commons. Disponível em: <https://scholar.smu.edu/datasciencereview>. Citado 1 vez na página 30.

KILIMCI, Z. H.; AKYUZ, A. O.; UYSAL, M.; AKYOKUS, S.; UYSAL, M. O.; BULBUL, B. A.; EKMIS, M. A. An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain. **Complexity**, Hindawi, v. 2019, article id 9067367, 15 pages, 2019. DOI: [10.1155/2019/9067367](https://doi.org/10.1155/2019/9067367). Disponível em: <https://doi.org/10.1155/2019/9067367>. Citado 1 vez na página 42.

KIM, H.-K.; RYAN, J. K. The Cost Impact of Using Simple Forecasting Techniques in a Supply Chain. en. **Naval Research Logistics**, Wiley, n. 50, p. 388–411, nov. 2003. Citado 1 vez na página 28.

KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, Elsevier, v. 32, n. 3, p. 669–679, 2016. DOI: [10.1016/j.ijforecast.2015.09.011](https://doi.org/10.1016/j.ijforecast.2015.09.011). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207015002070>. Citado 3 vezes nas páginas 67, 68.

KIM, S.; ALIZAMIR, M.; KIM, N. W.; KISI, O. Bayesian Model Averaging: A Unique Model Enhancing Forecasting Accuracy for Daily Streamflow Based on Different Antecedent Time Series. **Sustainability**, v. 12, p. 9720, 2020. Received: 19 September 2020; Accepted: 19 November 2020; Published: 21 November 2020. DOI: [10.3390/su12229720](https://doi.org/10.3390/su12229720). Disponível em: <https://www.mdpi.com/journal/sustainability>. Citado 1 vez na página 49.

KOLASSA, S. Evaluating predictive count data distributions in retail sales forecasting. **International Journal of Forecasting**, Elsevier, v. 32, n. 3, p. 788–803, 2016. DOI: <https://doi.org/10.1016/j.ijforecast.2015.12.004>. Citado 1 vez na página 37.

KUHN, M.; WESTON, S.; KEEFER, C.; COULTER, N.; QUINLAN, R. Cubist: rule-and instance based regression modeling. **R package version 0.0.18; CRAN: Vienna, Austria**, 2014. Citado 1 vez na página 46.

LAINDER, A. D.; WOLFINGER, R. D. Forecasting with gradient boosted trees: augmentation, tuning, and cross-validation strategies. **International Journal of Forecasting**, Elsevier, v. 38, p. 1426–1433, 2022. ISSN 0169-2070. DOI:

[10.1016/j.ijforecast.2021.12.003](https://doi.org/10.1016/j.ijforecast.2021.12.003). Disponível em:
<https://doi.org/10.1016/j.ijforecast.2021.12.003>. Citado 1 vez na página 60.

LI, J.; CAI, J.; LI, R.; LI, Q.; ZHENG, L. Wavelet Transforms Based ARIMA-XGBoost Hybrid Method for Layer Actions Response Time Prediction of Cloud GIS Services. **Journal of Cloud Computing: Advances, Systems and Applications**, Springer, v. 12, p. 11, jun. 2023. DOI: [10.1186/s13677-022-00360-z](https://doi.org/10.1186/s13677-022-00360-z). Citado 1 vez na página 35.

LINES, J.; TAYLOR, S.; BAGNALL, A. Time Series Classification with HIVE-COTE: The Hierarchical Vote Collective of Transformation-Based Ensembles. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM, v. 12, n. 5, p. 35, jul. 2018. DOI: [10.1145/3182382](https://doi.org/10.1145/3182382). Disponível em: <https://doi.org/10.1145/3182382>. Citado 1 vez na página 61.

LUO, T.; LIU, Y. Machine truth serum: a surprisingly popular approach to improving ensemble methods. **Machine Learning**, v. 112, p. 789–815, 2023. Received: 4 November 2021 / Revised: 11 February 2022 / Accepted: 9 May 2022 / Published online: 12 July 2022. DOI: [10.1007/s10994-022-06183-y](https://doi.org/10.1007/s10994-022-06183-y). Disponível em: <https://doi.org/10.1007/s10994-022-06183-y>. Citado 1 vez na página 40.

MADSEN, H. **Time Series Analysis**. 1st. Boca Raton, FL, USA: Chapman e Hall/CRC, 2007. Corresponding author: Henrik Madsen, Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, hmad@dtu.dk. Citado 1 vez na página 23.

MAKRIDAKIS, S. Accuracy measures: theoretical and practical concerns. **International Journal of Forecasting**, Elsevier, v. 9, n. 4, p. 527–529, 1993. Citado 1 vez na página 67.

MCAFEE, A.; BRYNJOLFSSON, E.; DAVENPORT, T. H.; PATIL, D. J.; BARTON, D. Big Data: The Management Revolution. **Harvard Business Review**, v. 90, n. 10, p. 60–68, 2012. Citado 1 vez na página 15.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. **Technometrics**, Taylor & Francis, v. 21, n. 2, p. 239–245, 1979. Citado 2 vezes nas páginas 62, 63.

MEHDIZADEH, S.; SALES, A. K.; SAFARI, M. J. S. Estimating the short-term and long-term wind speeds: implementing hybrid models through coupling machine learning and linear time series models. **SN Applied Sciences**, Springer Nature Switzerland AG, v. 2, n. 6, p. 1048, 2020. DOI: [10.1007/s42452-020-2830-0](https://doi.org/10.1007/s42452-020-2830-0). Citado 1 vez na página 49.

MENÉNDEZ-GARCÍA, L. A.; GARCÍA-NIETO, P. J.; GARCÍA-GONZALO, E.; SÁNCHEZ LASHERAS, F. Time series analysis for COMEX platinum spot price forecasting using SVM, MARS, MLP, VARMA and ARIMA models: A case study. en. **Resources Policy**, Elsevier, v. 95, p. 105148, jun. 2024. Citado 1 vez na página 49.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 3ª. São Paulo, Brazil: Blucher, 2018. Citado 1 vez na página 31.

MUEHLENSTAEDT, T.; MATTSON, C. A. Latin Hypercube Sampling in Practice. **Technometrics**, v. 48, n. 4, p. 499–510, 2006. DOI: [10.1198/004017005000000593](https://doi.org/10.1198/004017005000000593). Disponível em: <https://doi.org/10.1198/004017005000000593>. Citado 1 vez na página 62.

MURRAY, C.; CHAURASIA, P.; HOLLYWOOD, L.; COYLE, D. A Comparative Analysis of State-of-the-Art Time Series Forecasting Algorithms. In: IEEE. 2022 International Conference on Computational Science and Computational Intelligence (CSCI). [S.l.]: IEEE, 2022. DOI: [10.1109/CSCI58124.2022.00021](https://doi.org/10.1109/CSCI58124.2022.00021). Citado 1 vez na página 61.

NIE, H.; LIU, G.; LIU, X.; WANG, Y. Hybrid of ARIMA and SVMs for Short-Term Load Forecasting. **Energy Procedia**, Elsevier, v. 16, p. 1455–1460, 2012. Citado 1 vez na página 70.

NIKOLOPOULOS, K. We need to talk about intermittent demand forecasting. **European Journal of Operational Research**, 2020. Article history: Received 1 March 2019; Accepted 27 December 2019; Available online xxx. Keywords: Forecasting, Ubiquitousness, Intermittent demand, Special events, Foresight. Disponível em: <http://www.elsevier.com/locate/ejor>. Citado 1 vez na página 72.

NIU, W.-j.; FENG, Z.-k. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. **Sustainable Cities and Society**, Elsevier, v. 64, p. 102562, 2021. DOI: [10.1016/j.scs.2020.102562](https://doi.org/10.1016/j.scs.2020.102562). Citado 1 vez na página 42.

OHNO, T. **Toyota Production System: Beyond Large-Scale Production**. Portland, OR: Productivity Press, 1988. Citado 1 vez na página 28.

PAI, P.-F.; LIN, C.-S. A hybrid ARIMA and support vector machines model in stock price forecasting. **Neurocomputing**, v. 33, p. 497–505, 2004. Citado 1 vez na página 49.

PANDA, S. K.; MOHANTY, S. N. Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis. en. **Journal Name**, Publisher Name, Department of Artificial Intelligence et al., v. 11, 2023. Citado 1 vez na página 43.

PANIGRAHI, S.; BEHERA, H. A hybrid ETS–ANN model for time series forecasting. **Engineering Applications of Artificial Intelligence**, v. 66, p. 49–59, 2017. Disponível em: <https://www.elsevier.com/locate/engappai>. Citado 2 vezes nas páginas 34, 49.

PERREY, J.; SPILLECKE, D.; UMBLIJS, A. Smart Analytics: How Marketing Drives Short-Term and Long-Term Growth. **McKinsey Quarterly**, p. 00425–3, 2013. Citado 1 vez na página 15.

PTOTIĆ, M. P.; STOJANOVIĆ, M. B.; POPOVIĆ, P. M. A Review of Machine Learning Methods for Long-Term Time Series Prediction. In: IEEE. 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST). Ohrid, North Macedonia: IEEE, jun. 2022. DOI: [10.1109/ICEST55168.2022.9828618](https://doi.org/10.1109/ICEST55168.2022.9828618). Citado 1 vez na página 54.

PUNIA, S.; NIKOLOPOULOS, K.; SINGH, S. P.; MADAAN, J. K.; LITSIU, K. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. **International Journal of Production Research**, 2020. DOI: [10.1080/00207543.2020.1735666](https://doi.org/10.1080/00207543.2020.1735666). Disponível em: <https://doi.org/10.1080/00207543.2020.1735666>. Citado 1 vez na página 45.

QIAO, W.; MOAYEDI, H.; FOONG, L. K. Nature-inspired hybrid techniques of IWO, DA, ES, GA, and ICA, validated through a k-fold validation process predicting monthly natural gas consumption. **Energy & Buildings**, v. 217, p. 110023, 2020. Disponível em: <https://www.elsevier.com/locate/enbuild>. Citado 2 vezes nas páginas 38, 49.

RAAB, D.; THESSLER, A.; SPILIOPOULOU, M. XAI4EEG: spectral and spatio-temporal explanation of deep learning-based seizure detection in EEG time series. **Neural Computing and Applications**, v. 35, n. 100, p. 10051–10068, 2023.

DOI: [10.1007/s00521-022-07809-x](https://doi.org/10.1007/s00521-022-07809-x). Disponível em:
<https://doi.org/10.1007/s00521-022-07809-x>. Citado 1 vez na página 60.

RAFATI, L. The Bullwhip Effect in Supply Chains: Review of Recent Development. en. **Journal of Forecasting and Supply Chain Management**, Elsevier, 2022. Citado 1 vez na página 29.

RIAHI, Y.; SAIKOUK, T.; GUNASEKARAN, A.; BADRAOUI, I. Artificial Intelligence Applications in Supply Chain: A Descriptive Bibliometric Analysis and Future Research Directions. **Expert Systems With Applications**, Elsevier, v. 173, p. 114702, 2021. DOI: [10.1016/j.eswa.2021.114702](https://doi.org/10.1016/j.eswa.2021.114702). Citado 1 vez na página 15.

ROHAETI, E.; KURNIA, A.; SADIK, K.; SUMERTAJAYA, I. M. Generalized Linear Model Approach for Time Series Count Data on Number of Foreign Tourists Modeling in West Java. en. **Journal of Physics: Conference Series**, IOP Publishing, v. 1863, 2021. DOI: [10.1088/1742-6596/1863/1/012025](https://doi.org/10.1088/1742-6596/1863/1/012025). Citado 1 vez na página 38.

SAHOO, D.; SOOD, N.; RANI, U.; ABRAHAM, G.; DUTT, V.; DILEEP, A. Comparative Analysis of Multi-Step Time-Series Forecasting for Network Load Dataset. In: IEEE. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Kharagpur, India: IEEE, jul. 2020. DOI: [10.1109/ICCCNT49239.2020.9225449](https://doi.org/10.1109/ICCCNT49239.2020.9225449). Citado 1 vez na página 54.

SAHRAEI, M. A.; DUMAN, H.; ÇODUR, M. Y.; EYDURAN, E. Prediction of transportation energy demand: Multivariate Adaptive Regression Splines. **Energy**, Elsevier, v. 224, p. 120090, 2021. DOI: [10.1016/j.energy.2021.120090](https://doi.org/10.1016/j.energy.2021.120090). Citado 1 vez na página 49.

SANDMANN, W.; BOBER, O. **Stochastic Models for Intermittent Demands Forecasting and Stock Control**. Bamberg, Germany: University of Bamberg, Germany, 2010. Corresponding author: W. Sandmann, University of Bamberg, Department of Information Systems and Applied Computer Science, 96045 Bamberg, Feldkirchenstraße 21, Germany, werner.sandmann@uni-bamberg.de. Citado 1 vez na página 21.

SCHAER, O.; KOURENTZES, N.; FILDES, R. Demand forecasting with user-generated online information. en. **International Journal of Forecasting**, Elsevier, v. 35, p. 197–212, 2019. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0169207018300505>. Citado 1 vez na página 66.

SELIM, M.; ZHOU, R.; FENG, W.; ALAM, O. Studying Error Propagation for Energy Forecasting Using Univariate and Multivariate Machine Learning Algorithms. **IJCA**, ISCA, v. 27, n. 3, p. 1–9, 2020. Citado 2 vezes nas páginas 42, 56.

SHARMA, R.; KAMBLE, S. S.; GUNASEKARAN, A.; KUMAR, V.; KUMAR, A. A Systematic Literature Review on Machine Learning Applications for Sustainable Agriculture Supply Chain Performance. **Computers and Operations Research**, Elsevier, v. 119, p. 104926, 2020. DOI: [10.1016/j.cor.2020.104926](https://doi.org/10.1016/j.cor.2020.104926). Citado 1 vez na página 15.

SHEHZAD, M. K.; ROSE, L.; WESEMANN, S.; ASSAAD, M. ML-Based Massive MIMO Channel Prediction: Does It Work on Real-World Data? **IEEE Wireless Communications Letters**, IEEE, v. 11, n. 4, p. 811–815, abr. 2022. DOI: [10.1109/LWC.2022.3146230](https://doi.org/10.1109/LWC.2022.3146230). Citado 1 vez na página 54.

SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications: With R examples**. New York: Springer, 2010. Citado 1 vez na página 31.

SILVA, R. G. da; RIBEIRO, M. H. D. M.; LARCHER, J. H. K.; MARIANI, V. C.; COELHO, L. d. S. Artificial Intelligence and Signal Decomposition Approach Applied to Retail Sales Forecasting, 2020. Citado 1 vez na página 47.

SIMON, R. Supervised analysis when the number of candidate features (p) greatly exceeds the number of cases (n). **ACM SIGKDD Explorations Newsletter**, v. 5, n. 2, p. 31–36, 2003. Citado 1 vez na página 58.

SINA, L. B.; SECCO, C. A.; BLAZEVIC, M.; NAZEMI, K. Hybrid Forecasting Methods—A Systematic Review. **Electronics**, MDPI, Darmstadt, Germany, v. 12, n. 2019, 2023. Received: 17 March 2023; Revised: 12 April 2023; Accepted: 25 April 2023; Published: 27 April 2023. Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). DOI: [10.3390/electronics12092019](https://doi.org/10.3390/electronics12092019). Citado 1 vez na página 51.

SLUTSKY, E. The summation of random causes as the source of cyclic processes. en. **Econometrica**, Econometric Society, v. 5, n. 2, p. 105–146, abr. 1937. Citado 1 vez na página 18.

SOUSA, M.; TOMÉ, A. M.; MOREIRA, J. A General Framework for Multi-step Ahead Adaptive Conformal Heteroscedastic Time Series Forecasting. **Preprint submitted to Elsevier**, out. 2023. Disponível em: <https://github.com/Quilograma/AEnbMIMOCQR>. Citado 2 vezes nas páginas 36, 39.

SPILIOTIS, E.; MAKRIDAKIS, S.; SEMENOGLOU, A.-A.; ASSIMAKOPOULOS, V. Comparison of statistical and machine learning methods for daily SKU demand forecasting. **Operational Research**, p. 3, 2020. Original Paper. DOI: [10.1007/s12351-020-00605-2](https://doi.org/10.1007/s12351-020-00605-2). Disponível em: <https://doi.org/10.1007/s12351-020-00605-2>. Citado 1 vez na página 45.

STEIN, M.; BOOCHER, C. W. Search strategies for multicriterion optimization: An overview. **Mathematical Methods of Operations Research**, Springer, v. 61, n. 1, p. 141–173, 2005. Citado 1 vez na página 62.

SYNTETOS, A.; BOYLAN, J.; CROSTON, J. On the categorization of demand patterns. **Journal of the Operational Research Society**, v. 56, p. 495–503, 2005. Citado 4 vezes nas páginas 25–27.

TAGHIYEH, S.; LENGACHER, D. C.; SADEGHI, A. H.; SAHEBI-FAKHRABAD, A.; HANDFIELD, R. B. A novel multi-phase hierarchical forecasting approach with machine learning in supply chain management. **Supply Chain Analytics**, v. 3, p. 100032, 2023. DOI: [10.1016/j.sca.2023.100032](https://doi.org/10.1016/j.sca.2023.100032). Disponível em: <https://doi.org/10.1016/j.sca.2023.100032>. Citado 1 vez na página 60.

TAIEB, S. B.; SORJAMAA, A.; BONTEMPI, G. Multiple-output modeling for multi-step-ahead time series forecasting. **Neurocomputing**, Elsevier, v. 73, p. 1950–1957, jun. 2010. DOI: [10.1016/j.neucom.2009.11.030](https://doi.org/10.1016/j.neucom.2009.11.030). Citado 1 vez na página 56.

TASHMAN, L. J. Out-of-sample tests of forecasting accuracy: an analysis and review. **International Journal of Forecasting**, v. 16, n. 4, p. 437–450, 2000. Citado 1 vez na página 61.

TAYLOR, W. et al. Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. **Decision Support Systems**, Elsevier, v. 54, n. 1, p. 296–305, 2012. Citado 1 vez na página 48.

TEUNTER, R. H.; SYNTETOS, A. A.; BABAI, M. Z. Intermittent demand: linking forecasting to inventory obsolescence. **European Journal of Operational Research**, Elsevier, v. 214, n. 3, p. 606–615, 2011. Citado 2 vez na página 27.

TEUNTER, R. H.; SYNTETOS, A. A.; BABAI, Z. Intermittent demand: Linking forecasting to inventory obsolescence. **European Journal of Operational Research**, Elsevier, v. 214, n. 3, p. 606–615, 2011. Citado 1 vez na página 28.

TIAN, J.; OOKA, R.; LEE, D. Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review. **Journal of Cleaner Production**, v. 426, p. 139040, 2023. Available online 30 September 2023. © 2023 Elsevier Ltd. All rights reserved. ISSN 0959-6526. DOI: [DOI_here](#). Disponível em: [URL_here](#). Citado 1 vez na página 37.

TODOROV, I. B.; SÁNCHEZ LASHERAS, F. Forecasting Applied to the Electricity, Energy, Gas and Oil Industries: A Systematic Review. **Mathematics**, v. 10, n. 3930, 2022. DOI: [10.3390/math10213930](#). Disponível em: <https://doi.org/10.3390/math10213930>. Citado 1 vez na página 37.

TSAY, R. S. **Analysis of Financial Time Series**. Third. New York, NY: Wiley, 2013. The University of Chicago Booth School of Business, Chicago, IL. Citado 2 vezes nas páginas 23, 24.

TULI, S.; CASALE, G.; JENNINGS, N. R. TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data. **arXiv preprint arXiv:2201.07284v6**, mai. 2022. arXiv: [2201.07284v6 \[cs.LG\]](#). Citado 1 vez na página 71.

TÜRKMEN, A. C.; JANUSCHOWSKI, T.; WANG, Y.; CEMGIL, A. T. Forecasting intermittent and sparse time series: A unified probabilistic framework via deep renewal processes. Edição: Yossiri Adulyasak. **PLoS ONE**, Public Library of Science, Berlin, Germany; East Palo Alto, CA, United States of America; Istanbul, Turkey, v. 16, n. 11, e0259764, 2021. Received: September 16, 2020; Accepted: October 26, 2021; Published: November 29, 2021. Copyright: © 2021 Türkmen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License,

which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. DOI: [10.1371/journal.pone.0259764](https://doi.org/10.1371/journal.pone.0259764).

Citado 1 vez na página 21.

TWUMASI, C.; TWUMASI, J. Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana. **International Journal of Forecasting**, Elsevier, v. 38, p. 1258–1277, jun. 2022. DOI: [10.1016/j.ijforecast.2021.10.008](https://doi.org/10.1016/j.ijforecast.2021.10.008). Citado 1 vez na página 35.

VAN BELLE, J.; GUNS, T.; VERBEKE, W. Using shared sell-through data to forecast wholesaler demand in multi-echelon supply chains. **European Journal of Operational Research**, v. 288, p. 466–479, 2021. DOI: [10.1016/j.ejor.2020.06.003](https://doi.org/10.1016/j.ejor.2020.06.003). Disponível em: <https://doi.org/10.1016/j.ejor.2020.06.003>. Citado 1 vez na página 33.

VAN DER AUWERAER, S.; N. BOUTE, R.; A. SYNTETOS, A. Forecasting spare part demand with installed base information: A review. en. **International Journal of Forecasting**, Elsevier, n. 35, p. 181–196, mar. 2019. Citado 1 vez na página 19.

VANDERPUT, N. **Data Science for Supply Chain Forecasting**. 2. ed. [S.l.]: De Gruyter, 2017. Citado 1 vez na página 19.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. 1st. New York, NY: Springer-Verlag, 1995. ISBN 0-387-94559-8. Citado 1 vez na página 41.

VILLEGAS, M. A.; PEDREGAL, D. J. Automatic selection of unobserved components models for supply chain forecasting. en. **International Journal of Forecasting**, Elsevier, n. 35, p. 157–169, mar. 2021. Citado 1 vez na página 20.

VOLLMER, M. A.; GLAMPSON, B.; MELLAN, T.; MISHRA, S.; MERCURI, L.; COSTELLO, C.; KLABER, R.; COOKE, G.; FLAXMAN, S.; BHATT, S. A unified machine learning approach to time series forecasting applied to demand at emergency departments. en. **BMC Emergency Medicine**, v. 21, n. 9, 2021. DOI: [10.1186/s12873-020-00395-y](https://doi.org/10.1186/s12873-020-00395-y). Disponível em: <https://doi.org/10.1186/s12873-020-00395-y>. Citado 1 vez na página 40.

WALLSTRÖM, P.; SEGERSTEDT, A. Evaluation of forecasting error measurements and techniques for intermittent demand. **International Journal of Production Economics**, Elsevier, v. 128, n. 2, p. 625–636, 2010. Citado 1 vez na página 27.

WANG, G.; GUNASEKARAN, A.; NGAI, E. W.; PAPADOPOULOS, T. Big Data Analytics in Logistics and Supply Chain Management: Certain Investigations for Research and Applications. **International Journal of Production Economics**, v. 176, p. 98–110, 2016. Citado 1 vez na página 15.

WANG, G. Demand Forecasting of Supply Chain Based on Support Vector Regression Method. en. **Procedia Engineering**, Elsevier, School of Business Administration, South China University of Technology, Guangzhou 510641, China, 2011. Disponível em: www.elsevier.com/locate/procedia. Citado 1 vez na página 42.

WANG, J.; LU, S.; WANG, S.-H.; ZHANG, Y.-D. A review on extreme learning machine. **Multimedia Tools and Applications**, v. 81, p. 41611–41660, 2022. Received: 21 August 2020 / Revised: 26 February 2021 / Accepted: 5 May 2021 / Published online: 22 May 2021. DOI: [10.1007/s11042-021-11007-7](https://doi.org/10.1007/s11042-021-11007-7). Citado 1 vez na página 42.

WANG, Y.; ZHANG, Y.; QI, X.; WANG, M.; WANG, X. A Bi-Level Optimization Model for Virtual Power Plant Membership Selection Considering Load Time Series. **Sustainability**, MDPI, v. 15, p. 2138, 2023. Academic Editor: Sumedha Sharma. DOI: [10.3390/su15032138](https://doi.org/10.3390/su15032138). Disponível em: <https://doi.org/10.3390/su15032138>. Citado 1 vez na página 65.

WAPACHI, F. I.; DIAB, A. Time-series forecasting of a typical PWR system response under Control Element Assembly withdrawal at full power. **Nuclear Engineering and Design**, Elsevier, v. 413, p. 112472, 2023. DOI: [10.1016/j.nucengdes.2023.112472](https://doi.org/10.1016/j.nucengdes.2023.112472). Disponível em: <https://doi.org/10.1016/j.nucengdes.2023.112472>. Citado 1 vez na página 66.

WILLEMAIN, T. R.; SMART, C. N.; SCHWARZ, H. F. A new approach to forecasting intermittent demand for service parts inventories. **International Journal of Forecasting**, Elsevier, v. 20, n. 3, p. 375–387, 2004. Citado 1 vez na página 27.

WILLIAMS, T. Stock control with sporadic and slow-moving demand. **Journal of the Operational Research Society**, v. 35, p. 939–948, 1984. Citado 2 vez na página 25.

YANG, Y.; LIN, J.; LIU, G.; ZHOU, L. The Behavioural Causes of Bullwhip Effect in Supply Chains: A Systematic Literature Review. en. **International Journal of Production Economics**, Elsevier, v. 236, p. 108120, 2021. Citado 1 vez na página 29.

YIN, C.; BINH, L. T.; DUONG, T. A.; MAI, S. T.; LE, A.; NGUYEN, V.-H.; NGUYEN, V.-C.; XUAN TINH, N.; TANAKA, H.; TRUNG VIET, N.; NGUYEN, L. D.; DUONG, T. Q. Advanced Machine Learning Techniques for Predicting Nha Trang Shorelines. **IEEE Access**, IEEE, v. 9, p. 1258–1277, jun. 2021. DOI: [10.1109/ACCESS.2021.3095339](https://doi.org/10.1109/ACCESS.2021.3095339). Citado 1 vez na página 35.

YING, X. An Overview of Overfitting and its Solutions. **IOP Conf. Series: Journal of Physics: Conf. Series**, IOP Publishing, Building 1, Huizhong Tower, NO.1 Shangdi Seven Street, Haidian District, Beijing 100085, China, v. 1168, p. 022022, 2019. Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd. DOI: [10.1088/1742-6596/1168/2/022022](https://doi.org/10.1088/1742-6596/1168/2/022022). Citado 1 vez na página 39.

ZHANG, G. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, Elsevier, v. 50, p. 159–175, 2003. Citado 1 vez na página 35.

ZHANG, T.; ZHANG, Y.; CAO, W.; BIAN, J.; YI, X.; ZHENG, S.; LI, J. **Less Is More: Fast Multivariate Time Series Forecasting with Light Sampling-oriented MLP Structures**. [S.l.: s.n.], 2022. arXiv:2207.01186 [cs.LG]. DOI: [10.48550/arXiv.2207.01186](https://doi.org/10.48550/arXiv.2207.01186). arXiv: [2207.01186](https://arxiv.org/abs/2207.01186) [cs.LG]. Disponível em: <https://doi.org/10.48550/arXiv.2207.01186>. Citado 1 vez na página 39.

ZHANG, X.; LIU, C.-A. Model averaging prediction by K-fold cross-validation. **Journal of Econometrics**, v. 235, p. 280–301, 2023. DOI: [10.1016/j.jeconom.2022.04.004](https://doi.org/10.1016/j.jeconom.2022.04.004). Disponível em: www.elsevier.com/locate/jeconom. Citado 1 vez na página 59.

ZHUANG, X.; YU, Y.; CHEN, A. A combined forecasting method for intermittent demand using the automotive aftermarket data. **Data Science and Management**, KeAi Publishing, China, v. 5, p. 43–56, 2022. Contents lists available at ScienceDirect Data Science and Management journal homepage: www.keaipublishing.com/en/journals/data-science-and-management. DOI: [10.1016/j.dsm.2022.XXXX.XXX](https://doi.org/10.1016/j.dsm.2022.XXXX.XXX). Citado 1 vez na página 28.

ZOU, J.; ZHAO, Q.; JIAO, Y.; CAO, H.; LIU, Y.; YAN, Q.; ABBASNEJAD, E.; LIU, L.; SHI, J. Q. Stock Market Prediction via Deep Learning Techniques: A Survey. **arXiv preprint arXiv:YYYY.YYYY**, p. 34, fev. 2023. CCS Concepts: • Applied computing → Economics; • Computing methodologies → Artificial intelligence; Machine learning.

Additional Key Words and Phrases: Deep learning, Machine learning, Finance, AI in finance, Stock market prediction. Citado 1 vez na página 71.