UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

DOUGLAS LUZ DA SILVEIRA

MODELAGEM DA PD FORWARD LOOKING PARA UMA INSTITUIÇÃO FINANCEIRA BRASILEIRA

PORTO ALEGRE
2025

DOUGLAS LUZ DA SILVEIRA

MODELAGEM DA PD FORWARD LOOKING PARA UMA INSTITUIÇÃO FINANCEIRA BRASILEIRA

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia. Área de Concentração: Economia.

Orientador: Prof. Dr. Carlos Eduardo Schonerwald da Silva

Coorientadora: Profa. Dra. Letícia de Oliveira

PORTO ALEGRE

2025

CIP - Catalogação na Publicação

Silveira, Douglas Luz da
Modelagem da PD Forward Looking para uma instituição financeira brasileira / Douglas Luz da Silveira. -- 2025.
94 f.
Orientador: Carlos Eduardo Schonerwald da Silva.

Coorientadora: Letícia de Oliveira.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2025.

1. Perdas esperadas de crédito. 2. SARIMAX. 3. Provisões bancárias. 4. Variáveis macroeconômicas. 5. IFRS 9. I. Silva, Carlos Eduardo Schonerwald da, orient. II. Oliveira, Letícia de, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

DOUGLAS LUZ DA SILVEIRA

MODELAGEM DA PD FORWARD LOOKING PARA UMA INSTITUIÇÃO FINANCEIRA BRASILEIRA

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia. Área de Concentração: Economia.

Aprovada em: Porto Alegre, _____de _____de 2025.

BANCA EXAMINADORA:

Prof. Dr. Carlos Eduardo Schonerwald da Silva – Orientador

UFRGS

Profa. Dra. Letícia de Oliveira – Coorientadora

UFRGS

Prof. Dr. Alessandro Kahmann

UFRGS

Prof. Dr. Cristiano Lima Hackmann

UFRGS

Dedico esta conquista àqueles que me motivam e fazem parte do que venho plantando.

À minha esposa Jaqueline, aos nossos filhos Alice Janete e Heloísa, bem como à minha família, incluindo os que aqui não mais estão...

AGRADECIMENTOS

Agradeço especialmente a minha mãe, que, apesar de não estar mais presente fisicamente desde 2021, vítima da COVID-19, continua iluminando e guiando todos os meus passos com seu amor eterno.

Às minhas filhas, Alice Janete e Heloísa, que tornam os meus dias mais alegres, trazendo cor, energia e inspiração para seguir em frente em cada momento da minha jornada.

Ao meu pai, cuja dedicação incansável à minha educação foi inestimável, alfabetizando-me antes mesmo de eu entrar na escola e plantando em mim a semente do conhecimento e da curiosidade.

Minha gratidão também ao Banco do Estado do Rio Grande do Sul, que transformou minha vida ao proporcionar suporte e oportunidades fundamentais para a realização desta pesquisa.

Por fim, mas não menos importante, agradeço à minha esposa, cujo apoio incondicional e paciência foram essenciais durante todos os momentos de ausência e dedicação a este trabalho.

Muito obrigado a todos vocês que, direta ou indiretamente, fizeram parte desta caminhada.

RESUMO

A norma internacional IFRS 9 introduziu um modelo de perdas esperadas que visa aprimorar o cálculo de provisões bancárias ao considerar tanto as perdas já ocorridas quanto as potenciais perdas futuras. Esse modelo incentiva a realização de estudos sobre o impacto de variáveis macroeconômicas nos modelos utilizados para calcular as provisões bancárias. A identificação dessas variáveis, com relevância estatística para as alterações históricas das carteiras de crédito, busca capacitar os bancos a incorporá-las, aprimorando as estimativas de probabilidade de *default* (PD). Este trabalho investiga a relação entre fatores macroeconômicos e a probabilidade de *default* na carteira de crédito do Banrisul (Banco do Estado do Rio Grande do Sul) à luz da IFRS 9, com o intuito de fornecer informações que apoiem as instituições financeiras na adequação às exigências da norma. O objetivo principal é explorar métodos que avaliem essa interação e demonstrar como as oscilações macroeconômicas podem ser integradas às probabilidades de *default*. Os fatores analisados incluem o PIB, o IPCA, a taxa de desemprego, a taxa de câmbio e a taxa Selic. Os dados de *default* são baseados na carteira de crédito de pessoa física e pessoa jurídica do Banrisul, disponibilizados por base interna, e consideram atrasos superiores ou iguais a 90 dias no período de janeiro de 2018 a dezembro de 2023. Foram empregados modelos como SARIMA, SARIMAX, VAR, Regressão Linear e XGBoost para examinar a relação entre as variáveis macroeconômicas e a probabilidade de *default*. Os resultados apontam IPCA, taxa de desemprego e câmbio como fatores macroeconômicos mais significativos na explicação das variações na probabilidade de *default*, embora a própria série e suas defasagens tenham apresentado um poder preditivo considerável, contribuindo diretamente para o cálculo das perdas esperadas de crédito. Ao final, o modelo SARIMAX demonstrou equilíbrio entre precisão, estabilidade e interpretabilidade, destacando-se como a abordagem mais adequada para embasar decisões estratégicas no contexto da gestão de risco de crédito. A pesquisa também abre caminho para estudos futuros, que poderão ampliar o conjunto de variáveis explicativas e explorar novas técnicas de modelagem para fortalecer ainda mais o arcabouço preditivo conforme a

complexidade das dinâmicas econômicas evolui.

Palavras-chave: Perdas esperadas de crédito. IFRS 9. Probabilidade de descumprimento. Variáveis macroeconômicas. ARIMAX. SARIMAX. Regressão linear. VAR. XGBoost. Banrisul. Provisões bancárias.

ABSTRACT

The international standard IFRS 9 introduced an expected loss model aimed at improving the calculation of banking provisions by considering both incurred losses and potential future losses. This model encourages studies on the impact of macroeconomic variables on models used to calculate banking provisions. Identifying these variables, with statistical relevance to historical changes in credit portfolios, seeks to enable banks to incorporate them, enhancing estimates of probability of *default* (PD). This study investigates the relationship between macroeconomic factors and the probability of *default* in Banrisul's (Banco do Estado do Rio Grande do Sul) credit portfolio under IFRS 9, aiming to provide information that supports the financial institution in complying with the standard's requirements. The main objective is to explore methods that assess this interaction and demonstrate how macroeconomic fluctuations can be integrated into *default* probabilities. The factors analyzed include GDP, IPCA (Consumer Price Index), unemployment rate, exchange rate, and Selic rate. *Default* data are based on Banrisul's personal and corporate credit portfolios, provided by internal databases, and consider delays of 90 days or more during the period from January 2018 to December 2023. Models such as SARIMA, SARIMAX, VAR, Linear Regression, and XGBoost were employed to examine the relationship between macroeconomic variables and the probability of *default*. The results indicate that the IPCA (consumer price index), unemployment rate, and exchange rate are the most significant macroeconomic factors in explaining variations in the probability of *default*, although the *default* series itself and its lags also showed considerable predictive power, contributing directly to the calculation of expected credit losses. In the end, the SARIMAX model demonstrated a balance of accuracy, stability, and interpretability, standing out as the most suitable approach to support strategic decision-making in the context of credit risk management. This research also paves the way for future studies, which may expand the set of explanatory variables and explore new modeling techniques to further strengthen the predictive framework as the complexity of economic dynamics continues to evolve.

Keywords: Expected credit losses. IFRS 9. Probability of *default*. Macroeconomic variables. ARIMAX. SARIMAX. Linear Regression. VAR. XGBoost. Banrisul. Credit loss provisions.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ARIMAX	AutoRegressive Integrated Moving Average with Exogenous Variables
BCB	Banco Central do Brasil
CDS	Credit *Default* Swaps
EAD	Exposure At *Default* (Exposição em *Default*)
ECL	Expected Credit Loss (Perda de Crédito Esperada)
IAS	International Accounting Standards
IASB	International Accounting Standards Board
IFRS	International Financial Reporting Standards
IPEA	Instituto de Pesquisa Econômica Aplicada
IPCA	Índice de Preços ao Consumidor Amplo
ECL	Expected Credit Loss
LGD	Loss Given *Default* (Perda Dado o *Default*)
PD	Probality of *Default* (Probabilidade de *Default*)
PD LT	Probality of *Default* Lifetime (Probabilidade de *Default* Vida)
PIB	Produto Interno Bruto
PIT	Ponto no Tempo
RMSE	Raiz do Erro Quadrático Médio
TTC	Através do Ciclo
VAR	Vetorial AutoRegressivo
VECM	Modelo de Correção de Erros Vetoriais
VIF	Variance Inflation Factor

SUMÁRIO

1 INTRODUÇÃO

O tema central abordado será o desenvolvimento de metodologias para o cálculo da perda esperada de crédito conforme a norma IFRS 9, com particular atenção à integração de informações macroeconômicas e macroprudenciais. Este assunto é extremamente pertinente no cenário econômico atual, marcado pela volatilidade dos mercados financeiros globais e por transformações regulatórias que demandam uma gestão de risco de crédito mais precisa e proativa.

No Brasil, o Banco Central publicou a Resolução nº 2.682, em 21 de dezembro de 1999, que introduz critérios para a classificação das operações de crédito e para a constituição de provisões relacionadas a créditos de liquidação duvidosa. Essa norma busca simplificar a classificação de riscos bancários, estabelecendo percentuais de provisão de acordo com o nível de risco identificado, o que facilita o monitoramento e a comparabilidade entre instituições financeiras nacionais.

Com o avanço das normas internacionais, o Conselho de Normas Internacionais de Contabilidade (IASB) publicou, em 24 de julho de 2014, a versão final da norma IFRS 9, que redefine a contabilização de instrumentos financeiros. Desenvolvida em resposta à crise financeira de 2008, a IFRS 9 trouxe um modelo de perdas esperadas, o qual antecipa o reconhecimento de provisões, incorporando fatores prospectivos e modelos estatísticos. A norma exige que tanto perdas já incorridas quanto aquelas projetadas sejam adequadamente provisionadas.

A norma IFRS 9, implementada no Brasil em janeiro de 2018, trouxe mudanças significativas em relação à sua antecessora, a IAS 39, especialmente ao introduzir o modelo de perda de crédito esperada para toda a vida útil dos instrumentos financeiros. Esta abordagem busca uma representação mais realista e antecipatória do risco de crédito, que responde melhor às condições econômicas e diminui a prociclicidade no sistema financeiro.

No Brasil, a Resolução nº 4.966/21 do Banco Central do Brasil reforçou a necessidade de aprimorar os métodos de avaliação de risco e a estimativa de perdas esperadas. Essa evolução metodológica busca alinhar o setor financeiro nacional às

melhores práticas de mercado, especialmente no que diz respeito ao cálculo de provisões.

A importância deste tema é enfatizada pela instabilidade econômica global recente, onde a precisão na estimativa de perdas de crédito pode influenciar diretamente a estabilidade das instituições financeiras e, consequentemente, das economias em geral. A integração de variáveis macroeconômicas e macroprudenciais nas estimativas de ECL permite que bancos e outras instituições financeiras ajustem suas provisões para perdas de crédito de maneira mais eficaz diante de flutuações econômicas, o que é vital em períodos de incerteza.

Além de cumprir com os requisitos regulatórios, as metodologias focadas no IFRS 9 oferecem às instituições financeiras ferramentas aprimoradas para gestão de risco, promovem maior transparência para investidores e reguladores, e contribuem para a estabilidade financeira. Esse aspecto se torna crucial em contextos de crises financeiras, onde modelos preditivos inadequados podem exacerbar os problemas econômicos. Portanto, a discussão e desenvolvimento dessas metodologias não só atendem a uma exigência regulatória, como também representam uma necessidade prática para enfrentar os desafios financeiros contemporâneos, conferindo ao tema uma relevância significativa e atual.

Um dos desafios mais significativos nesse contexto é a estimativa da probabilidade de *default* com perspectiva futura, conhecida como *PD Forward Looking*. Tradicionalmente, modelos de cálculo de PD não incluem variáveis macroeconômicas, pois estas não estão diretamente ligadas às características individuais dos clientes. Contudo, a IFRS 9 introduziu a ideia de integrar fatores macroeconômicos nos modelos de estimativa de PD, promovendo análises mais abrangentes e alinhadas a cenários prospectivos.

Diante disso, o objetivo deste trabalho é investigar a relação entre variáveis macroeconômicas e a probabilidade de *default* (PD) da carteira de crédito do Banrisul, à luz da norma IFRS 9. Pretende-se desenvolver, testar e comparar diferentes metodologias que incorporam fatores econômicos ao cálculo da PD Forward Looking, avaliando sua capacidade preditiva, estabilidade e aplicabilidade prática.

A relevância desta pesquisa está na contribuição ao aperfeiçoamento dos

modelos de risco de crédito utilizados pelas instituições financeiras, especialmente no contexto brasileiro, onde há crescente demanda por abordagens mais dinâmicas e alinhadas a padrões internacionais.

O Banco do Estado do Rio Grande do Sul é uma das maiores instituições financeiras do Brasil, com uma trajetória sólida e reconhecida por seu papel estratégico no desenvolvimento econômico do estado do Rio Grande do Sul. Fundado em 1928, o Banrisul atua como um banco múltiplo, oferecendo uma ampla gama de produtos e serviços financeiros para pessoas físicas, empresas e setores governamentais. Sua relevância regional é evidenciada por sua capilaridade, com uma extensa rede de agências e correspondentes que atendem tanto áreas urbanas quanto rurais, além de sua forte atuação no mercado de crédito, especialmente em operações destinadas ao agronegócio e ao crédito pessoal.

A estrutura deste trabalho está organizada da seguinte forma: inicialmente, a Revisão de Literatura apresenta estudos anteriores sobre modelagem de risco de crédito com fatores macroeconômicos, além de fundamentos teóricos e regulatórios relevantes. Em seguida, a Fundamentação Teórica define os principais conceitos relacionados à taxa de *default* e detalha as metodologias empregadas. A Metodologia descreve o processo de modelagem, as ferramentas utilizadas e os efeitos esperados das variáveis macroeconômicas sobre a PD. A Análise dos Resultados compara os quatro modelos desenvolvidos, avaliando consistência dos coeficientes, multicolinearidade, resíduos e predições. A Simulação de Impacto investiga a estabilidade dos modelos frente a variações nas variáveis explicativas. Por fim, nas Considerações Finais, destaca o modelo com melhor desempenho, discute as limitações encontradas e propõe caminhos para pesquisas futuras.

2 REVISÃO DE LITERATURA

No contexto da evolução das normas contábeis e financeiras, especialmente com a implementação da IFRS 9, a revisão da literatura assume papel fundamental ao abordar o desenvolvimento de metodologias para o cálculo da PD com abordagem Forward Looking. Essa revisão não apenas traça a trajetória histórica e as práticas atualmente adotadas, mas também analisa como essas metodologias têm sido ajustadas às exigências regulatórias e às crescentes demandas do mercado financeiro global por maior precisão e responsividade.

Um aspecto particularmente relevante é a integração de variáveis macroeconômicas e macroprudenciais nas estimativas de PD. A literatura destaca como essas variáveis são incorporadas aos modelos preditivos, com o objetivo de refletir de forma mais acurada as condições econômicas vigentes. Essa abordagem busca garantir que as estimativas de risco de crédito respondam de maneira mais sensível e tempestiva às oscilações do ambiente macroeconômico.

Além disso, um ponto crítico amplamente explorado é a relação entre as condições macroeconômicas e as perdas de crédito esperadas (Expected Credit Losses – ECL). Diversos estudos evidenciam o impacto de diferentes cenários econômicos sobre as perdas projetadas, oferecendo subsídios valiosos para o ajuste dinâmico dos modelos de previsão. Esses trabalhos proporcionam insights sobre como calibrar os modelos de PD para refletir de forma adequada as mudanças no ciclo econômico, aumentando a eficácia das estratégias de gestão de risco e conformidade regulatória.

2.1 A NORMA IFRS 9 E SEUS OBJETIVOS

A norma IFRS 9 foi desenvolvida pelo International Accounting Standards Board (IASB) com o objetivo de superar as limitações do modelo contábil anterior, o IAS 39, cujas fragilidades tornaram-se especialmente evidentes durante a crise financeira global de 2008. Nesse cenário, a IFRS 9 representa um marco regulatório voltado à melhoria da qualidade das informações contábeis, promovendo maior transparência, relevância e tempestividade na mensuração e divulgação de

instrumentos financeiros.

Entre suas inovações centrais, destaca-se a introdução do modelo de perdas esperadas em substituição ao modelo de perdas incorridas anteriormente adotado. Essa mudança busca antecipar potenciais impactos financeiros, proporcionando maior previsibilidade e alinhamento entre as demonstrações financeiras e a realidade econômica enfrentada pelas entidades. Ademais, a norma simplifica os critérios de classificação e mensuração de ativos financeiros, incorporando uma abordagem baseada no modelo de negócios da entidade e nas características contratuais dos fluxos de caixa dos instrumentos financeiros.

Outro avanço significativo introduzido pela IFRS 9 refere-se à reformulação do hedge accounting, aproximando as práticas contábeis da efetiva gestão de riscos corporativos. Essa adequação visa garantir uma representação mais fiel das estratégias de proteção financeira adotadas pelas empresas, reforçando a utilidade das informações contábeis para os stakeholders.

Conforme ressalta o IASB (2014), a implementação da IFRS 9 reflete um esforço de alinhamento da contabilidade às práticas de mercado, com vistas a assegurar que as demonstrações financeiras ofereçam informações úteis, transparentes e relevantes aos usuários, fortalecendo a tomada de decisão e a confiança nos mercados financeiros.

2.2 ECL: ELEMENTOS DO CÁLCULO DE PERDA ESPERADA

Conforme a IFRS Foundation (2014), o cálculo da Expected Credit Loss (ECL) – perda esperada de crédito – é baseado em três componentes fundamentais:

PD (Probability of *Default*): Probabilidade de *default* do tomador em um determinado horizonte temporal.

LGD (Loss Given *Default*): Percentual da exposição que se espera perder no caso de *default*.

EAD (Exposure at *Default*): Valor da exposição ao risco no momento do *default*.

De acordo com a IFRS 9, os instrumentos financeiros são classificados em três estágios distintos, conforme a evolução do risco de crédito:

Estágio 1: Instrumentos que não apresentam deterioração significativa no risco de crédito desde a sua originação. Nessa fase, a estimativa de ECL considera a PD para os próximos 12 meses.

Estágio 2: Instrumentos com deterioração significativa no risco de crédito, ainda que não inadimplentes. O cálculo do ECL deve considerar a PD ao longo da vida total do instrumento.

Estágio 3: Instrumentos inadimplentes, para os quais o ECL também deve ser estimado considerando a PD para toda a vida do ativo financeiro.

Adicionalmente, a norma requer a incorporação de cenários macroeconômicos esperados no cálculo da perda esperada de crédito. Esses ajustes visam antecipar variações nas condições econômicas e mitigar os efeitos de choques sistêmicos, garantindo maior previsibilidade nas estimativas de perdas e na gestão de risco de crédito.

2.3 METODOLOGIAS PARA PD FORWARD LOOKING

Diversos estudos têm contribuído para o avanço das metodologias de estimativa da Probability of *Default* (PD), especialmente no contexto das exigências da IFRS 9, que demanda modelos mais dinâmicos e sensíveis às condições econômicas futuras.

Kauffmann (2017) combinou modelos de regressão logística com ARIMAX para integrar dados históricos e projeções econômicas, ajustando as estimativas de PD às tendências previstas.

Breeden e Crook (2022) desenvolveram um modelo de sobrevivência em tempo discreto, alinhado aos padrões da IFRS 9, que incorpora fatores como a idade do empréstimo, condições temporais e características do tomador. No entanto, a principal limitação do modelo é a sua dependência de padrões históricos, que podem não capturar adequadamente futuras mudanças no ambiente econômico.

Xu (2016) propôs a integração entre análise de sobrevivência e regressão bayesiana para desenvolver modelos mais dinâmicos e adaptáveis às flutuações econômicas. A abordagem utiliza indicadores macroeconômicos, como PIB e taxa de desemprego, embora sua eficácia dependa da disponibilidade de dados atualizados e

precisos.

Bednarek e Franke (2024) exploraram as dinâmicas das PDs reportadas trimestralmente por instituições financeiras alemãs, aplicando modelos de Cadeias de Markov para prever transições entre classificações de crédito ao longo do tempo.

Outras contribuições relevantes incluem Miao *et al.* (2018), que propuseram uma variação do modelo de distância ao *default* de Merton, incorporando volatilidade implícita e custo de capital implícito para prever inadimplência. Dainelli, Bet e Fabrizi (2024) desenvolveram um modelo baseado na credibilidade das projeções financeiras corporativas para estimar PD. Georgiou e Yannacopoulos (2023) apresentaram um modelo estocástico alinhado às exigências da IFRS 9, enquanto Grigutis (2023) forneceu uma abordagem probabilística específica para portfólios com baixa incidência de *default*s.

Kaushansky, Lipton e Reisinger (2018) derivaram fórmulas semi-analíticas para modelar o movimento do risco de crédito, contribuindo para o entendimento matemático do comportamento das PDs. Yang (2017) propôs um modelo baseado na probabilidade de sobrevivência futura, que integra variáveis macroeconômicas e pontuações de crédito. Já Jarrow e Turnbull (1995) introduziram um modelo estatístico de risco de crédito que permite estimativas de PD sem a dependência direta da estrutura de capital.

Por fim, Altman (1968) desenvolveu o famoso Z-score, um modelo baseado em indicadores financeiros que antecipa a falência de empresas em até dois anos, sendo amplamente utilizado em análises de risco de crédito corporativo.

Esses estudos evidenciam a importância de incorporar variáveis macroeconômicas e indicadores prospectivos na modelagem da PD. Apesar das limitações inerentes a cada abordagem, essas metodologias oferecem ferramentas valiosas para a gestão do risco de crédito e para o cumprimento das exigências regulatórias da IFRS 9.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, será apresentado o conceito de taxa de *default* adotado neste estudo, bem como o detalhamento da aplicação dos modelos utilizados para sua análise e previsão: regressão linear múltipla, ARIMAX, VAR (Vetores Auto-Regressivos) e regressão XGBoost.

3.1 DEFINIÇÃO DA TAXA DE INADIMPLÊNCIA

Neste estudo, a taxa de *default* será definida como o percentual das operações de crédito concedidas a pessoas físicas e jurídicas no Banrisul que apresentam atraso igual ou superior a 90 dias. Essa métrica será utilizada como indicador de inadimplência e será calculada conforme a seguinte fórmula:

$$\mathrm{Tx}_{inad_i} = \frac{\mathrm{Qtd}_{contratos\_atraso\_90_i}}{\mathrm{Qtd}_{contratos\_total_i}} \quad (1)$$

Onde:

Tx_{inad_i}: Representa a taxa de inadimplência no mês i;
$\mathrm{Qtd}_{contratos\_atraso\_90_i}$: Corresponde à quantidade de contratos com atrasos superiores a 90 dias no mês i;
$\mathrm{Qtd}_{contratos\_total_i}$: Refere-se ao total de contratos existentes no mês i.

A adoção dessa métrica segue a definição padrão amplamente utilizada em estudos relacionados ao crédito bancário, o que assegura comparabilidade e consistência metodológica. Além disso, permite a análise da taxa de *default* com base na estrutura temporal das operações de crédito, possibilitando a modelagem de sua dinâmica ao longo do tempo em resposta a variáveis econômicas e financeiras.

3.2 REGRESSÃO LINEAR MÚLTIPLA

A análise de regressão é amplamente utilizada tanto para a geração de previsões quanto para a verificação de hipóteses econômicas que envolvem a relação entre uma variável dependente e uma ou mais variáveis independentes. A escolha pela utilização da regressão neste estudo justifica-se pela facilidade de interpretação dos parâmetros estimados, o que possibilita uma compreensão mais clara dos efeitos de cada variável explicativa sobre a variável de interesse. Adicionalmente, os coeficientes obtidos permitem projetar probabilidades futuras de *default* com base nas estimativas das variáveis explicativas.

De forma geral, o modelo de regressão linear múltipla pode ser representado pela seguinte equação:

$$Y_i = \sum_{j=0}^{k} x_{ij}\beta_j + \varepsilon_i, \quad i = 1, \dots, n \quad (2)$$

Onde:

Y: Variável dependente que está associada às covariáveis estocásticas;

β_j: Coeficientes que indicam o peso de cada covariável estocástica na variável dependente;

x_{ij}: Covariáveis estocásticas que explicam a variável dependente;

ε_i: Termo residual que representa o erro do modelo;

n: Número de observações;

k: Número de covariáveis.

Neste trabalho, será aplicado o modelo de regressão linear múltipla, considerando como variável dependente a taxa de *default* e, como variáveis explicativas, indicadores macroeconômicos selecionados. Essa abordagem permite investigar o impacto dessas variáveis sobre a inadimplência, bem como realizar previsões da taxa de *default* com base em projeções econômicas.

Embora amplamente utilizada em estudos econômicos e financeiros, a regressão linear múltipla apresenta limitações importantes quando aplicada à previsão de *PD Forward Looking*. Uma das principais restrições reside na suposição de linearidade entre a variável dependente e as explicativas. No contexto de dados financeiros e macroeconômicos, as relações frequentemente assumem formas não lineares, o que pode limitar a capacidade do modelo de capturar adequadamente a dinâmica subjacente (Gujarati; Porter, 2011).

Outro desafio relevante é a presença de multicolinearidade entre as variáveis explicativas — um fenômeno comum entre indicadores macroeconômicos — que dificulta a interpretação dos coeficientes estimados e compromete a precisão das previsões. A correlação elevada entre variáveis pode provocar instabilidade nos parâmetros, reduzindo a confiabilidade dos resultados (Wooldridge, 2020).

Adicionalmente, a regressão linear é sensível a outliers. Em séries financeiras, onde choques econômicos são frequentes, valores extremos podem distorcer significativamente as estimativas, enviesando os resultados e comprometendo a assertividade do modelo (Montgomery; Peck; Vining, 2012).

Além disso, o modelo assume que os resíduos seguem distribuição normal e apresentam variância constante (homocedasticidade). Contudo, séries temporais financeiras costumam violar essas premissas, exibindo heterocedasticidade, além de padrões sazonais ou cíclicos que não são capturados adequadamente pela regressão linear convencional (Greene, 2018).

Outro ponto crítico é a limitação do modelo em capturar a dinâmica temporal de forma eficiente. Apesar da possibilidade de inclusão de defasagens como variáveis explicativas, a regressão linear não é projetada para modelar interdependências temporais de maneira estruturada, como ocorre em modelos específicos de séries temporais, a exemplo do ARIMA ou do VAR (Stock; Watson, 2019).

Por fim, a qualidade dos dados exerce influência significativa sobre a eficácia da regressão linear múltipla. Erros, omissões ou mudanças estruturais nas variáveis explicativas — como alterações em políticas econômicas ou choques exógenos — podem comprometer a estabilidade e a validade das estimativas geradas (Hamilton, 1994).

3.3 MODELO ARIMAX

Os modelos ARIMA (AutoRegressive Integrated Moving Average) são amplamente empregados na análise de séries temporais, especialmente em contextos nos quais a variável de interesse depende unicamente de seu comportamento passado e da estrutura temporal dos dados. Um exemplo clássico de aplicação consiste na previsão de preços de ativos financeiros, como ações, cuja dinâmica é fortemente influenciada por valores históricos.

O modelo ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) representa uma extensão do ARIMA tradicional, ao incorporar variáveis exógenas como elementos explicativos adicionais. Essa característica permite que o modelo considere não apenas a evolução histórica da variável dependente, mas também o impacto de fatores externos que influenciam sua dinâmica. Como resultado, a abordagem ARIMAX tende a apresentar desempenho superior em relação ao ARIMA puro, especialmente em ambientes onde variáveis macroeconômicas ou setoriais exercem influência significativa sobre a série temporal analisada.

A formulação matemática do modelo ARIMAX é expressa da seguinte forma:

$$y_i = \beta_1 X_i + \sum_{j=1}^{p} \phi_j y_{i-j} + \varepsilon_i + \sum_{j=1}^{q} \theta_j \varepsilon_{i-j} \quad (3)$$

Onde:

y_i: Variável dependente, que neste estudo corresponde à taxa de *default* no i-ésimo mês;

X_i: Variáveis exógenas, como Produto Interno Bruto (PIB) ou inflação, no i-ésimo mês;

β_1 : Coeficiente associado às variáveis exógenas;

ϕ_j: Coeficiente autoregressivo para a j-ésima defasagem;

θ_j: Coeficiente de média móvel para a j-ésima defasagem;

ε_i: Termo de erro associado ao modelo de média móvel no i-ésimo mês.

As variáveis exógenas utilizadas em modelos ARIMAX podem abranger uma ampla gama de indicadores econômicos, como taxas de inflação, índices de preços,

variáveis categóricas que distinguem dias da semana, entre outros fatores externos. A principal vantagem do ARIMAX reside justamente em sua flexibilidade, permitindo a incorporação de qualquer variável externa considerada relevante para influenciar a dinâmica da taxa de *default*.

No entanto, uma limitação importante do modelo é a suposição de linearidade na relação entre a variável dependente e as variáveis exógenas. Embora essa hipótese seja válida em muitos contextos, ela pode ser insuficiente para capturar padrões complexos e não lineares, comumente observados em dados financeiros e macroeconômicos. Isso pode resultar em modelagens incompletas e em previsões com menor precisão (Greene, 2018).

Além disso, o modelo ARIMAX enfrenta dificuldades significativas quando há multicolinearidade entre as variáveis exógenas — uma condição frequente em séries macroeconômicas, nas quais variáveis como inflação, PIB e taxa de juros costumam apresentar altas correlações. Esse cenário pode comprometer a estabilidade dos coeficientes estimados e dificultar a interpretação dos resultados (Gujarati; Porter, 2011).

Outro aspecto crítico é a sensibilidade do modelo à especificação incorreta das ordens autorregressiva (p), de diferenciação (d) e de média móvel (q). A definição inadequada desses parâmetros pode levar a um modelo mal ajustado, incapaz de capturar a dinâmica real da série temporal. Tal desafio se intensifica no contexto de *PD Forward Looking*, em que choques econômicos podem gerar comportamentos não usuais, dificultando a correta parametrização do modelo (Hamilton, 1994).

Adicionalmente, o ARIMAX assume que os resíduos do modelo se comportam como *white noise*, ou seja, que são não correlacionados, com média zero e variância constante. Contudo, séries temporais financeiras frequentemente violam essa premissa, apresentando heterocedasticidade e autocorrelação nos resíduos, o que pode comprometer a validade das inferências e reduzir a eficácia preditiva do modelo (Montgomery; Peck; Vining, 2012).

Por fim, o modelo pode apresentar limitações no tratamento de bases de dados extensas ou com elevado número de variáveis exógenas. À medida que a dimensionalidade das variáveis aumenta, cresce também a complexidade

computacional, exigindo técnicas rigorosas de seleção de variáveis e ajustes finos para evitar problemas como overfitting e a consequente perda de capacidade preditiva (Stock; Watson, 2019).

3.4 MODELO VAR (VETOR AUTO-REGRESSIVO)

O modelo Vetor Auto-Regressivo é uma extensão dos modelos autorregressivos univariados, amplamente utilizado na análise de séries temporais multivariadas. Sua principal característica é a capacidade de analisar e prever o comportamento conjunto de múltiplas variáveis interdependentes ao longo do tempo. Ao contrário dos modelos univariados, o VAR trata todas as variáveis do sistema como potencialmente endógenas, permitindo que cada variável seja explicada não apenas por suas próprias defasagens, mas também pelas defasagens das demais variáveis incluídas no modelo.

Essa abordagem permite capturar de forma mais abrangente a dinâmica temporal e as inter-relações entre variáveis econômicas, sendo particularmente útil em contextos onde os efeitos de retroalimentação e causalidade mútua são relevantes, como na análise de risco de crédito, política monetária e ciclos econômicos.

O modelo VAR de ordem p pode ser representado pela seguinte equação matricial:

$$\mathrm{Y_t} = \mathrm{A_1 Y_{t-1}} + A_2 \mathrm{Y_{t-2}} + \cdots + \mathrm{A_p Y_{t-p}} + \varepsilon_\mathrm{t} \quad (4)$$

Onde:

Y_t: Vetor de variáveis endógenas no período t;

A_1, A_2, A_p: Matrizes de coeficientes que relacionam as variáveis endógenas e suas defasagens;

ε_t: Vetor de termos de erro, que são assumidos como ruído branco (white noise), sem correlação serial e com média zero.

No contexto deste estudo, o modelo Vetor Auto-Regressivo (VAR) será empregado para analisar a interação dinâmica entre a taxa de *default* e variáveis macroeconômicas selecionadas. Essa abordagem é particularmente eficaz para

capturar relações bidirecionais, permitindo avaliar tanto o impacto de choques macroeconômicos sobre a taxa de *default* quanto os efeitos de retroalimentação da taxa de *default* sobre o ambiente macroeconômico.

Uma das principais vantagens do VAR é que ele não impõe, a priori, relações causais específicas entre as variáveis. Em vez disso, permite que os dados revelem as interações predominantes no sistema, proporcionando uma análise empírica mais flexível. Ademais, o modelo possibilita investigações aprofundadas por meio da decomposição da variância dos erros de previsão e da análise de funções impulso-resposta, que medem o efeito de choques exógenos em cada variável ao longo do tempo.

Apesar de sua utilidade, o modelo VAR apresenta algumas limitações importantes. Uma delas é sua elevada sensibilidade à dimensionalidade dos dados. À medida que aumenta o número de variáveis e de defasagens consideradas, cresce também o número de parâmetros a serem estimados, o que pode levar ao overfitting e comprometer a capacidade de generalização do modelo, especialmente em amostras pequenas ou moderadas (Lütkepohl, 2005).

Outro desafio está na ausência de relações causais explícitas. Embora o VAR seja eficiente para capturar correlações e interdependências temporais, ele não permite inferências diretas sobre causalidade econômica. Por isso, interpretações mais completas exigem análises complementares, como os testes de causalidade de Granger (Hamilton, 1994).

Adicionalmente, o modelo VAR assume que todas as variáveis inseridas são estacionárias, ou seja, que possuem média e variância constantes ao longo do tempo. No entanto, variáveis macroeconômicas e financeiras, como a taxa de *default*, frequentemente apresentam tendências ou comportamentos não estacionários, o que torna necessário aplicar transformações como diferenciação antes da modelagem. A não observância dessa condição pode levar a estimativas inválidas ou inconsistentes (Stock; Watson, 2019).

A especificação do número adequado de defasagens representa outro ponto crítico. Um número insuficiente pode omitir informações relevantes e gerar viés nas estimativas, enquanto um número excessivo pode introduzir multicolinearidade e reduzir

a eficiência do modelo. A seleção das defasagens deve ser criteriosa, baseada em métricas como o Critério de Informação de Akaike (AIC) ou o Critério de Informação Bayesiano (BIC). (Lütkepohl, 2005).

Por fim, o VAR não é projetado para capturar relações não lineares entre variáveis, o que pode limitar sua aplicabilidade em contextos financeiros mais complexos, nos quais as interações entre variáveis macroeconômicas e inadimplência muitas vezes seguem padrões não lineares ou sujeitos a mudanças estruturais (Greene, 2018).

3.5 MODELO XGBREGRESSOR

O XGBRegressor é um modelo de regressão baseado no algoritmo de aprendizado de máquina Gradient Boosting, integrado à biblioteca XGBoost (eXtreme Gradient Boosting. Desenvolvido para lidar com problemas de regressão em contextos de alta complexidade, o XGBRegressor destaca-se pela sua capacidade de processar grandes volumes de dados com elevada precisão e velocidade de treinamento.

O funcionamento do algoritmo baseia-se em uma abordagem iterativa, na qual árvores de decisão são construídas sequencialmente. A cada iteração, o modelo busca minimizar os erros residuais cometidos pelas árvores anteriores, por meio da otimização de uma função objetivo específica. Esse processo de boosting progressivo permite que o modelo aprenda padrões complexos e não lineares presentes nos dados, adaptando-se de forma eficaz a diferentes estruturas de variáveis explicativas.

A flexibilidade do XGBRegressor o torna particularmente útil em cenários onde os modelos tradicionais, como regressões lineares ou modelos de séries temporais, enfrentam limitações para capturar relações não lineares, interações complexas entre variáveis e estruturas de dados com ruídos ou valores atípicos.

A fórmula geral do modelo pode ser descrita como:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \quad (5)$$

Onde:

$\hat{y}_i$: Predição final para o i-ésimo exemplo;

k: Número total de árvores no modelo;f

$f_k(x_i)$: Representa a k-ésima árvore de decisão aplicada aos dados x_i;

$\mathcal{F}$: Conjunto de todas as árvores possíveis.

No contexto deste estudo, o XGBRegressor será utilizado para prever a taxa de *default* com base em variáveis macroeconômicas e fatores exógenos. A escolha por esse modelo justifica-se por suas diversas vantagens em relação aos métodos tradicionais:

Capacidade de lidar com não linearidades: O XGBRegressor é capaz de capturar relações complexas e não lineares entre a taxa de *default* e as variáveis explicativas, superando as limitações dos modelos lineares tradicionais.

Regularização integrada: O modelo incorpora técnicas de regularização L1 e L2, que reduzem o risco de overfitting, aumentando a capacidade de generalização para novos dados.

Importância das variáveis: O XGBRegressor fornece métricas como ganho, cobertura e frequência, permitindo avaliar a importância relativa das variáveis explicativas e auxiliando na interpretação dos resultados.

Flexibilidade na modelagem: O algoritmo permite ajustes precisos de hiperparâmetros, como taxa de aprendizado, profundidade das árvores e número de iterações, otimizando o desempenho preditivo.

Apesar dessas vantagens, o XGBRegressor apresenta algumas limitações importantes. Uma delas é sua alta complexidade computacional. O processo de construção sequencial de múltiplas árvores de decisão, inerente à metodologia de gradient boosting, demanda recursos computacionais significativos, especialmente em conjuntos de dados extensos ou com grande número de variáveis explicativas (Chen; Guestrin, 2016). Essa característica pode restringir sua aplicação em contextos com infraestrutura tecnológica limitada.

Outro desafio relevante é a interpretação dos resultados. Embora o modelo forneça medidas de importância das variáveis, como ganho e cobertura, a interpretação dos efeitos individuais de cada variável sobre a taxa de *default* não é tão direta quanto

em modelos lineares, o que pode ser uma desvantagem em contextos como o da *PD Forward Looking*, onde a interpretação econômica das relações é tão relevante quanto a acurácia preditiva (Friedman, 2001).

Além disso, o XGBoost é suscetível ao risco de overfitting, especialmente quando configurado com muitas árvores, profundidade elevada ou taxa de aprendizado muito baixa. Apesar das técnicas internas de regularização, esse risco permanece em cenários com alta variabilidade ou ruído nos dados, como é comum em séries temporais financeiras (Hastie; Tibshirani; Friedman, 2010).

Outro ponto de atenção é o tratamento da estrutura temporal dos dados. O XGBRegressor não é nativamente projetado para lidar com séries temporais, exigindo que o usuário inclua manualmente defasagens, médias móveis ou outras variáveis derivadas para capturar as dinâmicas temporais. Esse requisito demanda pré-processamento rigoroso e aumenta a complexidade do desenvolvimento do modelo (Chen; Guestrin, 2016).

Por fim, a eficácia do modelo é altamente dependente da qualidade dos dados de entrada. Dados ausentes, outliers ou mal normalizados podem comprometer o desempenho do modelo, exigindo tratamento cuidadoso. Ademais, a escolha adequada dos hiperparâmetros é crucial, pois configurações inadequadas podem reduzir significativamente a performance preditiva (Friedman, 2001).

4 METODOLOGIA

Este capítulo apresenta de forma detalhada as etapas realizadas neste estudo para a implementação do modelo de *PD Forward Looking*. Serão descritas as ferramentas utilizadas no processo de modelagem, bem como os procedimentos adotados para o tratamento das variáveis.

4.1 BASE DE MODELAGEM

Para a realização desta pesquisa, será utilizada uma base de dados fornecida pelo Banrisul, contendo, em frequência mensal, o percentual de contratos inadimplentes de uma carteira de crédito da instituição. A amostra abrange o período de janeiro de 2018 a dezembro de 2023. Para fins de modelagem, os dados de janeiro de 2018 a setembro de 2022 serão utilizados na etapa de desenvolvimento do modelo, enquanto o período de outubro de 2022 a dezembro de 2023 será reservado para a validação, compondo a chamada "safra de teste".

A fim de aplicar a técnica de regressão múltipla de forma eficaz, foi necessário realizar transformações nas variáveis macroeconômicas, visando tratar a sazonalidade e a tendência presentes nas séries temporais. Neste estudo, optou-se pela utilização do conceito de retorno contínuo, calculado com base em defasagens mensais (lags) de 1, 3, 6 e 12 meses, de modo a capturar a dinâmica temporal e os efeitos de variações passadas sobre a taxa de *default*.

As séries macroeconômicas podem ser acessadas por meio do site IPEADATA nos seguintes links:

a) IPCA: http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38391;
b) Desemprego: http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=1347352645;
c) PIB: http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=521274780&module=M;
d) Selic: http://www.ipeadata.gov.br/exibeserie.aspx?serid=38402;
e) Câmbio: http://www.ipeadata.gov.br/exibeserie.aspx?serid=38389.

4.2 MULTICOLINEARIDADE

A multicolinearidade ocorre quando duas ou mais variáveis explicativas em um modelo de regressão apresentam elevada correlação entre si, o que dificulta a identificação dos efeitos individuais de cada variável sobre a variável dependente. Esse fenômeno compromete a independência das informações fornecidas por cada variável preditora, afetando negativamente a precisão das estimativas dos coeficientes (Gujarati; Porter, 2011).

Embora a multicolinearidade não viole diretamente os pressupostos fundamentais de modelos lineares, ela pode gerar implicações significativas na análise dos resultados, como instabilidade nos coeficientes estimados e dificuldades de interpretação econômica. Em situações de alta multicolinearidade, os coeficientes podem assumir valores não intuitivos ou inconsistentes com a teoria econômica, além de apresentar p-valores distorcidos, que não refletem adequadamente a significância estatística das variáveis (Montgomery; Peck; Vining, 2012).

A avaliação da multicolinearidade é comumente realizada por meio do Fator de Inflação da Variância (Variance Inflation Factor – VIF), que quantifica o grau de correlação de uma variável explicativa com as demais variáveis do modelo. Neste estudo, foi adotado um critério mais flexível, considerando aceitáveis os valores de VIF inferiores a 5. Esse limite é amplamente utilizado na literatura para detectar multicolinearidade severa, embora, em modelos que requerem maior precisão, recomenda-se a adoção de limites mais conservadores (Wooldridge, 2020).

Os critérios gerais para interpretação do VIF são os seguintes:

VIF ≤ 1: Indica ausência de correlação entre a variável explicativa e as demais.
1 < VIF ≤ 5: Representa uma correlação moderada, geralmente aceitável.
5 < VIF ≤ 10: Indica uma correlação mais elevada, mas ainda pode ser tolerável dependendo do contexto.
VIF > 10: Sugere uma multicolinearidade severa, o que compromete a confiabilidade das estimativas.

É importante ressaltar que a multicolinearidade possui impacto limitado em

modelos VAR. Nesse tipo de modelagem, o foco principal não está na interpretação individual dos coeficientes, mas sim na análise das interações dinâmicas entre variáveis ao longo do tempo. Por essa razão, a presença de correlação entre variáveis explicativas não inviabiliza a aplicação do modelo, desde que este esteja corretamente especificado em termos de número de defasagens e que as séries utilizadas sejam estacionárias (Lütkepohl, 2005).

No entanto, é fundamental que a seleção das variáveis e defasagens esteja alinhada com a teoria econômica e com a evidência empírica, de modo a garantir a consistência dos resultados. Caso a multicolinearidade seja considerada problemática, especialmente em situações de especificações amplas com muitas variáveis e defasagens, algumas estratégias podem ser adotadas para mitigar seus efeitos:

a) **redução do número de variáveis explicativas altamente correlacionadas:** remover ou combinar variáveis;
b) **utilização de regularização:** métodos como Ridge ou LASSO ajudam a mitigar os efeitos da multicolinearidade em modelos de regressão;
c) **ajuste no pré-processamento dos dados:** realizar transformações que reduzam a correlação entre variáveis;
d) **aumento da amostra:** um maior número de observações pode atenuar o problema.

4.3 ANÁLISE DAS VARIÁVEIS MACROECONÔMICAS

O uso de defasagens nas variáveis macroeconômicas é essencial na modelagem da *PD Forward Looking*, uma vez que os efeitos das mudanças nos indicadores econômicos não se manifestam de forma imediata sobre a probabilidade de *default*. Em geral, esses efeitos se propagam ao longo do tempo, influenciando gradualmente o comportamento financeiro de indivíduos e empresas. A incorporação de defasagens permite capturar essas dinâmicas temporais, aumentando a precisão das estimativas (Hamilton, 1994).

Na modelagem de séries temporais, é fundamental reconhecer que as variáveis macroeconômicas possuem uma estrutura temporal própria e que seu impacto sobre a variável dependente pode ocorrer com defasagens. A utilização sistemática de

informações históricas por meio das defasagens contribui para previsões mais assertivas e fundamentadas (Stock; Watson, 2019). Além disso, ao considerar defasagens múltiplas, é possível avaliar o impacto de diferentes horizontes temporais, possibilitando a identificação de efeitos de curto e longo prazo sobre a taxa de *default*.

A ausência de defasagens pode acarretar a omissão de informações relevantes, resultando em estimativas enviesadas ou inconsistentes. Isso ocorre porque o modelo deixa de considerar o efeito acumulado das variáveis macroeconômicas ao longo do tempo, o que compromete a confiabilidade dos coeficientes estimados e, por consequência, das previsões (Gujarati; Porter, 2011).

Do ponto de vista técnico, as defasagens possibilitam a incorporação da dependência temporal sem violar os pressupostos fundamentais dos modelos econométricos. Em modelos como o VAR e o ARIMAX, as defasagens são particularmente relevantes, permitindo descrever com maior precisão as relações dinâmicas entre variáveis e realizar previsões diante de choques exógenos ou mudanças nos indicadores econômicos (Hamilton, 1994).

A tabela abaixo reúne as correlações entre as variáveis macroeconômicas para os lags 1, 3, 6 e 12 e a taxa de *default* a ser explorada:

Tabela 1 – Correlação com a taxa de *default*

LAG (Defasagem)	PIB	IPCA	SELIC	CÂMBIO	DESEMPREGO
1	-0,657	-0,068	0,460	-0,696	0,108
3	-0,650	0,015	0,514	-0,653	-0,008
6	-0,636	0,077	0,595	-0,628	-0,195
12	-0,647	0,251	0,639	-0,582	-0,541

Fonte: elaboração própria com base em IPEA (2024) e Banrisul (2024)[1].

Foi realizado um levantamento na literatura econômica com o objetivo de analisar as correlações entre variáveis macroeconômicas e a taxa de *default*, visando compreender o efeito esperado de cada variável sobre a variável-alvo. Essa análise teórica fundamentou a seleção e o tratamento das variáveis explicativas, permitindo antecipar o sinal esperado das relações e orientar a interpretação dos resultados obtidos nos modelos preditivos.

[1] Dados obtidos por consulta ao percentual de contratos em *default* em base interna do Banrisul.

4.3.1 PIB

A relação entre o Produto Interno Bruto (PIB) e a probabilidade de *default* no Banrisul reflete a influência do desempenho econômico geral sobre a capacidade de adimplência dos tomadores de crédito. Com base nos dados da tabela, há uma correlação negativa forte entre o PIB e a taxa de inadimplência em todas as defasagens, começando em -0,657 em 1 mês, -0,650 em 3 meses, -0,636 em 6 meses e -0,647 em 12 meses. Essa consistência ao longo do tempo reforça que um crescimento econômico está associado a uma redução significativa nas taxas de *default*.

O Banrisul, com forte atuação no agronegócio e no crédito consignado, está diretamente exposto aos impactos do PIB. Em períodos de crescimento econômico, há um aumento na geração de renda e na atividade econômica, o que melhora a capacidade de pagamento tanto de empresas quanto de indivíduos. O setor agropecuário, em particular, se beneficia de ciclos econômicos favoráveis, com maior demanda interna e externa por produtos agrícolas, o que fortalece a solvência dos produtores rurais. Além disso, o crescimento do PIB contribui para a estabilidade do mercado de trabalho e, consequentemente, para a renda dos tomadores de crédito consignado, reduzindo ainda mais os riscos de inadimplência.

Por outro lado, em períodos de recessão ou desaceleração econômica, a queda no PIB impacta negativamente as condições financeiras dos tomadores de crédito. No agronegócio, uma redução na atividade econômica pode significar menos demanda por produtos agrícolas ou preços menos favoráveis, enquanto no crédito consignado, o aumento do desemprego ou da pressão sobre os salários reduz a renda disponível para cobrir outras dívidas.

A literatura econômica reforça essas dinâmicas. Bernanke, Gertler e Gilchrist (1999) introduziram o conceito do Financial Accelerator, destacando como o crescimento do PIB reduz riscos financeiros ao aumentar a renda e o valor colateral dos ativos, fortalecendo a solvência de famílias e empresas. Mishkin (2007) também aponta que o PIB é um dos indicadores mais importantes para medir a saúde econômica geral e sua influência sobre a inadimplência.

No contexto brasileiro, Santolin e Gama (2021) mostram que a relação entre renda per capita (PIB dividido pelo número de habitantes) e inadimplência é particularmente forte. A alta da inadimplência influencia negativamente o crescimento de renda no país.

A correlação negativa entre o PIB e a taxa de *default* do Banrisul reflete a importância do desempenho econômico na capacidade de pagamento dos tomadores de crédito. Em períodos de crescimento econômico, tanto o agronegócio quanto o crédito consignado se beneficiam de uma maior geração de renda e de condições financeiras mais favoráveis, resultando em menores taxas de inadimplência. Por outro lado, períodos de retração econômica amplificam os riscos, especialmente em segmentos mais sensíveis à renda e à atividade econômica.

4.3.2 IPCA

A análise da relação entre o IPCA (Índice Nacional de Preços ao Consumidor Amplo) e a probabilidade de *default* no Banrisul deve considerar a estrutura de atuação do banco, que inclui forte presença no crédito consignado e no agronegócio. Na tabela, a correlação entre o IPCA e a inadimplência é baixa em defasagens curtas, começando em -0,068 na defasagem de 1 mês, praticamente inexistente, e subindo para 0,015 em 3 meses. O impacto da inflação começa a ser mais perceptível em horizontes maiores, com correlações de 0,077 em 6 meses e 0,251 em 12 meses, indicando um efeito moderado, mas ainda relevante, da inflação sobre a inadimplência.

Essa dinâmica pode ser explicada pela maneira como a inflação afeta o custo de vida e, consequentemente, a capacidade de pagamento dos clientes do banco. No caso do crédito consignado, que é amplamente oferecido pelo Banrisul, as parcelas são descontadas diretamente da folha de pagamento, o que reduz a sensibilidade imediata à inflação. No entanto, em períodos prolongados de alta inflação, a renda disponível dos clientes é reduzida, pois os salários geralmente não acompanham os aumentos nos preços de bens e serviços. Isso pode levar a um aumento na inadimplência, principalmente entre aqueles que possuem múltiplos compromissos financeiros.

Além disso, a inflação também afeta as operações do agronegócio, outra área relevante para o Banrisul. Embora os produtos agrícolas exportados possam se

beneficiar de preços mais altos em dólar, custos crescentes de insumos dolarizados, como fertilizantes e máquinas, podem pressionar a margem de lucro dos produtores. Isso pode resultar em maior dificuldade para honrar compromissos financeiros, aumentando o risco de inadimplência em horizontes de médio e longo prazo.

A literatura econômica oferece suporte para essas observações. Mishkin (2007) argumenta que a inflação afeta a estabilidade financeira de maneira indireta, pressionando o poder de compra das famílias e as margens de lucro das empresas. Friedman (1968) também destaca que a inflação inesperada gera incertezas que podem levar a ajustes financeiros mais lentos, aumentando os riscos de inadimplência em períodos prolongados.

Gomes *et al.* (2024), analisaram o impacto de alguns indicadores econômicos, entre eles o IPCA, na quantidade de contratos e no valor dos empréstimos consignados no período de 2014 à 2021 no Brasil. Puderam observar que, no longo prazo, a inflação exerceu pressão na renda dos aposentados, aumentando seu envididamento.

A relação entre o IPCA e a taxa de *default* do Banrisul é moderada e mais evidente em defasagens longas. Enquanto o crédito consignado oferece uma certa resiliência contra os efeitos diretos da inflação, a perda de renda disponível dos clientes em períodos prolongados de alta inflação pode contribuir para o aumento da inadimplência. No agronegócio, o impacto da inflação é mais misto, dependendo das margens de lucro e dos custos dos insumos. Essa dinâmica reforça a importância de estratégias de gestão de risco para mitigar os impactos de longo prazo da inflação.

4.3.3 SELIC

A SELIC, sendo a taxa básica de juros da economia brasileira, exerce uma influência direta no custo do crédito e na dinâmica financeira de empresas e indivíduos. Na tabela apresentada, a correlação entre a SELIC e a taxa de *default* é positiva e consistente ao longo do tempo, começando em 0,460 na defasagem de 1 mês e crescendo gradualmente para 0,514 em 3 meses, 0,595 em 6 meses e alcançando 0,639 em 12 meses. Esses números indicam que, à medida que a SELIC aumenta, a inadimplência também tende a crescer, especialmente em horizontes de médio e longo prazo.

Essa correlação positiva pode ser explicada pelo impacto da SELIC nos custos de financiamento e na renda disponível. No caso do crédito consignado, uma modalidade de empréstimo muito utilizada pelo Banrisul, a taxa de inadimplência tende a ser baixa, pois as parcelas são descontadas diretamente na folha de pagamento. No entanto, a elevação da SELIC pode reduzir a demanda por novos financiamentos e pressionar a renda disponível dos tomadores, especialmente em contextos de aumento da inflação e custos gerais de vida. Isso pode gerar um efeito indireto na capacidade de adimplência dos clientes, particularmente aqueles que possuem outras dívidas ou dependem de ajustes salariais em defasagem com a inflação.

Além disso, a SELIC impacta diretamente as condições macroeconômicas, afetando o custo de capital das empresas e o orçamento das famílias. Em cenários de aumento da SELIC, o custo do crédito no mercado se eleva, e isso pode afetar tomadores de crédito em outras modalidades, como o crédito pessoal ou empresarial, que compõem parte da carteira do Banrisul. Para o agronegócio, por exemplo, o crédito rural subsidiado pode mitigar os efeitos da alta da SELIC, mas o impacto indireto nas famílias e empresas urbanas pode levar a um aumento gradual na inadimplência.

A literatura econômica oferece suporte para essa análise. Mishkin (2007) explica que as taxas de juros desempenham um papel central na determinação das condições financeiras, influenciando diretamente os custos do crédito e a solvência financeira dos agentes econômicos. No caso brasileiro, Bernanke e Gertler (1995) destacam que economias emergentes são particularmente sensíveis às variações das taxas de juros devido à estrutura do mercado de crédito e à alavancagem de empresas e famílias.

A correlação positiva entre a SELIC e a taxa de *default* do Banrisul reflete o impacto cumulativo da política monetária sobre o custo do crédito e as condições econômicas gerais. Embora o crédito consignado ofereça proteção contra inadimplência no curto prazo, os efeitos indiretos de uma SELIC elevada sobre o orçamento familiar e empresarial tornam-se mais evidentes ao longo do tempo, especialmente em horizontes de 6 a 12 meses. Essa dinâmica ressalta a importância da gestão de risco e da diversificação da carteira de crédito do banco em períodos de alta dos juros.

4.3.4 Câmbio

A análise da relação entre o câmbio e a taxa de *default* ganha ainda mais clareza quando consideramos que a taxa de inadimplência observada refere-se ao Banrisul, um banco que tem atuação forte no financiamento ao agronegócio, setor altamente influenciado pelas variações cambiais. Na tabela apresentada, a correlação negativa significativa entre o câmbio e a probabilidade de *default* ao longo de todas as defasagens reforça como as dinâmicas do mercado de câmbio afetam diretamente o setor agropecuário e, por extensão, a carteira de crédito do banco.

Na defasagem de 1 mês, a correlação de -0,696 já evidencia que uma depreciação do câmbio (enfraquecimento do real em relação ao dólar) está associada a uma redução significativa na taxa de inadimplência. Essa relação se mantém forte em defasagens maiores, com coeficientes de -0,653 aos 3 meses, -0,628 aos 6 meses e -0,582 aos 12 meses. Isso reflete o impacto direto que o câmbio exerce sobre a rentabilidade e a solvência do agronegócio, principal setor financiado pelo banco.

O setor agropecuário brasileiro é amplamente exportador, beneficiando-se diretamente de um câmbio depreciado. Quando o real se desvaloriza, os produtos agrícolas tornam-se mais competitivos no mercado internacional, aumentando as receitas dos produtores. Esse aumento de receita fortalece a capacidade dos agricultores de honrar seus compromissos financeiros, resultando em uma redução nas taxas de *default* das operações de crédito rural e empresarial. Adicionalmente, os ganhos cambiais ajudam a compensar eventuais pressões inflacionárias decorrentes do aumento no custo de insumos importados.

Por outro lado, em períodos de valorização do real, o setor agropecuário pode sofrer com margens reduzidas, já que os preços recebidos pelos produtores no mercado internacional, quando convertidos para a moeda nacional, tornam-se menos atrativos. Essa dinâmica pode aumentar a pressão sobre a inadimplência, especialmente entre produtores mais alavancados ou menos eficientes.

A relação entre o câmbio e o desempenho financeiro de setores exportadores, como o agronegócio, é amplamente discutida na literatura econômica. Bernanke e Gertler (1995) destacam que o câmbio afeta diretamente a solvência de empresas em economias emergentes, onde os setores exportadores desempenham um papel-chave. Mishkin (2007) reforça que as variações cambiais podem ter um efeito estabilizador ou

desestabilizador, dependendo do grau de exposição dos agentes econômicos ao mercado externo.

No contexto brasileiro, a depreciação cambial tende a beneficiar os produtores rurais exportadores, uma vez que aumenta sua receita em moeda nacional e melhora sua capacidade de pagamento de dívidas. Segundo Camuri (2016), a valorização do dólar frente ao real eleva a competitividade dos produtos agropecuários no mercado externo, impactando positivamente a rentabilidade do setor. Esse efeito é especialmente relevante para instituições financeiras regionais com maior concentração de crédito rural em suas carteiras, pois a maior solidez financeira dos produtores reduz os riscos de inadimplência e fortalece a saúde das carteiras de crédito agrícola.A relação inversa entre o câmbio e a taxa de *default* do Banrisul reflete a influência predominante do agronegócio na economia gaúcha e na carteira de crédito do banco. A depreciação cambial favorece a rentabilidade dos produtores rurais, reduzindo a inadimplência, enquanto uma valorização do real pode comprometer as margens financeiras do setor e aumentar o risco de crédito. Essa dinâmica ressalta a importância do câmbio como variável-chave na gestão de riscos de bancos com forte exposição ao agronegócio.

4.3.5 Desemprego

A análise da correlação entre o desemprego e a taxa de *default* revela uma relação que se intensifica com o passar do tempo, especialmente em defasagens mais longas. Na defasagem de 1 mês, a correlação é de 0,108, praticamente inexistente, indicando que, no curto prazo, a taxa de desemprego não exerce influência significativa sobre a inadimplência. Contudo, conforme os meses avançam, a correlação se torna gradualmente mais negativa: -0,008 na defasagem de 3 meses, -0,195 em 6 meses e -0,541 em 12 meses. Esse padrão sugere que o impacto do desemprego na probabilidade de *default* é mais perceptível no médio e longo prazo.

Essa relação negativa entre o desemprego e a inadimplência, pode ser explicada pela composição da carteira de crédito do banco e a forte influência do setor agropecuário na economia gaúcha. O agronegócio, amplamente financiado pelo banco, tende a ser menos sensível ao desemprego urbano, uma vez que sua dinâmica está

mais relacionada ao desempenho do mercado externo e à produtividade do campo do que às condições do mercado de trabalho nas áreas urbanas. Assim, o impacto do desemprego nas taxas de *default* do banco não é imediato e aparece de forma mais significativa apenas quando o desemprego persiste por períodos prolongados, afetando o consumo geral, os fluxos financeiros e a capacidade de pagamento das famílias.

Além disso, a dinâmica inversa entre desemprego e inadimplência pode ser influenciada por políticas públicas e subsídios ao crédito rural, que costumam ser implementados em períodos de dificuldade econômica, como forma de proteger o setor agrícola. Esses mecanismos ajudam a suavizar os impactos negativos do desemprego no curto prazo, mas, à medida que os efeitos acumulados do desemprego se propagam pela economia, a inadimplência pode aumentar, especialmente entre setores urbanos mais dependentes de renda estável.

A literatura econômica também fornece suporte para essa análise. Mishkin (2007) observa que o desemprego afeta o crédito de forma indireta, reduzindo o consumo e a renda disponível, mas os impactos variam de acordo com a estrutura econômica de cada país ou região. Bernanke, Gertler e Gilchrist (1999) destacam que o desemprego prolongado pode comprometer a solvência de famílias e empresas, especialmente em setores que dependem do mercado doméstico, como o varejo e os serviços.

Por outro lado, bancos regionais com perfil de carteira voltado ao agronegócio, como o Banrisul, tendem a apresentar maior resiliência a choques de desemprego e recessão. Isso ocorre porque o setor agrícola, tradicionalmente menos exposto às flutuações do mercado de trabalho urbano, mantém relativa estabilidade na geração de receitas mesmo em cenários macroeconômicos adversos. Dados da Serasa Experian (2024) mostram que, enquanto a inadimplência de pessoas físicas no Brasil ultrapassava 30%, no setor rural a taxa era de apenas 7,7%, com os pequenos produtores registrando os menores índices (6,9%). Além disso, a região Sul — onde o Banrisul possui forte atuação — apresentou a menor taxa regional, com 5,0% de inadimplência entre produtores rurais.

O Banrisul possui uma significativa participação do crédito consignado em sua carteira ativa, com forte exposição a servidores públicos, aposentados e pensionistas,

cujo perfil é caracterizado por estabilidade de renda e baixo risco de inadimplência (Banrisul, 2023). Como o crédito consignado é descontado diretamente da folha de pagamento, sua inadimplência tende a ser estruturalmente menor, mesmo em períodos de adversidade econômica.

Em cenários de aumento do desemprego, a inadimplência nos produtos de crédito tradicional pode subir, mas os efeitos sobre o consignado são limitados, sobretudo entre servidores públicos e beneficiários do INSS. Além disso, o aumento do desemprego pode levar os bancos a adotar posturas mais conservadoras na concessão de crédito, o que contribui para preservar a qualidade da carteira (Gomes *et al.,* 2024).

Nesse contexto, a correlação negativa entre a taxa de desemprego e a inadimplência observada na carteira do Banrisul pode ser explicada pelo perfil resiliente dos tomadores e pelo comportamento anticíclico das instituições financeiras com foco em crédito consignado (IPEA, 2022).

4.4 MÉTRICAS DE AVALIAÇÃO DOS MODELOS

No processo de estimar a PD de uma carteira de crédito, é essencial avaliar a qualidade das previsões geradas pelos modelos desenvolvidos. Entre as métricas amplamente utilizadas para este propósito estão o Erro Percentual Médio Absoluto (MAPE) e o Raiz do Erro Quadrático Médio (RMSE). Ambas possuem características específicas que permitem analisar a precisão das previsões, oferecendo insights importantes para a tomada de decisão.

4.4.1 Erro Percentual Médio Absoluto (MAPE)

O MAPE mede a precisão do modelo ao quantificar a média do erro absoluto em termos percentuais, o que o torna intuitivo e útil para comunicar resultados a públicos não técnicos. Ele é calculado pela fórmula:

$$[\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \times 100] \quad (6)$$

Onde:

y_i: representa o valor observado da variável alvo;
$\widehat{y_i}$: representa o valor previsto pelo modelo;
n: é o número total de observações.

A métrica expressa o erro como uma porcentagem média das observações reais, o que permite uma comparação direta entre diferentes modelos ou conjuntos de dados, independentemente da escala. No entanto, o MAPE pode ser sensível a valores muito próximos de zero em y_i, levando a resultados distorcidos.

4.4.2 Raiz do Erro Quadrático Médio (RMSE)

Por sua vez, o RMSE avalia a magnitude do erro em termos absolutos, dando maior peso a desvios maiores devido à sua estrutura quadrática. Sua fórmula é:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \quad (7)$$

Onde os termos são definidos da mesma forma que no MAPE. A métrica é expressa na mesma unidade da variável dependente, tornando-a útil para compreender o impacto real dos erros no contexto analisado. Por penalizar erros maiores, o RMSE é particularmente adequado quando desvios extremos nas previsões são críticos para o processo decisório.

4.4.3 Critério de Informação de Akaike (AIC)

O Critério de Informação de Akaike (AIC) é amplamente utilizado em estatística para avaliar a qualidade relativa de modelos estatísticos ajustados ao mesmo conjunto de dados. Proposto por Hirotsugu Akaike em 1974, o AIC introduziu uma abordagem prática para a escolha de modelos, equilibrando o ajuste aos dados com a simplicidade do modelo. Segundo Akaike, "o AIC é uma medida da qualidade do modelo que considera tanto a bondade do ajuste quanto a penalização por complexidade, ajudando

a evitar o problema do overfitting" (Akaike, 1974).

Matematicamente, o AIC é definido como:

$$\mathrm{AIC} = 2\mathrm{k} - 2\ln(\mathrm{L}) \quad (8)$$

Onde k é o número de parâmetros estimados no modelo, e L é a verossimilhança máxima do modelo, uma medida de quão bem ele ajusta os dados. A inclusão de $2k$ na fórmula é fundamental, pois penaliza a complexidade do modelo. Modelos mais complexos, com mais parâmetros, podem se ajustar melhor aos dados observados, mas podem sofrer de overfitting, isto é, capturar ruídos específicos dos dados em vez de padrões generalizáveis (Burnham; Anderson, 2002).

Um dos principais atrativos do AIC é sua capacidade de comparar diferentes modelos ajustados ao mesmo conjunto de dados. A regra geral é que, entre dois modelos, aquele com o menor valor de AIC deve ser preferido. Contudo, o AIC não oferece uma escala absoluta; ele serve apenas como critério comparativo. Segundo Burnham e Anderson (2002), "o AIC não busca o modelo verdadeiro, mas sim o modelo mais próximo da realidade dentro do conjunto avaliado".

Apesar de sua popularidade, o AIC apresenta limitações. Ele não considera diretamente a capacidade preditiva do modelo, mas sim sua adequação aos dados observados. Além disso, em situações com dados não independentes, como algumas séries temporais, o uso do AIC pode ser inadequado se o modelo não for bem especificado (Claeskens; Hjort, 2008).

4.4.4 R² Ajustado

O R² ajustado é uma métrica estatística que desempenha um papel fundamental na avaliação da qualidade de ajuste de modelos de regressão múltipla. Sua principal virtude é ir além do R² simples, ajustando a proporção da variação explicada pelo modelo com base no número de variáveis independentes e no tamanho da amostra. Essa correção é crucial para evitar interpretações equivocadas, especialmente em modelos com muitas variáveis.

Supomos que um modelo está sendo desenvolvido e decide-se incluir novas

variáveis para tentar melhorar a explicação da variável dependente. O R^2 simples, nesse caso, aumentará automaticamente, independentemente de as novas variáveis realmente contribuírem de forma significativa. É aí que o R^2 ajustado se faz presente: ele penaliza a inclusão de variáveis irrelevantes, oferecendo uma medida mais confiável da capacidade explicativa do modelo.

No campo da econometria, Gujarati e Porter (2011) destacam que o R^2 ajustado é indispensável para evitar o superajuste, um problema em que o modelo se torna excessivamente complexo sem ganho real de explicação. De forma semelhante, Montgomery, Peck e Vining (2012) enfatizam que ele é especialmente útil em análises comparativas, ajudando na escolha de um modelo mais parcimonioso. Já Hair *et al.* (2019) reforçam que sua aplicação é essencial em estudos de regressão múltipla, garantindo que os resultados sejam interpretados com rigor.

A interpretação do R^2 ajustado em modelos econômicos e financeiros requer uma compreensão contextual, especialmente porque essas áreas são caracterizadas por alta complexidade e múltiplos fatores exógenos que afetam as variáveis analisadas. De acordo com (Gujarati; Porter, 2011), em ciências sociais e econômicas, é comum que o R^2 ajustado seja relativamente baixo, pois muitos dos fatores que influenciam as variáveis dependentes não podem ser capturados totalmente em modelos econométricos. Isso ocorre porque variáveis como comportamento humano, choques econômicos e mudanças de política são frequentemente difíceis de mensurar ou prever.

Além disso, Wooldridge (2020) destaca que "modelos preditivos em economia e finanças devem equilibrar a simplicidade e a explicação. Embora um R^2 ajustado baixo possa ser visto como uma limitação, ele não implica necessariamente que o modelo seja inútil, especialmente em cenários onde a previsão é mais relevante que a explicação total". Essa perspectiva é especialmente válida para modelos que buscam prever PD, os quais lidam com dados que refletem não apenas condições econômicas, mas também fatores comportamentais e psicológicos dos tomadores de crédito.

Blanchard (2018) abordou críticas aos modelos macroeconômicos modernos em seu artigo "On the future of macroeconomic models", publicado no Oxford Review of Economic Policy. Nesse trabalho, Blanchard discute as deficiências dos modelos DSGE (Dynamic Stochastic General Equilibrium), destacando que eles muitas vezes omitem

características essenciais do comportamento de firmas e indivíduos, o que pode comprometer sua capacidade de capturar a realidade econômica de forma eficaz.

Por fim, Stock e Watson (2019) explicam que o objetivo do R^2 ajustado não é maximizar a explicação de variância, mas garantir que o modelo seja parcimonioso e útil para a análise. "Em áreas como crédito e finanças, onde há alta variabilidade externa, um R^2 ajustado modesto é esperado, e a análise deve focar na significância das variáveis explicativas e na qualidade do modelo como um todo".

4.4.5 Resíduos

A análise de resíduos é crucial para validar os pressupostos de um modelo de regressão. Testes como Jarque-Bera, Ljung-Box e Breusch-Pagan avaliam aspectos fundamentais como normalidade, independência e homocedasticidade dos resíduos, respectivamente. A interpretação correta desses testes é essencial para garantir a validade estatística do modelo.

4.4.5.1 Teste de Jarque-Bera

Objetivo: Avaliar se os resíduos seguem uma distribuição normal.

Hipóteses:

H_0 : Os resíduos seguem uma distribuição normal.

H_1: Os resíduos não seguem uma distribuição normal.

Estatística e p-valor: O teste utiliza a assimetria (skewness) e a curtose (kurtosis) para calcular uma estatística JB. Um p-valor menor que o nível de significância (α, geralmente 0,05) rejeita H_0, indicando não normalidade dos resíduos.

Implicações: Gujarati e Porter (2011) destacam que "a normalidade dos resíduos é essencial para a aplicação de testes de hipótese confiáveis e a construção de intervalos de confiança válidos". Resíduos não normais podem apontar a presença

de outliers ou variáveis omitidas.

4.4.5.2 Teste de Ljung-Box

Objetivo: Verificar a presença de autocorrelação nos resíduos.

Hipóteses:

H_0: Não há autocorrelação nos resíduos (resíduos independentes).

H_1: Há autocorrelação nos resíduos.

Estatística e p-valor: O teste calcula a estatística Q, baseada nas autocorrelações dos resíduos para diferentes defasagens ($lags$). Greene (2018) observa que “um número moderado de defasagens, como 10, é suficiente para detectar dependências significativas em resíduos de séries temporais que não apresentam alta frequência.” Se o p-valor for menor que α, rejeitamos H_0, indicando autocorrelação significativa.

Implicações: Wooldridge (2020) afirma que "a autocorrelação nos resíduos pode comprometer a validade das inferências estatísticas e sugere que o modelo não capturou completamente as relações temporais". Nesses casos, é necessário incluir termos autorregressivos ou defasados no modelo.

4.4.5.3 Teste de Breusch-Pagan

Objetivo: Avaliar se os resíduos apresentam homocedasticidade (variância constante).

Hipóteses:

H_0: Os resíduos possuem variância constante.

H_1: A variância dos resíduos varia em função de variáveis explicativas.

Estatística e p-valor: O teste regressa os resíduos quadrados em relação às variáveis explicativas e utiliza a estatística BP. Um p-valor menor que α rejeita H_0, indicando heterocedasticidade.

Implicações: Greene (2018) explica que "a heterocedasticidade leva a estimativas de mínimos quadrados ineficientes e erros padrão incorretos".

Interpretação dos p-valores

- p-valor > 0,05: Não rejeitamos H_0, indicando que os resíduos atendem ao pressuposto avaliado pelo teste.

- p-valor <= 0,05: Rejeitamos H_0, sugerindo a necessidade de ajustes no modelo.

4.4.6 Testes estatísticos

Descrição dos testes estatísticos utilizados na pesquisa como subsídio para a especificação de modelos de séries temporais.

4.4.6.1 Função de Autocorrelação (ACF)

O teste ACF é amplamente utilizado no estudo de séries temporais, sendo uma ferramenta essencial para compreender os padrões de dependência temporal nos dados. Segundo Box *et al.* (2015), em sua obra seminal *Time Series Analysis: Forecasting and Control*, a autocorrelação é definida como uma medida estatística que identifica a relação linear entre valores da série temporal em diferentes momentos, denominados lags.

A função ACF permite explorar a estrutura interna da série temporal, revelando não apenas padrões de sazonalidade e persistência, mas também auxiliando na

identificação de componentes autoregressivos (AR) e de média móvel (MA) em modelos como o ARIMA. Como afirmam Cryer e Chan (2008) em *Time Series Analysis: With Applications in R*, o ACF "não apenas mede a força da relação entre os valores da série ao longo do tempo, mas também fornece insights sobre a natureza das flutuações e repetições nos dados".

O procedimento para realizar o teste ACF envolve o cálculo dos coeficientes de autocorrelação para uma sequência de lags. Esses coeficientes variam entre -1 e 1, onde valores próximos a 1 indicam forte correlação positiva, enquanto valores próximos a -1 apontam para uma forte correlação negativa. O gráfico da função ACF, por sua vez, fornece uma representação visual desses coeficientes, com limites de significância que ajudam a determinar se as correlações são estatisticamente relevantes. Sobre isso, Hyndman e Athanasopoulos (2018) destacam, em *Forecasting: Principles and Practice*, que "a análise do ACF é crucial para entender se os resíduos de um modelo ajustado são aleatórios ou possuem estrutura não capturada pelo modelo".

Além disso, o ACF é frequentemente utilizado em conjunto com a Função de Autocorrelação Parcial (PACF), que, diferentemente da ACF, mede a correlação direta entre os valores da série, excluindo os efeitos dos lags intermediários. Essa complementaridade é essencial, conforme reforçam Wei (2006) em *Time Series Analysis: Univariate and Multivariate Methods*: "Enquanto o ACF expõe padrões globais de dependência, o PACF revela dependências diretas, tornando a análise conjunta fundamental para a identificação de modelos de séries temporais".

4.4.6.2 Função de Autocorrelação Parcial (PACF)

O teste PACF é uma ferramenta fundamental no campo da análise de séries temporais, sendo amplamente utilizado para identificar relações diretas entre valores defasados de uma variável. De acordo com Box *et al.* (2015), a PACF mede a autocorrelação parcial, ou seja, a correlação direta entre uma variável e seus lags, eliminando os efeitos das correlações intermediárias.

Enquanto a função ACF mede a correlação total entre os valores defasados, a PACF é projetada para isolar o impacto de um lag específico, controlando a influência de lags anteriores. Cryer e Chan (2008), explicam que "a PACF é particularmente útil para identificar a ordem do componente autoregressivo (AR) em modelos de séries temporais, ao determinar o número de lags que têm influência direta e significativa sobre o valor atual".

O teste PACF é frequentemente representado graficamente, com os coeficientes de autocorrelação parcial no eixo Y e os lags no eixo X. Os valores são avaliados em relação a limites de significância estatística, geralmente representados por linhas pontilhadas. Hyndman e Athanasopoulos (2018), destacam que "o gráfico PACF apresenta cortes distintos nos coeficientes além de certo lag, indicando o ponto em que os efeitos diretos de lags adicionais se tornam irrelevantes para a previsão".

A análise do PACF é essencial no contexto de modelagem de séries temporais, especialmente na identificação da ordem do componente AR em modelos ARIMA. Wei (2006), em Time Series Analysis: Univariate and Multivariate Methods, afirma que "enquanto o ACF é útil para capturar padrões gerais de dependência, a PACF oferece um método mais preciso para determinar a influência específica de cada lag, fornecendo insights detalhados para a construção de modelos autoregressivos".

Além disso, a PACF é frequentemente usada em conjunto com a ACF para distinguir entre componentes AR e MA em uma série temporal. Quando a PACF apresenta um corte claro após um determinado lag, isso sugere a ordem do componente AR. Por outro lado, padrões mais difusos indicam a necessidade de modelar outros componentes, como a média móvel.

4.4.6.3 Dickey-Fuller Aumentado (ADF)

O teste ADF é uma das ferramentas mais amplamente utilizadas na análise de séries temporais para verificar a presença de raiz unitária, ou seja, determinar se uma série temporal é estacionária. Estacionariedade, como destacado por Box *et al.* (2015), é uma propriedade essencial para a modelagem de séries temporais, especialmente em métodos como ARIMA, pois implica que as propriedades estatísticas da série (como

média e variância) são constantes ao longo do tempo.

O teste Dickey-Fuller, desenvolvido por Dickey e Fuller (1979), foi aprimorado para a versão aumentada (ADF) para lidar com possíveis problemas de autocorrelação nos resíduos. Segundo Enders (2015) em *Applied Econometric Time Series*, o teste ADF introduz termos defasados adicionais na equação de teste para capturar essas autocorrelações, tornando a análise mais confiável.

A hipótese nula do teste ADF é que a série temporal possui uma raiz unitária, ou seja, é não estacionária. A hipótese alternativa varia dependendo do modelo especificado, podendo indicar estacionariedade pura ou estacionariedade em torno de uma tendência ou de um nível constante. Cryer e Chan (2008), explicam que "o teste ADF é particularmente valioso porque permite avaliar diferentes formas de estacionariedade, considerando tanto tendências determinísticas quanto componentes estocásticos".

A interpretação do teste baseia-se no valor da estatística-t associado a γ. Um valor estatisticamente significativo rejeita a hipótese nula de não estacionariedade. Como Hyndman e Athanasopoulos (2018) destacam, "a rejeição da hipótese nula indica que a série é estacionária, o que é uma condição necessária para aplicar vários métodos de modelagem preditiva".

O teste ADF desempenha um papel essencial na preparação de dados para modelagem de séries temporais, especialmente ao decidir se a diferenciação (ou outras transformações) é necessária para tornar a série estacionária. Wei (2006), reforça que "o teste ADF fornece uma base objetiva para determinar o nível de diferenciação necessário, eliminando a subjetividade e aumentando a precisão das análises".

Em síntese, o teste ADF é um pilar na análise de séries temporais, permitindo verificar a estacionariedade. Conforme Enders (2015) conclui, "a importância do teste ADF não está apenas em sua aplicação prática, mas também em sua capacidade de fornecer insights sobre a dinâmica fundamental de uma série temporal".

4.5 FERRAMENTAS UTILIZADAS

A escolha do Python para a modelagem PD Forward Looking foi fundamentada por sua flexibilidade e capacidade de atender às demandas da análise de dados e

modelagem preditiva. Reconhecido como uma das principais linguagens para ciência de dados, Python é amplamente adotado em análises financeiras devido à sua eficiência e capacidade de executar cálculos avançados de maneira escalável. Conforme destacado por McKinney (2012), Python oferece uma abordagem prática para manipulação e análise de dados, permitindo que pesquisadores e profissionais desenvolvam modelos reprodutíveis.

No contexto desse estudo, Python foi essencial para a implementação de modelos de regressão linear múltipla, SARIMAX, VAR e Regressão XGBoost, utilizados para capturar diferentes aspectos das variáveis econômicas e financeiras. A clareza e a simplicidade da linguagem, conforme apontado por VanderPlas (2016), facilitaram o desenvolvimento de scripts que garantem a reprodutibilidade dos resultados. Essa característica é particularmente relevante em pesquisas que requerem validação rigorosa e compartilhamento de metodologias.

Com o intuito de garantir a transparência e a reprodutibilidade da pesquisa, os scripts utilizados neste trabalho estão disponíveis publicamente no repositório: https://github.com/douglasl3/pd_fwl. O repositório serve não apenas como uma documentação do processo, mas também como um recurso para futuras análises ou adaptações do modelo.

Assim, a escolha do Python foi motivada não apenas por suas capacidades técnicas, mas também por sua capacidade de atender às melhores práticas em pesquisa e ciência de dados, alinhando-se à necessidade de reprodutibilidade e acessibilidade dos métodos empregados.

5 ANÁLISE DOS RESULTADOS

Nesta seção, serão descritos os resultados e discussões sobre as quatro abordagens explorados. Cada um deles teve alguma particularidade no processo, que serão descritas bem como suas implicações. As análises serão feitas de forma padronizada conforme resíduos e métricas de qualidade descritas na seção anterior. São eles: Regressão linear múltipla, SARIMAX, VAR e Regressão XGBoost.

5.1 REGRESSÃO LINEAR

Foi utilizada uma função com o objetivo de iterar entre as diferentes combinações de variáveis exógenas em busca do menor AIC e com VIF menor que 5 (multicolinearidade dentro do aceitável). Após diversas iterações, o seguinte modelo foi construído:

Figura 1- Resultado Regressão Linear

OLS Regression Results

Dep. Variable:	PD	R-squared:	0.280
Model:	OLS	Adj. R-squared:	0.254
Method:	Least Squares	F-statistic:	10.52
Date:	Sat, 04 Jan 2025	Prob (F-statistic):	0.000139
Time:	08:52:59	Log-Likelihood:	11.587
No. Observations:	57	AIC:	-17.17
Df Residuals:	54	BIC:	-11.04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.5934	0.045	79.360	0.000	3.503	3.684
IPCA_LAG_1	-0.0137	0.005	-2.938	0.005	-0.023	-0.004
IPCA_LAG_3	-0.0134	0.005	-2.641	0.011	-0.024	-0.003

Omnibus:	1.503	Durbin-Watson:	0.596
Prob(Omnibus):	0.472	Jarque-Bera (JB):	1.147
Skew:	-0.347	Prob(JB):	0.564
Kurtosis:	3.001	Cond. No.	18.3

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.1.1 Coeficientes

A um nível de 5%, todas as variáveis são significativas para calcular a PD. A relação entre a elevação do IPCA e a redução PD em uma carteira de crédito pode ser compreendida à luz da literatura econômica e das especificidades dos segmentos que compõem a carteira do Banrisul, predominantemente formada por crédito consignado e crédito rural. Essa hipótese se fundamenta no impacto que a inflação exerce sobre o comportamento econômico dos tomadores de crédito e sobre o ambiente

macroeconômico.

Oliveira e Wolf (2016), destacam que o crédito consignado apresenta um risco de inadimplência muito baixo, dado que o pagamento das parcelas é descontado diretamente da folha de pagamento ou dos benefícios previdenciários, o que reduz significativamente o risco para os bancos. Além disso, a maior parte dos empréstimos consignados é direcionada a servidores públicos e aposentados e pensionistas que recebem via INSS, ou seja, indivíduos com renda previsível, fator que diminui ainda mais os riscos para os bancos.

Por outro lado, o crédito rural, outro segmento predominante na carteira do Banrisul, também apresenta características que permitem uma redução da PD em cenários de aumento da inflação. A inflação, muitas vezes acompanhada por desvalorização cambial, beneficia os produtores rurais, principalmente aqueles voltados para a exportação, ao aumentar a competitividade dos produtos brasileiros no mercado internacional e as receitas em moeda local. Reinhart e Rogoff (2009) destacam que, em economias exportadoras, a inflação combinada com desvalorização cambial pode fortalecer o setor agropecuário, melhorando sua capacidade de honrar compromissos financeiros.

Além disso, o crédito rural no Brasil é amplamente subsidiado, com taxas de juros controladas e abaixo das praticadas no mercado, em função de políticas públicas que visam fomentar a produção agropecuária nacional (IPEA, 2019). Esses subsídios são viabilizados tanto por meio de recursos obrigatórios quanto por programas como o Pronaf e o Pronamp. Ademais, há instrumentos formais de mitigação de risco, como as políticas de renegociação de dívidas autorizadas pelo Conselho Monetário Nacional (CMN), que permitem a prorrogação de parcelas em casos de adversidades climáticas ou dificuldades de comercialização (CMN, 2024). Esse arcabouço institucional reduz significativamente a probabilidade de inadimplência, especialmente entre pequenos e médios produtores. Como destaca Souza (2019), há um histórico de renegociações recorrentes no setor, o que reforça a percepção de menor risco nas operações de crédito rural. Outro ponto relevante é que o aumento do IPCA reflete, muitas vezes, uma elevação nos preços agrícolas, especialmente em períodos de inflação de custos, como aumento no valor dos insumos ou da logística. Esse ajuste nos preços dos

produtos agrícolas, somado a políticas de apoio ao setor, pode fortalecer as finanças dos produtores rurais, reduzindo o risco de inadimplência. Esse fenômeno é reforçado em contextos em que os preços dos alimentos, que compõem uma parcela significativa do IPCA, se elevam e beneficiam diretamente os produtores, aumentando sua liquidez.

A relação entre inflação e inadimplência, no entanto, não é linear e depende do contexto econômico mais amplo. Enquanto a inflação moderada pode ter efeitos benéficos, como os descritos acima, níveis descontrolados de inflação podem prejudicar a confiança econômica e gerar instabilidade financeira. Contudo, no caso da carteira do Banrisul, onde o crédito consignado e rural possuem características estruturais que os tornam menos sensíveis às flutuações adversas, é plausível argumentar, com base na literatura econômica, que a elevação do IPCA pode, sim, estar associada à redução da PD.

5.1.2 Ajuste do Modelo

Conforme apresentado, o valor do R^2 ajustado é de 0.254, indicando que aproximadamente 25,4% da variação na variável dependente (PD) pode ser explicada pelas variáveis independentes selecionadas no modelo, já considerando a penalização pela inclusão de múltiplos preditores.

5.1.3 Multicolinearidade

Conforme observado na tabela, nenhum coeficiente do modelo apresentou um nível de multicolinearidade em nível elevando, conforme a literatura exposta anteriormente.

Tabela 2- Coeficientes Regressão Linear

Variável	VIF
Constante	2,8395
IPCA_LAG_1	1,0721
IPCA_LAG_3	1,0721

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

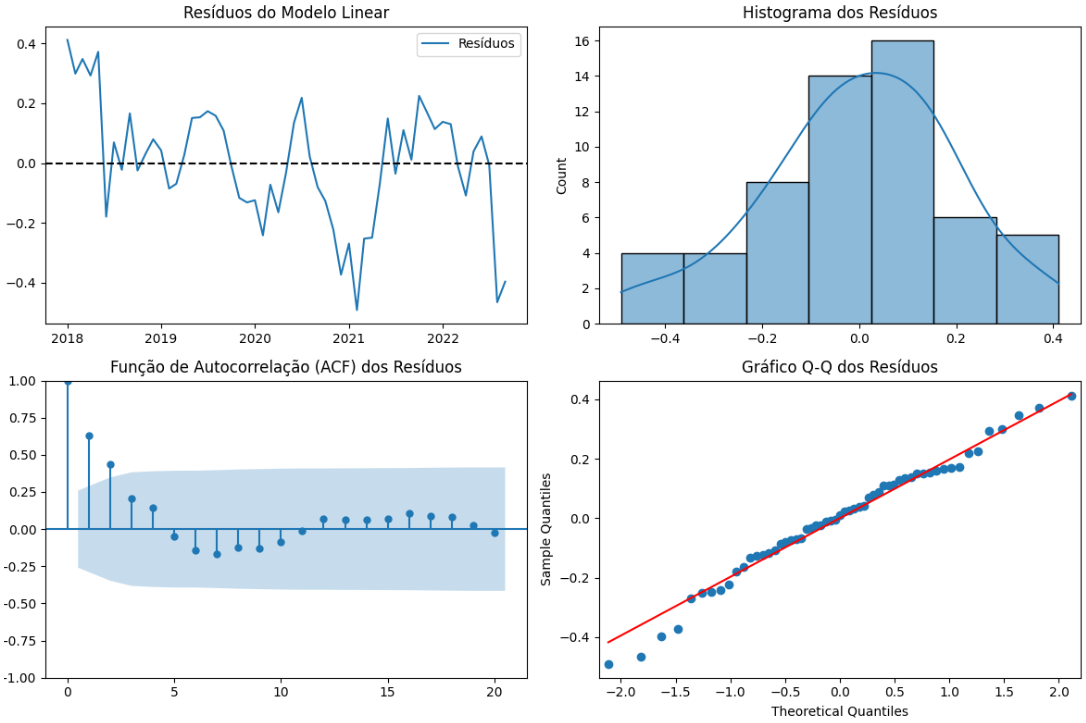
5.1.4 Resíduos

O Teste de Jarque-Bera indicou que os resíduos seguem uma distribuição normal, uma vez que o p-valor de 0,5636 é maior que o nível de significância usual de 0,05. Isso significa que o pressuposto de normalidade foi atendido, permitindo que os intervalos de confiança e os testes de hipótese baseados no modelo sejam confiáveis.

Por outro lado, o Teste de Ljung-Box, aplicado com 10 defasagens, revelou autocorrelação significativa nos resíduos. O p-valor de 0,0000 é extremamente baixo, rejeitando a hipótese nula de ausência de autocorrelação. Este resultado sugere que o modelo não capturou adequadamente a dependência temporal nos dados, o que pode comprometer a validade das inferências estatísticas. É provável que ajustes no modelo sejam necessários, como a inclusão de componentes autorregressivos ou defasados.

Já o Teste de Breusch-Pagan não apontou problemas de heterocedasticidade, com p-valores superiores a 0,05 tanto para a estatística LM (0,1412) quanto para a estatística F (0,1464). Isso confirma que os resíduos possuem variância constante, satisfazendo o pressuposto de homocedasticidade. Como resultado, as estimativas dos parâmetros do modelo são eficientes e os erros padrão são confiáveis.

Gráfico 1 - Resíduos Regressão Linear



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.1.4.1 Backtest do modelo

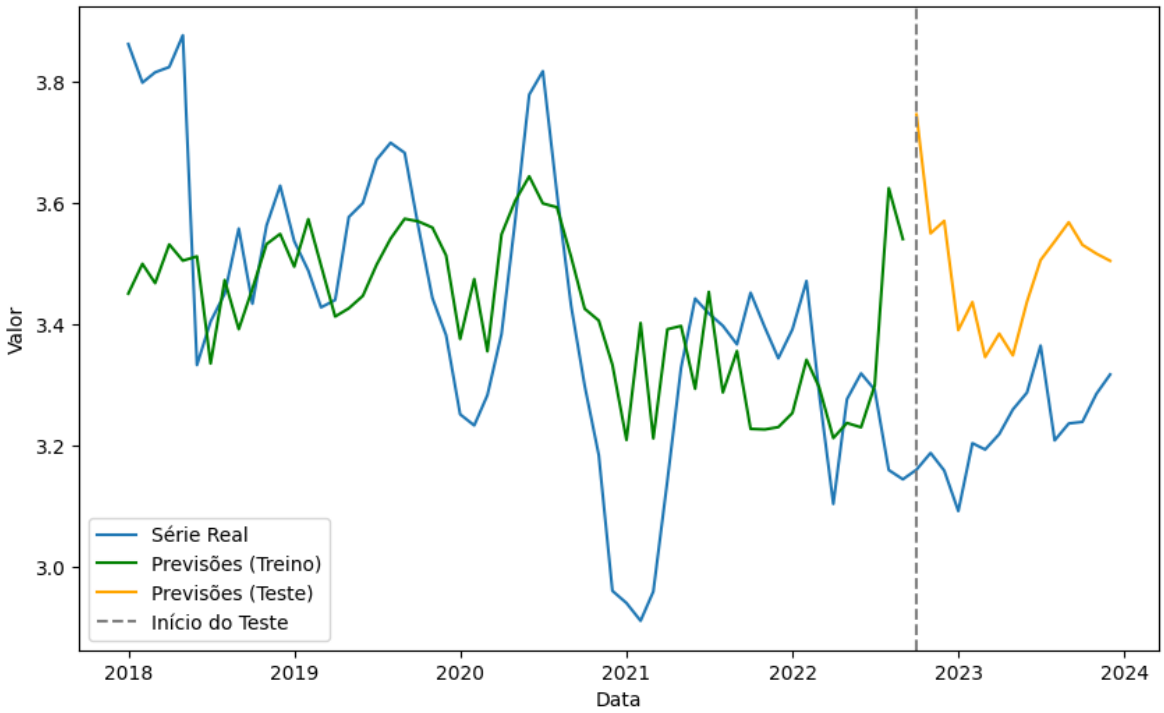
Para avaliar a performance dos modelos de forma mais completa, foi analisada a correspondência entre os valores previstos pelo modelo e os valores reais em dois intervalos de tempo distintos:

a) desempenho do modelo durante o período de desenvolvimento (de janeiro de 2018 a setembro de 2022);

b) desempenho do modelo durante o período de validação (de outubro de 2022 a dezembro de 2023).

O gráfico a seguir apresenta os resultados obtidos para o modelo final de regressão.

Gráfico 2 - Backtest Regressão Linear



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

Durante o período de treino, o modelo apresentou um RMSE de 0,1975, e um MAPE de 4,60%. Esses valores indicam que o modelo foi capaz de capturar as tendências da série e produzir previsões com erros baixos e precisos, mantendo uma diferença média de apenas 4,60% entre os valores previstos e os observados. Esse desempenho demonstra que, no conjunto de dados utilizado para desenvolvimento, o modelo conseguiu se ajustar bem.

No entanto, ao ser testado no período de validação, o modelo apresentou um desempenho inferior. O RMSE aumentou para 0,2922, e o MAPE ficou em 8,23%. Apesar de o modelo ainda conseguir capturar as tendências gerais, o erro percentual médio quase dobrou em relação ao período de treino. Isso indica que o modelo pode ter dificuldade em se adaptar a novas condições ou variações que não estavam completamente representadas no período de desenvolvimento.

O R^2 ajustado de 0,254 indica que parte expressiva da variação da PD permanece sem explicação, possivelmente devido à ausência de estrutura temporal no modelo. Essa fragilidade também foi evidenciada pelo Teste de Ljung-Box, que apontou autocorrelação significativa nos resíduos, reforçando a necessidade de incorporar a

dependência temporal, conforme sugerem Stock e Watson (2019). Tais limitações são recorrentes na literatura, como observado por Simões (2022), que também utilizou regressões lineares para modelagem de inadimplência e relatou dificuldades em capturar adequadamente a dinâmica da série, recomendando a adoção de modelos que incorporam componentes autorregressivos.

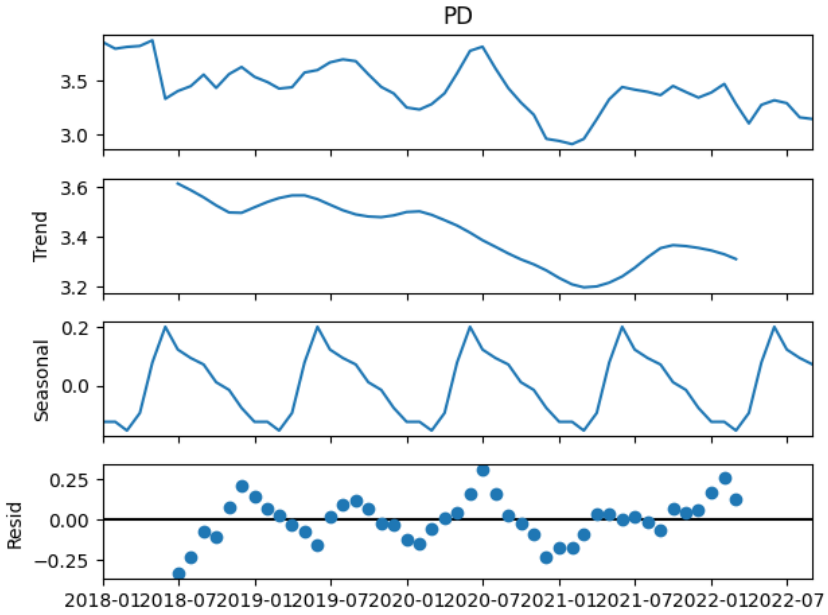
O gráfico reforça essa interpretação, evidenciando que o modelo se ajusta bem às flutuações mais suaves da série, mas apresenta algumas discrepâncias em picos ou mudanças bruscas, especialmente no período de validação. Esses desvios podem ser reflexo de características específicas da série que não foram plenamente capturadas no processo de modelagem.

5.2 SARIMAX

Para construir um modelo de séries temporais, é preciso antes identificar se a série é estacionária, através do teste de Dickey-Fuller e também suas ordens autorregresivas, médias móveis e sazonalidade a partir dos gráficos de Autocorrelação (ACF) e Autocorrelação Parcial (PACF).

A série foi decomposta conforme segue:

Gráfico 3 - Decomposição Série PD



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

A análise da série temporal descreve uma história técnica e estruturada sobre os seus comportamentos subjacentes. A trajetória da série é composta por três elementos principais: a tendência, a sazonalidade e os resíduos, cada um desempenhando um papel na trajetória da PD.

A tendência nos revela o movimento de longo prazo da PD. No início, há um declínio suave, como se o mercado estivesse se ajustando a condições mais favoráveis. Em seguida, surge uma recuperação gradual, com a tendência subindo lentamente, refletindo mudanças que podem ser estruturais ou econômicas. Contudo, ao final do período, nota-se um leve declínio, indicando que fatores recentes podem estar pressionando a estabilidade da série.

A sazonalidade, por sua vez, traz à tona padrões cíclicos claros e consistentes. Esses ciclos regulares, sugerem que a PD é influenciada por fatores previsíveis que se repetem ao longo do tempo. Seja pela sazonalidade do mercado de crédito ou por padrões de comportamento financeiro dos tomadores, essa componente nos ajuda a entender as flutuações recorrentes na série.

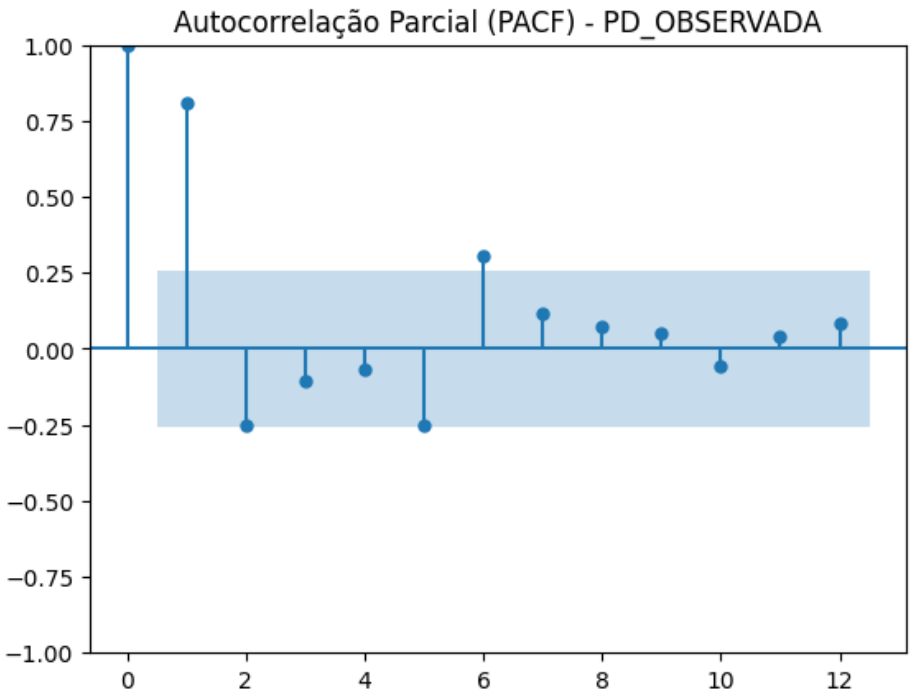
Por fim, os resíduos representam os desvios da trajetória principal. Durante grande parte do período, eles permanecem próximos de zero, indicando que o modelo capturou bem os padrões da série. No entanto, ao final, os resíduos mostram maior volatilidade, sugerindo que eventos inesperados ou novos fatores começaram a impactar a dinâmica da PD.

Ao aplicar o teste Dickey-Fuller, obtivemos uma estatística de -2.9072 e um p-valor de 0,0445, indicando que a série é estacionária e não necessita de diferenciação. No entanto, o teste detecta estacionariedade estatística com base na hipótese nula de presença de raiz unitária. Contudo, mesmo com um p-valor significativo (como no caso, p=0.044), a estacionariedade observada pode não ser suficiente para eliminar totalmente tendências ou flutuações de longo prazo que influenciam a modelagem. A inclusão de uma diferenciação pode ser uma escolha para garantir uma série completamente estacionária em todos os sentidos, especialmente em dados mais complexos.

Para estimar sa ordens do ARIMA, foi criada uma função que itera sobre uma lista de valores e escolhe o ajuste com o menor AIC. Os testes de autocorrelação

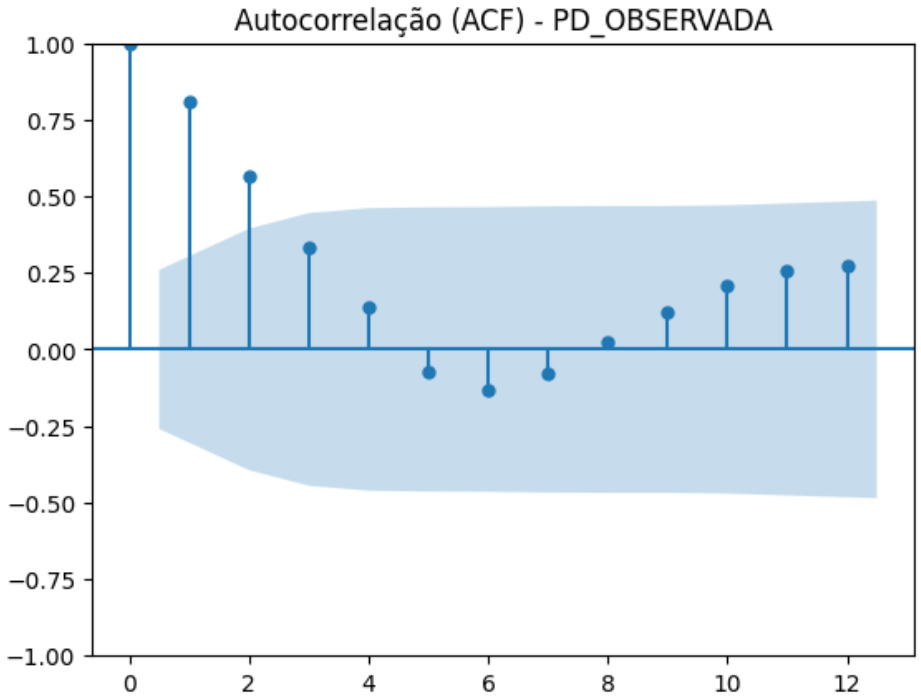
indicaram a seguinte distribuição:

Gráfico 4 - Autocorrelação Parcial (PACF) da PD



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

Gráfico 5 - Autocorrelação (ACF) da PD



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

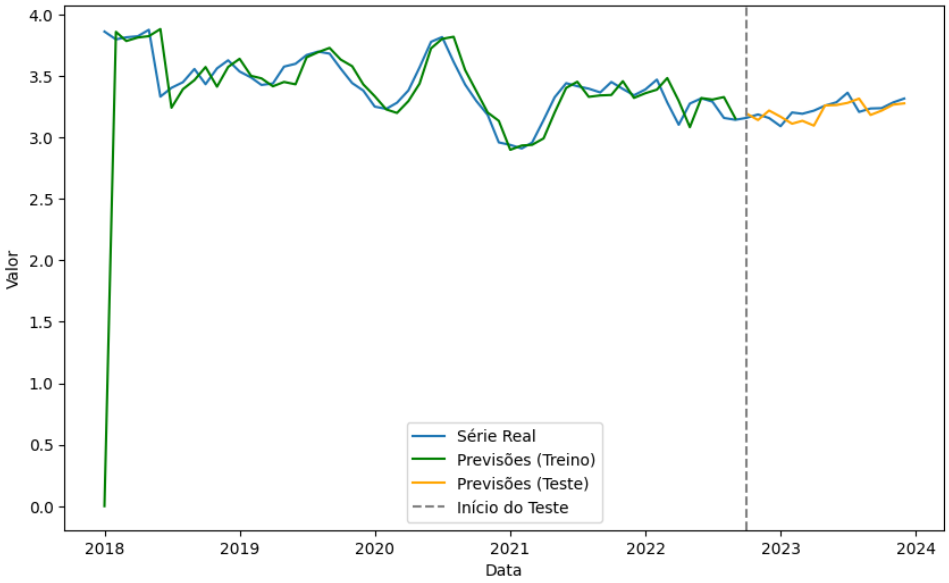
O modelo ARIMA(0,1,0)(2,0,0)[3] foi configurado com base nos padrões observados nos gráficos de ACF e PACF, destacando características específicas da série.

Na parte não sazonal, o gráfico da ACF apresenta um decaimento gradual, indicando a presença de uma tendência que foi corrigida com uma diferenciação simples (d = 1). Por outro lado, o gráfico da PACF não exibe lags significativos, justificando a escolha de p = 0, ou seja, a ausência de termos autoregressivos.

Na parte sazonal, o gráfico da PACF mostra dois lags significativos no intervalo correspondente ao período sazonal (multiplicado por 3, considerando a periodicidade definida), o que motivou a inclusão de dois termos autoregressivos sazonais (P = 2). Já no gráfico da ACF, o rápido decaimento dos lags sazonais indica que não há necessidade de incluir termos de média móvel sazonal, o que valida a escolha de Q = 0. Além disso, a ausência de padrões sazonais persistentes nos resíduos explica a decisão de não aplicar diferenciação sazonal (D = 0).

Ao aplicar as ordens do ARIMA e fazer as projeções, temos o seguinte resultado:

Gráfico 6 - Backtest Modelo SARIMA



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

É possível observar que a série sozinha, com seus componentes sazonais, autorregressivos e de médias móveis, consegue capturar de forma adequada a PD, tanto no período de desenvolvimento quanto no de teste. No entanto, como precisamos incorporar variáveis macroeconômicas, o faremos a partir do ARIMA modelado.

Assim como no modelo de regressão linear, foi construída uma função com o objetivo de iterar entre as diferentes combinações de variáveis exógenas em busca do menor AIC e com VIF menor que 5 (multicolinearidade dentro do aceitável). Após diversas iterações, o seguinte modelo foi construído:

Figura 2 - Resultado SARIMAX

SARIMAX Results

Dep. Variable:	PD	No. Observations:	57
Model:	SARIMAX(0, 1, 0)x(2, 0, 0, 3)	Log Likelihood	42.641
Date:	Sun, 17 Nov 2024	AIC	-77.283
Time:	18:13:40	BIC	-69.182
Sample:	01-01-2018	HQIC	-74.142
	- 09-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
IPCA_LAG_6	-0.0033	0.003	-1.038	0.299	-0.010	0.003
ar.S.L3	-0.1793	0.172	-1.040	0.298	-0.517	0.159
ar.S.L6	-0.5563	0.094	-5.903	0.000	-0.741	-0.372
sigma2	0.0123	0.002	5.601	0.000	0.008	0.017

Ljung-Box (L1) (Q):	0.30	Jarque-Bera (JB):	22.04
Prob(Q):	0.58	Prob(JB):	0.00
Heteroskedasticity (H):	0.48	Skew:	-0.78
Prob(H) (two-sided):	0.12	Kurtosis:	5.65

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.2.1 Coeficientes

Os coeficientes estimados oferecem algumas pistas importantes. A variável IPCA_LAG_6, que representa o IPCA defasado em seis períodos, apresenta um coeficiente de -0.0033, indicando que, teoricamente, um aumento no IPCA defasado reduz ligeiramente a PD. No entanto, o p-valor associado (p=0.299) mostra que essa relação não é estatisticamente significativa. Em outras palavras, o impacto do IPCA defasado sobre a PD não pode ser considerado confiável no contexto deste modelo.

Por outro lado, os termos sazonais fornecem uma visão mais ampla. O coeficiente para o componente autorregressivo sazonal de seis períodos (ar.S.L6) é -0.5563, com um p-valor muito baixo (p < 0.001), o que indica uma forte influência sazonal na PD.

Esses achados estão em consonância com os resultados obtidos por Simões (2022), que também empregou modelos com componentes autorregressivos para explicar a inadimplência bancária e identificou a estrutura temporal como fator determinante na modelagem preditiva.

5.2.2 Ajuste do Modelo

No que diz respeito à qualidade do ajuste, o modelo apresenta um R^2 ajustado de 0.726, indicando que cerca de 72% da variação na PD é explicada pelas variáveis incluídas no modelo. Esses valores são elevados para um modelo econômico, sugerindo que ele captura bem as dinâmicas da série temporal, mesmo considerando a penalização pela inclusão de múltiplos preditores.

5.2.3 Multicolinearidade

No contexto do modelo apresentado, o teste de multicolinearidade, como o cálculo do VIF (Variance Inflation Factor), não se aplica devido à ausência de múltiplas variáveis explicativas fixas no modelo. A multicolinearidade é um problema que ocorre quando duas ou mais variáveis independentes estão altamente correlacionadas, resultando em dificuldades para estimar com precisão os coeficientes individuais de cada variável. No entanto, o modelo SARIMAX utilizado conta com apenas uma variável exógena fixa, o IPCA_LAG_6, além dos componentes sazonais e autorregressivos, que não entram no cálculo da multicolinearidade de forma tradicional.

A lógica por trás do VIF é avaliar o grau de correlação entre as variáveis explicativas no modelo. Em modelos com uma única variável explicativa fixa, essa correlação simplesmente não pode existir, já que não há outra variável com a qual a primeira possa interagir. Como resultado, a ideia de "colinearidade" entre variáveis perde seu significado.

Além disso, a ausência de múltiplas variáveis fixas elimina o risco de inflar a variância dos coeficientes devido à redundância entre variáveis explicativas, que é o principal objetivo do teste de multicolinearidade.

5.3.3 Resíduos

Os resíduos do modelo, visualizados por meio dos gráficos e testes estatísticos, fornecem informações valiosas sobre o ajuste e as suposições do modelo. O gráfico dos resíduos ao longo do tempo mostra que eles estão centrados em torno de zero e

não apresentam padrões aparentes, indicando que o modelo capturou bem a dinâmica da série temporal.

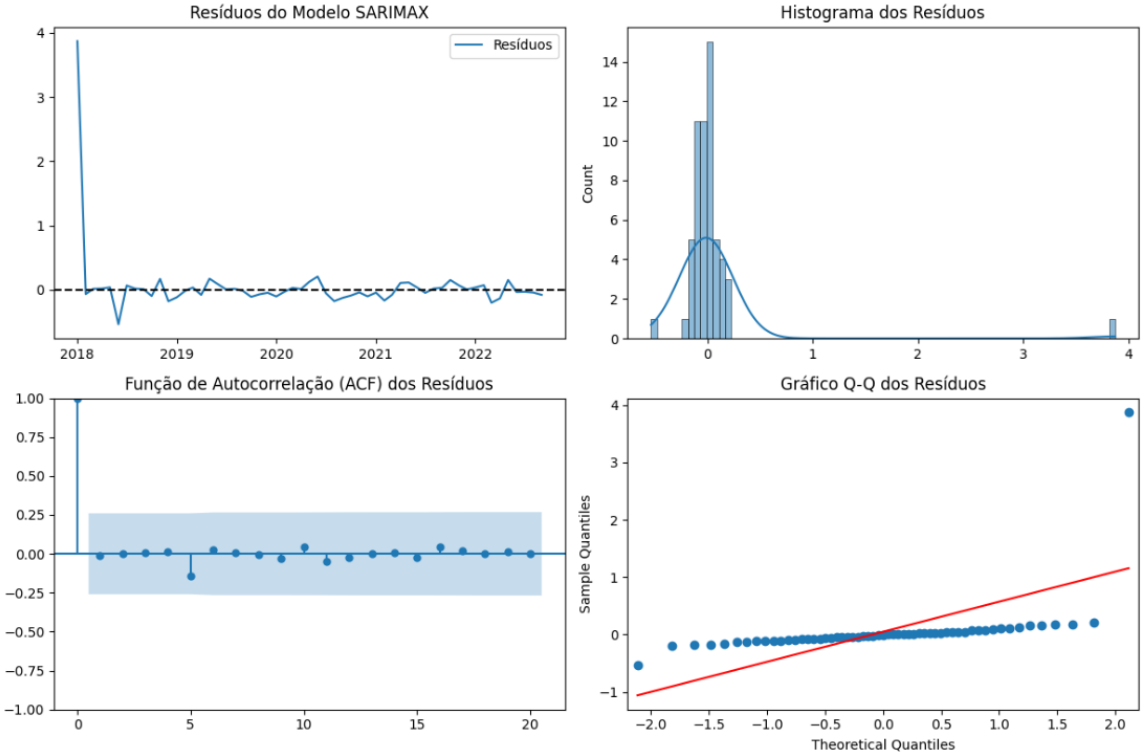
O histograma dos resíduos e o gráfico Q-Q sugerem uma violação da normalidade. Isso é corroborado pelo teste de Jarque-Bera, cujo p-valor (p = 0.0000) confirma que os resíduos não seguem uma distribuição normal. Embora isso possa ser uma limitação para algumas análises estatísticas, a normalidade dos resíduos não é um requisito estrito para prever com eficácia em modelos de séries temporais.

O teste de Ljung-Box, com p-valor elevado (p = 0.9985), indica que não há autocorrelação significativa nos resíduos, o que é desejável, pois sugere que o modelo capturou bem as dependências temporais da série.

Por fim, o teste de Breusch-Pagan aponta para homocedasticidade nos resíduos (p = 0.5906), indicando que a variância dos resíduos é constante ao longo do tempo.

Em resumo, os resíduos não apresentam autocorrelação nem heterocedasticidade, sugerindo um bom ajuste do modelo. No entanto, a falta de normalidade nos resíduos deve ser considerada, especialmente se análises posteriores dependerem de suposições estritas de normalidade.

Gráfico 7 - Resíduos SARIMAX

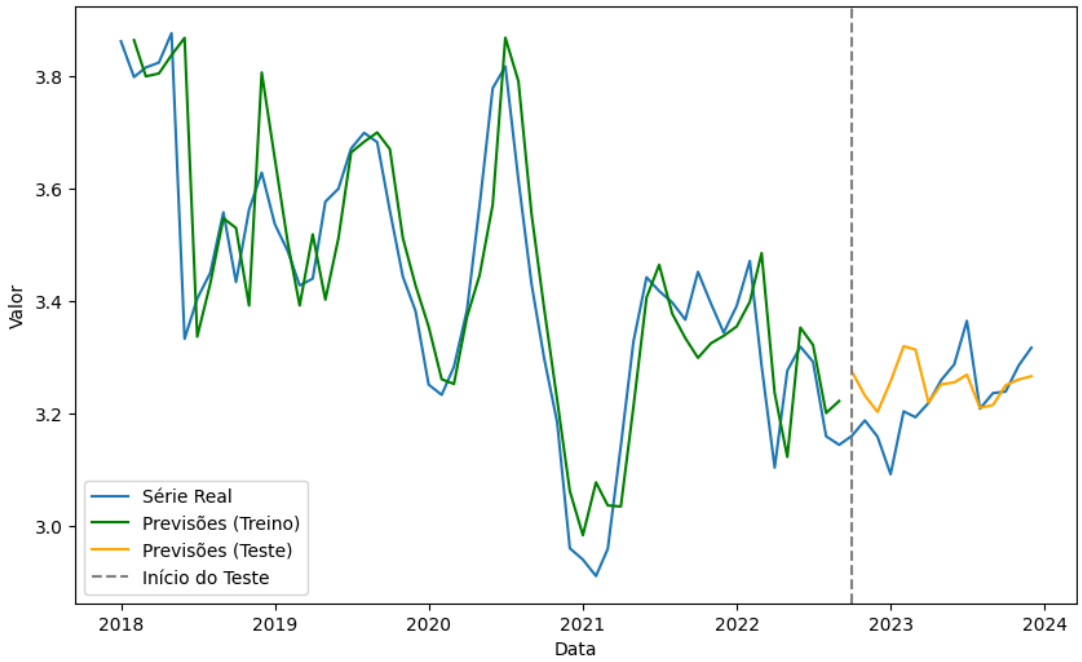


Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.3.4 Backtest do modelo

O gráfico a seguir apresenta os resultados obtidos para o modelo SARIMAX.

Gráfico 8 - Backtest Modelo SARIMAX



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

Durante o período de treino, o modelo mostrou alta precisão, com um RMSE de 0.1179 e um MAPE de 2,50%, indicando que os erros absolutos e percentuais eram baixos. O gráfico reflete isso, mostrando que as previsões do treino (linha verde) seguem de forma bastante alinhada à série real (linha azul), capturando bem as tendências e sazonalidades do período.

No período de teste, os resultados foram sensivelmente melhores do que em treino, com o RMSE 0.0666 e o MAPE caindo para 2,07%, indicando erros menores. No gráfico, as previsões do teste (linha laranja) começam a se distanciar mais da série real (linha azul), especialmente em alguns picos e vales, sugerindo que o modelo não conseguiu captar totalmente as flutuações ou eventos inesperados ocorridos nesse período. No entanto, as previsões ainda mantêm um alinhamento geral com a tendência da série real, o que demonstra a utilidade do modelo para capturar padrões gerais, embora com limitações em casos mais específicos.

Em resumo, o modelo apresentou um ajuste muito bom no em treino e teste, como mostrado pelo gráfico e métricas. No entanto, esses resultados apontam para a necessidade de ajustes, como a inclusão de variáveis exógenas adicionais ou refinamentos na especificação sazonal do SARIMAX, visando melhorar a precisão das

previsões, especialmente em novos dados.

5.3 VAR

Os modelos de Vetores Autorregressivos (VAR) são amplamente utilizados na análise de séries temporais multivariadas para capturar as relações dinâmicas entre variáveis ao longo do tempo. No entanto, um pressuposto fundamental para sua aplicação é que todas as séries incluídas no modelo sejam estacionárias. Isso significa que suas propriedades estatísticas, como média, variância e autocorrelação, devem ser constantes ao longo do tempo. Essa exigência garante que os parâmetros estimados sejam estáveis e que as inferências realizadas sejam válidas.

A não estacionariedade das séries pode comprometer significativamente os resultados de um modelo VAR. Segundo Hamilton (1994), séries não estacionárias podem levar a estimativas inconsistentes e dificultar a interpretação das relações entre as variáveis. Esse problema ocorre porque tendências ou volatilidades crescentes ao longo do tempo podem induzir correlações espúrias, gerando resultados estatisticamente inválidos. Lutkepohl (2005) reforça que a presença de séries não estacionárias no VAR pode tornar os coeficientes estimados instáveis, prejudicando a previsão e a análise de impacto entre variáveis.

Para lidar com séries não estacionárias, a transformação dos dados é uma etapa essencial. Enders (2015) destaca que, em casos como esses, a diferenciação das séries pode ser uma solução eficaz, pois elimina tendências e garante a estacionariedade necessária para o modelo. Essa abordagem, apesar de reduzir a interpretação de longo prazo, permite que o VAR capture com precisão as dinâmicas de curto prazo entre as variáveis.

Além disso, quando as séries possuem uma relação de equilíbrio de longo prazo, como no caso de cointegração, o modelo VAR pode não ser o mais adequado. Johansen (1995) aponta que, nesses casos, o uso do Modelo de Correção de Erros Vetoriais (VECM) é mais apropriado, pois ele combina a análise de longo prazo com a estacionariedade das diferenças, garantindo resultados mais aderentes.

A escolha de utilizar o teste Dickey-Fuller Aumentado com seleção automática do número de lags baseada no Critério de Informação de Akaike (AIC) encontra respaldo

na literatura, especialmente em análises multivariadas. Em séries temporais, o número de lags desempenha um papel fundamental na captura de dinâmicas importantes entre as observações. Uma escolha inadequada de defasagens pode levar a erros de especificação, comprometendo a validade dos resultados. Por outro lado, incluir defasagens excessivas reduz o poder do teste, tornando a análise menos eficiente.

No contexto de modelos multivariados, como o VAR, a escolha do número de lags é ainda mais crítica. Esses modelos analisam relações dinâmicas entre variáveis ao longo do tempo, e os efeitos defasados são essenciais para a qualidade das estimativas. Como apontado por Lütkepohl (2005), uma seleção apropriada de lags garante que as interações entre as séries sejam capturadas com precisão, evitando problemas de sub ou superestimação dos parâmetros.

Essa prática é especialmente relevante no teste ADF, pois as propriedades da série podem variar consideravelmente. Com o AIC, é possível ajustar automaticamente o número de lags ao comportamento dinâmico da série, sem necessidade de seleção manual, o que reduz a possibilidade de vieses ou inconsistências. Hamilton (1994) reforça que uma seleção inadequada do número de lags pode comprometer a validade de análises, especialmente em contextos multivariados, onde as relações entre as variáveis são sensíveis à escolha de defasagens.

Após a execução do teste de estacionariedade, ficamos com as seguintes variáveis: PD, IPCA_LAG_1, IPCA_LAG_3, IPCA_LAG_6, IPCA_LAG_12, DESEMPREGO_LAG_12, DIFF_SELIC_LAG_12, DIFF_CAMBIO_LAG_1, DIFF_CAMBIO_LAG_3, DIFF_CAMBIO_LAG_6 e DIFF_CAMBIO_LAG_12.

O modelo foi construído através de uma função que realiza uma busca sistemática pela melhor configuração de um modelo VAR, avaliando diferentes combinações de variáveis exógenas e números de defasagens. O objetivo é identificar o modelo que minimiza o RMSE na previsão da PD.

O processo envolve testar combinações de variáveis em diferentes tamanhos de subconjuntos, ajustando modelos VAR para cada uma delas e utilizando o critério de informação AIC para determinar o número ideal de lags. A função avalia a capacidade preditiva de cada modelo e retorna a melhor combinação de variáveis, o menor RMSE encontrado e o número de lags correspondente. Após diversas iterações, o seguinte

modelo foi construído:

Figura 3 - Resultado VAR

Results for equation PD

	coefficient	std. error	t-stat	prob
const	0.394713	0.323805	1.219	0.223
L1.DESEMPREGO_LAG_12	0.017295	0.016816	1.029	0.304
L1.DIFF_CAMBIO_LAG_1	0.141417	0.086378	1.637	0.102
L1.DIFF_CAMBIO_LAG_3	0.040751	0.094470	0.431	0.666
L1.PD	0.814005	0.077353	10.523	0.000

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.3.1 Coeficientes

Na equação para PD, a constante apresentou um coeficiente positivo, mas não significativo, indicando que seu impacto na explicação da variável pode ser irrelevante. A variável DESMPREGO_LAG_12, que representa a defasagem de 12 períodos da taxa de desemprego, também não foi significativa, sugerindo que essa variável tem pouca influência direta sobre PD nesse horizonte de tempo.

Por outro lado, a variável DIFF_CAMBIO_LAG_1, que reflete a variação cambial defasada em um período, apresentou um coeficiente marginalmente significativo, apontando para uma possível influência moderada no comportamento de PD. Já a defasagem de três períodos da mesma variável (DIFF_CAMBIO_LAG_3) não foi estatisticamente significativa, indicando uma menor relevância no modelo.

O destaque está na própria variável PD, cuja defasagem de um período (L1.PD) apresentou um coeficiente altamente significativo. Isso evidencia uma forte dependência da série em relação aos seus próprios valores passados, caracterizando um comportamento autocorrelacionado que é típico em muitas séries temporais. Tais achados são consistentes com a literatura, que reconhece a sensibilidade de variáveis macroeconômicas à escolha do horizonte de defasagem em modelos VAR (Lütkepohl, 2005; Kilian; Lütkepohl, 2017).

5.3.2 Ajuste do Modelo

O R^2 ajustado para a variável PD foi de 0.7022, indicando que cerca de 70,22% da variabilidade de PD é explicada pelo modelo, mesmo considerando a penalização pela inclusão de variáveis explicativas. Esse resultado demonstra que o modelo VAR possui uma boa capacidade de capturar a dinâmica da série de PD, especialmente considerando que o R^2 ajustado corrige eventuais ganhos artificiais de explicação devido à adição de variáveis irrelevantes.

5.3.3 Multicolinearidade

A multicolinearidade pode dificultar a interpretação dos coeficientes individuais, tornando-os instáveis ou até inflacionando sua variância. No entanto, no VAR, esses coeficientes não são o foco central da análise. Como apontado por Lütkepohl (2005), os modelos VAR são mais utilizados para examinar funções de impulso-resposta, decomposições de variância e previsões conjuntas das variáveis do sistema. Esses resultados dinâmicos são derivados das interações entre as variáveis ao longo do tempo, e não dependem exclusivamente da precisão de um único coeficiente. Portanto, mesmo na presença de multicolinearidade, o modelo pode gerar previsões confiáveis e insights úteis sobre a relação entre as variáveis.

Além disso, a presença de multicolinearidade não compromete a consistência das estimativas em um VAR, embora possa aumentar a variância dos coeficientes. Essa característica permite que o modelo continue sendo uma ferramenta eficaz para prever séries temporais interdependentes, mesmo quando as variáveis do sistema apresentam alguma colinearidade.

5.3.4 Resíduos

A análise dos resíduos do modelo indica que o ajuste capturou adequadamente a dinâmica temporal da série, mas apresenta algumas limitações em relação à distribuição dos resíduos. A série dos resíduos mostra oscilações em torno de zero, sem padrões evidentes ou tendências, o que é um indicativo de que o modelo conseguiu retirar boa parte da estrutura temporal da série original. Isso é reforçado pela análise da ACF, que revela que a maior parte dos lags está dentro dos limites de

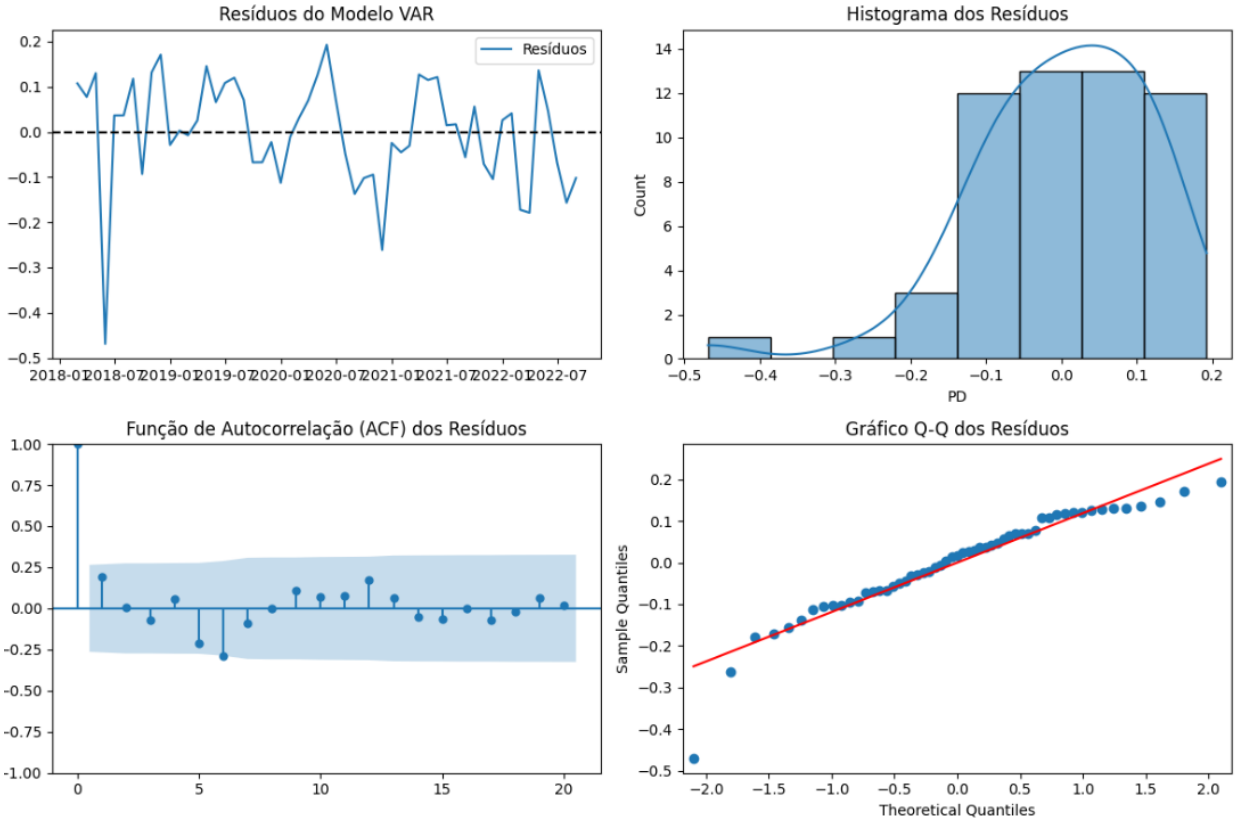
significância, com exceção do primeiro lag, sugerindo que os resíduos não possuem autocorrelação significativa.

No entanto, ao avaliar a distribuição dos resíduos, o histograma sugere uma leve simetria, mas com desvios da normalidade, conforme também observado no gráfico Q-Q, que revela discrepâncias nas caudas da distribuição. Esse resultado é confirmado pelo teste de Jarque-Bera, que rejeita a hipótese de normalidade com um p-valor de 0.0000. Isso indica que os resíduos não seguem uma distribuição normal, o que pode limitar algumas análises estatísticas, embora não comprometa necessariamente a qualidade preditiva do modelo.

Por outro lado, o teste de Ljung-Box, confirma que não há autocorrelação significativa nos resíduos, com um p-valor de 0.2371. Isso reforça que o modelo conseguiu capturar adequadamente as relações temporais na série, eliminando padrões que poderiam comprometer sua eficácia.

Nos modelos VAR, o teste de heterocedasticidade não é essencial porque o foco está na análise das interações dinâmicas entre variáveis endógenas e na previsão, não na interpretação precisa dos coeficientes individuais. Enders (2015) destaca que, embora a heterocedasticidade possa reduzir a eficiência das estimativas, ela não compromete a consistência dos coeficientes obtidos por Mínimos Quadrados Ordinários (OLS). Além disso, estatísticos como Sims (1980), ao introduzirem o uso dos VARs na macroeconomia empírica, enfatizaram que sua utilidade reside mais na previsão e na análise de impulso-resposta do que na interpretação direta dos coeficientes.

Gráfico 9 - Resíduos VAR

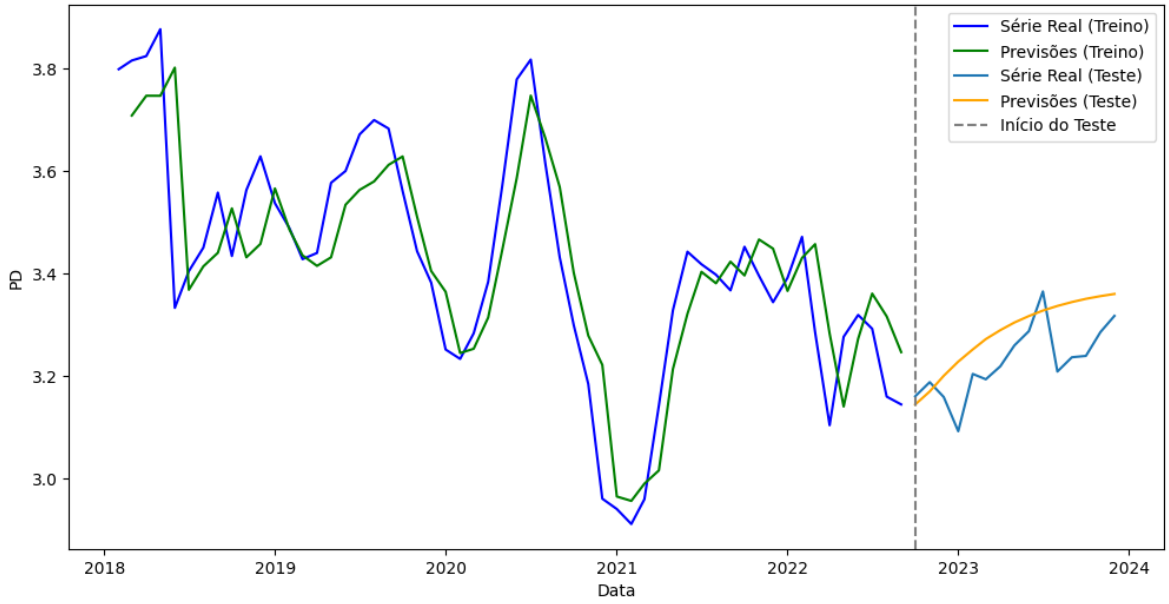


Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.3.5 Backtest do modelo

O gráfico a seguir apresenta os resultados obtidos para o modelo VAR.

Gráfico 10 - Backtest Modelo VAR



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

O gráfico mostra a comparação entre os valores reais e previstos para a variável PD nos períodos de treino e teste. Durante o período de treino, as previsões (linha verde) seguem de forma muito próxima os valores reais (linha azul), indicando que o modelo conseguiu capturar bem a dinâmica histórica da série. Após o início do período de teste, representado pela linha tracejada, as previsões (linha laranja) também acompanham bem a tendência dos valores reais (linha azul-claro), embora com pequenas variações, demonstrando a capacidade do modelo de generalizar para dados fora da amostra.

No conjunto de treino, o RMSE foi de 0.1186, enquanto o MAPE, que mede o erro percentual médio, foi de apenas 2,72%, indicando que o modelo é preciso na reprodução da série histórica. No conjunto de teste, os resultados são melhores, com um RMSE de 0.0756 e um MAPE de 2,03%, o que demonstra que o modelo conseguiu prever novos dados com erros ainda menores.

5.4 REGRESSÃO XGBOOST

Para a criação do modelo, foram utilizadas todas as variáveis disponíveis no conjunto de dados, assim como suas defasagens, permitindo capturar tanto os efeitos

contemporâneos quanto os efeitos temporais dessas variáveis sobre a variável-alvo.

Além disso, os dados foram normalizados antes do treinamento. A normalização transforma os valores das variáveis para que tenham média zero e desvio padrão unitário, garantindo que todas as variáveis fiquem na mesma escala. Esse processo é especialmente relevante para modelos como o XGBoost, pois ajuda a evitar que variáveis com valores maiores dominem o treinamento, o que poderia levar a uma atribuição de pesos inadequada.

Para construção do modelo, foi utilizada uma função que busca otimizar o desempenho do modelo XGBoost ajustando seus parâmetros por meio de uma técnica de busca aleatória. Essa etapa é essencial porque os parâmetros controlam como o modelo aprende e se adapta aos dados. Parâmetros bem ajustados ajudam o modelo a encontrar o equilíbrio ideal entre capturar padrões complexos e evitar a memorização excessiva dos dados, garantindo que ele funcione bem tanto no conjunto de treino quanto no de teste.

O ajuste considera aspectos como a profundidade das árvores, o número de iterações do modelo, a taxa com que ele aprende e a proporção de dados e variáveis usadas em cada etapa do treinamento. Esses elementos são importantes porque determinam o quanto o modelo consegue capturar variações nos dados e como ele lida com padrões complexos. Além disso, eles ajudam a evitar problemas como overfitting, que ocorre quando o modelo se ajusta demais aos dados de treino e perde desempenho em novos dados.

A busca aleatória permite testar diversas combinações desses parâmetros de forma eficiente, sem necessidade de avaliar todas as possibilidades, o que seria computacionalmente caro.

5.4.1 Coeficientes

A análise de importância das variáveis no modelo XGBoost revela quais fatores e suas defasagens foram mais relevantes para a previsão da variável-alvo. Dentre as variáveis disponíveis, destaca-se PIB_LAG_3, que apresentou a maior contribuição, com um valor de importância de 0.4836. Isso indica que o comportamento do PIB três períodos atrás é o principal fator influenciando a variável-alvo, o que está em

consonância com estudos como o de Xu, Li e Wu (2024), desenvolveram um modelo de previsão de risco financeiro sistêmico utilizando o algoritmo XGBoost e destacaram a relevância de variáveis macroeconômicas defasadas, como o crescimento do PIB, na previsão de riscos financeiros.

Outras variáveis que também se mostraram relevantes incluem CAMBIO_LAG_6 (0.0804) e SELIC_LAG_12 (0.0707), evidenciando que o câmbio com defasagem de seis períodos e a taxa SELIC com defasagem de doze períodos desempenham papéis importantes. Esses resultados refletem a sensibilidade do modelo a fatores macroeconômicos tanto de curto quanto de médio prazo — padrão também observado em Duan *et al.* (2022), que demonstraram a eficácia de modelos de aprendizado de máquina, como o XGBoost, na previsão de riscos econômicos mesmo em cenários altamente voláteis, reforçando sua aplicabilidade para contextos de instabilidade financeira. Por outro lado, variáveis como IPCA_LAG_12 (0.0060) e PIB_LAG_6 (0.0073) apresentaram contribuições mínimas, indicando que seus efeitos são praticamente irrelevantes no contexto deste modelo. De maneira similar, outras defasagens do IPCA, como IPCA_LAG_3 e IPCA_LAG_6, também tiveram baixa importância, sugerindo que o comportamento da inflação possui um impacto limitado na previsão da variável-alvo.

Esses resultados apontam para uma predominância de variáveis relacionadas ao PIB, câmbio e SELIC como os principais determinantes no modelo, com uma preferência por defasagens intermediárias e longas. Isso reflete uma relação temporal mais prolongada, onde os efeitos passados dessas variáveis desempenham um papel significativo na previsão.

Tabela 3- Importância Variáveis

Variável	Importância
PIB_LAG_3	0,483599
CAMBIO_LAG_6	0,080434
SELIC_LAG_12	0,070745
PIB_LAG_1	0,053152
SELIC_LAG_1	0,040424
CAMBIO_LAG_12	0,037428
SELIC_LAG_3	0,031082
CAMBIO_LAG_1	0,029239
DESEMPREGO_LAG_12	0,026037
CAMBIO_LAG_3	0,024273
SELIC_LAG_6	0,024192
DESEMPREGO_LAG_1	0,017246
DESEMPREGO_LAG_6	0,015083
IPCA_LAG_1	0,01385
PIB_LAG_12	0,012752
DESEMPREGO_LAG_3	0,01228
IPCA_LAG_3	0,007503
IPCA_LAG_6	0,007441
PIB_LAG_6	0,007263
IPCA_LAG_12	0,005976

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.4.2 Ajuste do Modelo

O R^2 ajustado do modelo foi de 0.7558, o que significa que aproximadamente 75,58% da variabilidade da variável-alvo no conjunto de treino é explicada pelas variáveis incluídas no modelo, levando em consideração a penalização pela quantidade de variáveis explicativas utilizadas. Essa métrica ajusta o R^2 (que foi de 0.8430) para refletir o impacto de possíveis variáveis desnecessárias ou redundantes, garantindo uma avaliação mais realista da qualidade do modelo.

5.4.3 Multicolinearidade

A multicolinearidade é uma preocupação comum em modelos de regressão linear, mas tem um impacto significativamente menor em modelos como o XGBoost. Isso se deve ao fato de o XGBoost utilizar árvores de decisão como base para suas

previsões, analisando cada variável de forma independente em cada divisão. Mesmo que variáveis sejam altamente correlacionadas, o algoritmo seleciona apenas aquela que melhor reduz a função de perda em cada nó, eliminando naturalmente a redundância (Chen; Guestrin, 2016).

Adicionalmente, o XGBoost incorpora mecanismos de regularização, como L_1 e L_2, que penalizam variáveis menos úteis, ajudando a mitigar os efeitos da multicolinearidade e a controlar a complexidade do modelo. Isso é particularmente útil em situações em que múltiplas variáveis correlacionadas poderiam levar a modelos excessivamente complexos sem ganho preditivo (Chen; Guestrin, 2016). Na prática, essas características tornam o XGBoost menos sensível à multicolinearidade, minimizando seus efeitos negativos sobre o desempenho.

Além disso, no caso de variáveis normalizadas, o impacto da escala é eliminado. A normalização garante que todas as variáveis sejam avaliadas em uma base justa, impedindo que variáveis com valores numéricos maiores dominem o aprendizado. Isso é importante porque as divisões em árvores de decisão podem ser influenciadas pela magnitude das variáveis, especialmente em modelos baseados em gradiente, como o XGBoost (Hastie; Tibshirani; Friedman, 2010).

5.4.4 Resíduos

A análise dos resíduos do modelo XGBoost mostra resultados mistos. O teste de Jarque-Bera indica que os resíduos seguem uma distribuição normal, com uma estatística muito baixa (0,0456) e um p-valor elevado (0,9774), confirmando que a hipótese nula de normalidade não pode ser rejeitada. Isso é corroborado pelo histograma e pelo gráfico Q-Q, que mostram uma boa adequação dos resíduos à normalidade, com pequenas discrepâncias nas caudas.

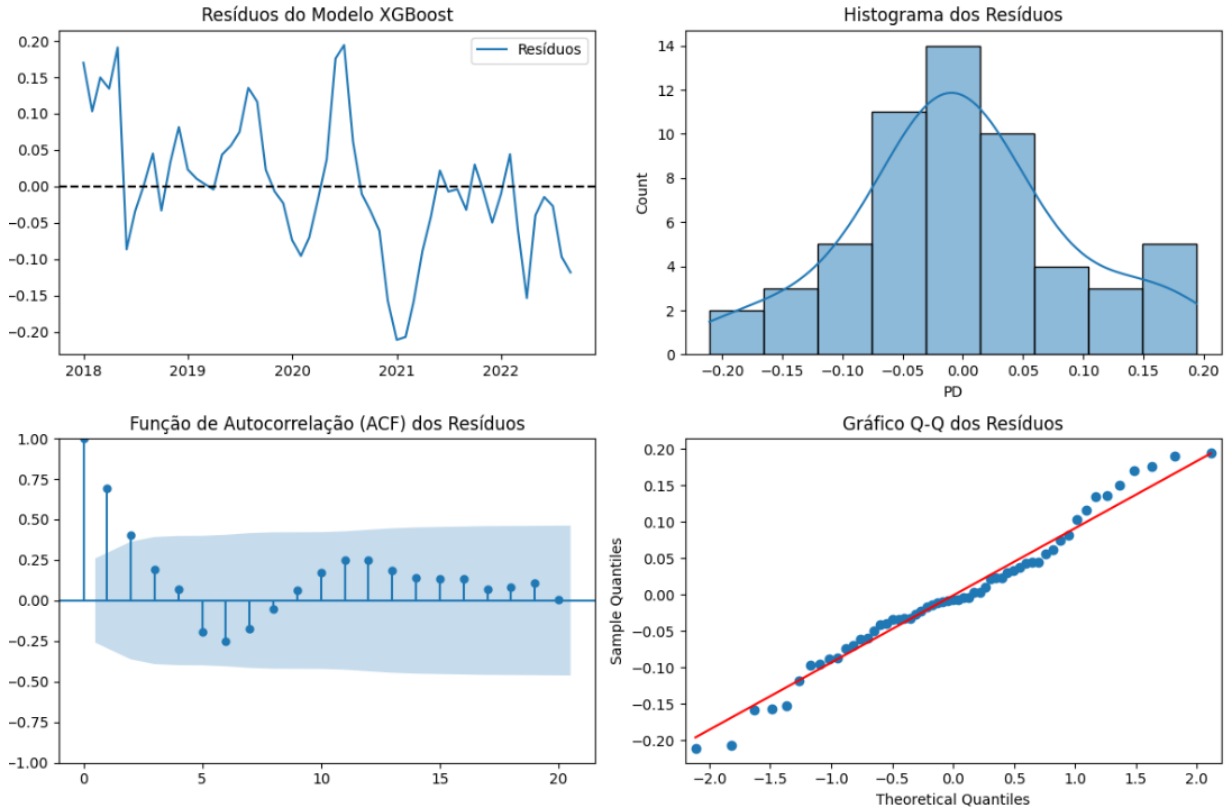
Por outro lado, o teste de Ljung-Box, aponta a presença de autocorrelação significativa nos resíduos, com uma estatística de 52,4738 e p-valor de 0,0000, rejeitando a hipótese nula de ausência de autocorrelação. Isso é reforçado pelo gráfico da ACF, que mostra correlações persistentes em vários lags, especialmente no primeiro, sugerindo que a estrutura temporal da série não foi plenamente.

Além disso, o comportamento dos resíduos ao longo do tempo (gráfico superior

esquerdo) apresenta oscilações regulares, o que pode ser um reflexo das características temporais não capturadas pelo modelo.

Em modelos baseados em árvores de decisão, como o XGBoost, a suposição de homocedasticidade não é necessária. O XGBoost não se baseia nos erros residuais para estimar os coeficientes, mas sim em uma função de perda que minimiza o erro geral em cada etapa de aprendizado.

Gráfico 11 - Resíduos Modelo XGBoost

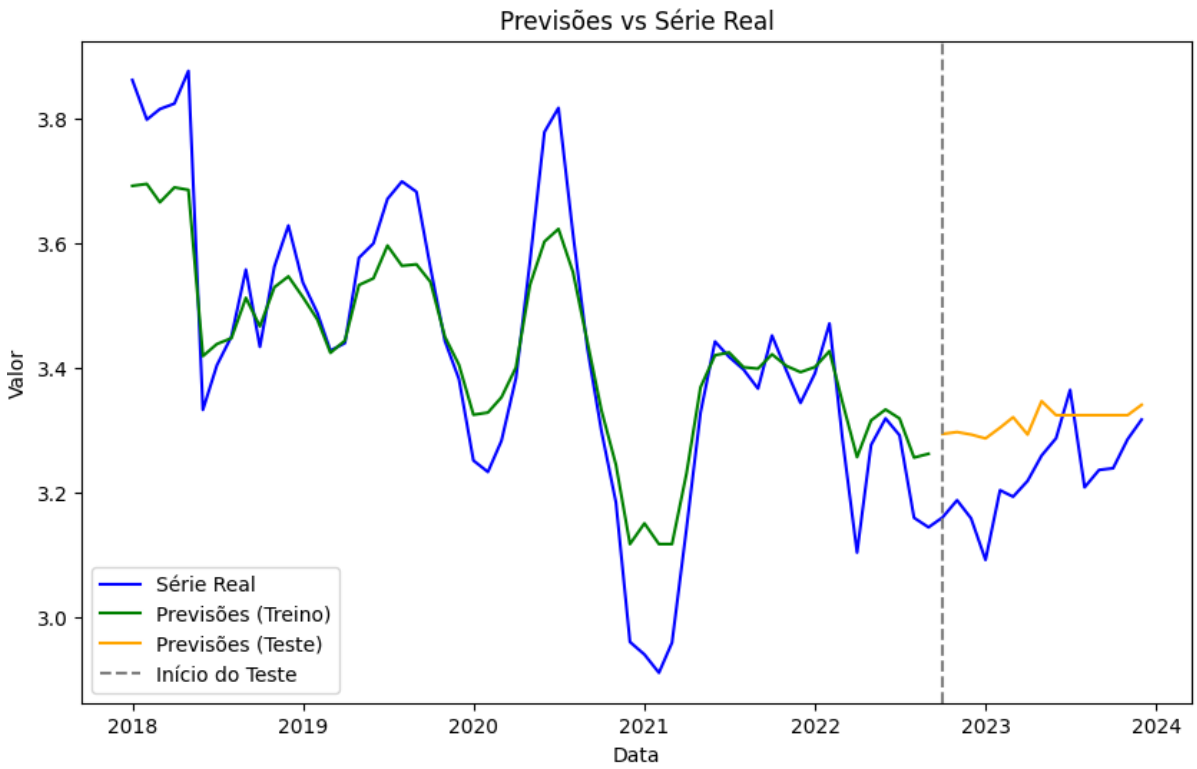


Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

5.4.5 Backtest do modelo

O gráfico a seguir apresenta os resultados obtidos para o modelo XGBoost.

Gráfico 12 - Backtest Modelo XGBoost



Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

O modelo XGBoost apresentou previsões que acompanharam os valores reais da variável-alvo tanto no período de treino quanto no de teste. Durante o treino, as previsões, representadas pela linha verde, seguiram a série real, indicada pela linha azul, capturando as variações temporais e as relações entre as variáveis. No teste, após a linha tracejada, as previsões (linha laranja) mantiveram uma trajetória próxima à da série real (linha azul), com algumas variações observadas em determinados pontos.

As métricas mostram que, no conjunto de treino, o $RMSE$ foi de 0.0922, enquanto o $MAPE$ foi de 2,06%, indicando erros absolutos e percentuais baixos. No conjunto de teste, o $RMSE$ foi de 0.1030, e o $MAPE$ foi de 2,90%, com um aumento em relação ao treino, mas mantendo consistência nos valores.

Os resultados apontam que o modelo conseguiu capturar as dinâmicas temporais presentes nos dados de treino e aplicá-las de forma semelhante aos dados de teste, indicando generalização sem variações significativas nos erros apresentados.

6 SIMULAÇÃO DE IMPACTO

Neste capítulo, será analisado como alterações nas variáveis macroeconômicas podem impactar a projeção da PD. Para isso, foi avaliado o efeito de variações hipotéticas, como uma mudança de 5%, nas variáveis macroeconômicas sobre a taxa de *default*.

A análise foi conduzida por meio de quatro cenários, com alterações de -10%, -5%, +5% e +10% sobre as variáveis macroeconômicas dos modelos. Essas variações foram aplicadas ao período de validação entre outubro de 2022 e dezembro de 2023, abrangendo todos os modelos considerados: Regressão Linear Múltipla, SARIMAX, VAR e XGBoost.

A abordagem da simulação seguiu os seguintes passos:

a) para ambos os modelos, foram introduzidos “choques” que representaram aumentos ou reduções nas variáveis macroeconômicas e suas defasagens durante o período de outubro de 2022 a dezembro de 2023;
b) em seguida, foi calculada a média da taxa de *default* projetada pelos modelos no intervalo analisado, comparando os resultados obtidos com o valor real (sem choques) e com as projeções ajustadas após os choques.

Segue a apresentação dos resultados da variação das variáveis macroecônicas na projeção da PD:

6.1 REGRESSÃO LINEAR MÚLTIPLA

Tabela 4 - Impacto Coeficientes Regressão Linear

Regressão Linear	
Variação (%)	**Variação PD (%)**
-10%	0,301939%
-5%	0,150969%
5%	-0,150969%
10%	-0,301939%

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

Quando há uma redução de 10% nas variáveis explicativas, a PD projetada sofre um aumento de 0,301939%. Por outro lado, ao aumentar as variáveis explicativas em

10%, a PD projetada tem uma redução em 0,301939. Esse comportamento evidencia que o modelo responde de forma tímida aos choques nas variáveis exógenas.

6.2 SARIMAX

Tabela 5 - Impacto Coeficientes SARIMAX

SARIMAX	
Variação (%)	**Variação PD (%)**
-10%	0,046167%
-5%	0,023084%
5%	-0,023084%
10%	-0,046167%

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

Quando as variáveis explicativas são reduzidas em 10%, a PD projetada cresce apenas 0,04%. Já para uma redução de 5%, o impacto na PD é ainda menor, com um crescimento de 0,02%. De maneira semelhante, um aumento de 5% nas variáveis explicativas resulta em uma redução de 0,02% na PD, e, com um aumento de 10%, a queda na PD chega a apenas 0,04%.

Essa análise indica que o modelo SARIMAX apresenta baixa elasticidade, ou seja, a PD é pouco influenciada por alterações nas variáveis explicativas. Esse comportamento pode ser atribuído à forma como o modelo utiliza componentes temporais e interdependências, priorizando padrões e tendências ao longo do tempo, em vez de reagir fortemente a flutuações momentâneas nas variáveis.

O resultado sugere que, no contexto do SARIMAX, as variáveis explicativas têm um papel mais marginal, enquanto fatores temporais e sazonais são provavelmente os principais direcionadores da PD. Isso torna o modelo mais estável para projeções em cenários de stress, mas pode limitar sua capacidade de capturar relações causais diretas entre variáveis explicativas e a PD.

6.3 VAR

Tabela 6 - Impacto Coeficientes VAR

VAR	
Variação (%)	**Variação PD (%)**
-10%	0%
-5%	0%
5%	0%
10%	0%

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

O modelo VAR não apresentou variação na PD projetada em nenhum dos cenários. Isso acontece porque nesse tipo de modelo todas variáveis são endógenas. Logo, se o mesmo choque for dado em todas ao mesmo tempo, o efeito na variação da projeção será nulo. Em alternativa, poderia-se simular impacto em apenas determinada variável do modelo, não sendo aplicado neste trabalho.

6.4 XGBOOST

Tabela 7 - Impacto Coeficientes XGBoost

XGBoost	
Variação (%)	**Variação PD (%)**
-10%	0,084506%
-5%	0,016680%
5%	-0,036715%
10%	-0,094776%

Fonte: Elaborado pelo autor a partir dos dados utilizados para modelagem.

Já no modelo XGBoost, para reduções de 10% e 5%, a PD projetada aumenta em 0,0845% e 0,0166%, respectivamente, enquanto aumentos de 5% e 10% geram um impacto negativo de 0,0367% e 0,0947%.

Essa variação mais pesada na PD em cenários mais extremos indica que o modelo é um pouco mais sensível a choques mais drásticos nas variáveis macroeconômicas.

7 CONSIDERAÇÕES FINAIS

Esta dissertação teve como foco principal a modelagem da PD Forward Looking, um tema amplamente debatido no mercado financeiro brasileiro. Para alcançar esse objetivo, foram exploradas quatro abordagens distintas, com o intuito de guiar o leitor no processo de construção do modelo e apresentar alternativas para a obtenção de uma equação final.

Todos os modelos, com exceção do modelo de regressão linear, demonstraram um bom desempenho durante o período de desenvolvimento, conseguindo captar as oscilações na taxa de *default* com base em variáveis macroeconômicas. Contudo, o modelo SARIMAX se destacou por sua maior precisão nas projeções realizadas, ficando somente abaixo do XGBoost.

Nas simulações de impacto, ao provocar alterações nas variáveis explicativas, o modelo de regressão apresentou uma interpretação mais direta ao traduzir as mudanças nas variáveis macroeconômicas para as variações na taxa de *default*. No entanto, a regressão linear apresentou limitações significativas, como a suposição de linearidade, sensibilidade a outliers, multicolinearidade entre variáveis macroeconômicas e baixa capacidade de capturar estruturas temporais complexas, o que compromete sua qualidade preditiva.

Já o SARIMAX, por sua natureza autoregressiva, utilizou grande parte da informação histórica de *default*, o que reduziu a sensibilidade às variações das variáveis macroeconômicas em suas projeções. Ainda assim, ele se destacou por apresentar previsões estáveis e interpretáveis. Como limitação, observa-se sua suposição de linearidade com variáveis exógenas, sensibilidade à escolha dos parâmetros e maior complexidade computacional quando se lida com muitos preditores correlacionados.

O modelo VAR, embora eficiente na análise de interdependências temporais entre variáveis macroeconômicas, mostrou-se sensível à dimensionalidade da base e à necessidade de estacionariedade, além de não capturar relações causais explícitas nem comportamentos não lineares. Sua aplicabilidade depende fortemente de transformações adequadas nos dados e de escolhas criteriosas do número de defasagens.

Por sua vez, o XGBoost obteve o melhor desempenho preditivo entre os modelos testados, mas apresentou limitações relacionadas à interpretabilidade, alta demanda computacional, risco de overfitting e necessidade de engenharia de variáveis temporais para capturar a dinâmica da série. Além disso, a eficácia do modelo depende fortemente da qualidade dos dados e da calibração adequada dos hiperparâmetros.

Outro aspecto que pesou na escolha do modelo final foi a estabilidade das previsões. Ao contrário de modelos como a regressão linear, que demonstram maior sensibilidade, o SARIMAX ofereceu resultados consistentes, mesmo diante de mudanças nas variáveis explicativas. Isso garante segurança ao tomar decisões estratégicas baseadas nas projeções.

A interpretabilidade do modelo também foi um ponto decisivo. Diferentemente de algoritmos mais complexos, como o XGBoost, que podem ser difíceis de entender e explicar, o SARIMAX proporciona clareza sobre os coeficientes e as influências das variáveis. Isso facilita a comunicação dos resultados para diferentes públicos, incluindo gestores e equipes executivas.

Por fim, o SARIMAX está alinhado com as características do mercado de crédito, onde as dinâmicas temporais desempenham um papel central. Ele é a escolha ideal para equilibrar precisão e clareza, fornecendo projeções confiáveis para embasar decisões estratégicas e minimizar riscos no gerenciamento de crédito.

Futuras análises poderiam incorporar outras técnicas estatísticas, a fim de expandir o estudo. Além disso, a inclusão de novas covariáveis, como o índice de confiança do consumidor, podem ser incorporadas, já que refletem o otimismo ou pessimismo da população em relação à economia, afetando diretamente a propensão a pagar dívidas. O nível de endividamento das famílias e a renda média real também são indicadores importantes, pois mostram o grau de comprometimento financeiro e a capacidade de consumo da população.

A saúde do setor produtivo também merece atenção, com variáveis como a produção industrial e o índice de preços ao produtor (IPP), que ajudam a compreender a dinâmica econômica em níveis mais setoriais. Já o spread bancário e a taxa de concessão de crédito podem revelar as condições de oferta e demanda no mercado de crédito, fundamentais para projetar inadimplência.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, New York, v. 19, n. 6, p. 716–723, 1974.

ALTMAN, E. I**.** Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The Journal of Finance***,* New York, v. 23, n. 4, p. 589-609, Sept. 1968. Disponível em: https://doi.org/10.2307/2978933. Acesso em: 7 abr. 2025.

BANRISUL. **Relatório da administração e demonstrações financeiras**: 2023. Porto Alegre, 2023. Disponível em: https://ri.banrisul.com.br. Acesso em: 10 maio 2025.

BEDNAREK, P.; FRANKE, G. **Dynamics of probabilities of *default***. Frankfurt am Main, June 2024. (Deutsche Bundesbank Discussion Paper, n. 32/2024). Disponível em: Disponível em: https://doi.org/10.2307/2978933. Acesso em: 7 maio 2025.

BERNANKE, B.; GERTLER, M. Inside the black box: the credit channel of monetary policy transmission. **Journal of Economic Perspectives**, Nashville, v. 9, n. 4, p. 27-48, 1995. Disponível em: https://www.aeaweb.org/articles?id=10.1257/jep.9.4.27. Acesso em: 7 maio 2025.

BERNANKE, B.; GERTLER, M.; GILCHRIST, S. The financial accelerator in a quantitative business cycle framework. *In*: TAYLOR, J. B.; WOODFORD, M. (ed.). **Handbook of macroeconomics**. Amsterdam: Elsevier, 1999. v. 1, p. 1341-1393.

BLANCHARD, O. J. On the future of macroeconomic models. **Oxford Review of Economic Policy**, Oxford, v. 34, n. 1-2, p. 70–106, 2018.

BONOMO, M.; MARTINS, B.; PINTO, E. Debt composition and exchange rate balance sheet effect in Brazil: a firm level analysis. **Emerging Markets Review***,* Amsterdam, v. 4, n. 4, p. 368-396,Dec. 2003.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis**: forecasting and control. 5th ed. Hoboken: John Wiley & Sons, 2015.

BREEDEN, J. L.; CROOK, J. Multihorizon discrete time survival models. **Journal of the Operational Research Society**, Oxford, v. 73, n. 1, p. 56-69, 2022. Disponível em: https://doi.org/10.1080/01605682.2020.1777907. Acesso em: 7 maio 2025.

BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multimodel inference**: a practical information-theoretic approach. 2nd ed. New York: Springer, 2002.

CAMURI, P. A. Taxa de câmbio e rentabilidade do setor agropecuário exportador: perspectivas para o primeiro ano de governo Temer. **CNA Brasil**, 2016. Disponível em: https://www.cnabrasil.org.br/assets/arquivos/artigostecnicos/33-artigo_0.12111300%201514912074.pdf. Acesso em: 10 maio 2025.

CHEN, T.; GUESTRIN, C. XGBoost: a scalable tree boosting system. *In:* ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, San Francisco. **Proceedings [...]**. San Francisco: ACM, 2016. p. 785-794. Disponível em: https://dl.acm.org/doi/10.1145/2939672.2939785. Acesso em: 7 maio 2025.

CLAESKENS, G.; HJORT, N. L. **Model selection and model averaging.** Cambridge: Cambridge University Press, 2008.

CMN. CMN altera norma sobre renegociação de dívidas do crédito rural ao amparo do Pronaf. **Notas à Imprensa**, Brasília, 28 mar. 2024. Disponível em: https://www.gov.br/fazenda/pt-br/canais_atendimento/imprensa/notas-a-imprensa/2024/marco/cmn-altera-norma-sobre-renegociacao-de-dividas-do-credito-rural-ao-amparo-do-pronaf. Acesso em: 10 maio 2025.

CRYER, J. D.; CHAN, K.-S. **Time series analysis**: with applications in R. New York: Springer, 2008.

DAINELLI, F.; BET, G.; FABRIZI, E. The financial health of a company and the risk of its *default*: back to the future. **International Review of Financial Analysis**, Greenwich, v. 25, part B, Oct. 2024. Disponível em: https://doi.org/10.1016/j.irfa.2024.103449. Acesso em: 7 maio 2025.

DICKEY, D. A.; FULLER, W. A**.** Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, New York, v. 74, n. 366, p. 427-431, 1979. Disponível em: https://doi.org/10.2307/2286348. Acesso em: 7 maio 2025.

DUAN, Y.; GOODELL, J. W.; LI, H.; LI, X. Assessing machine learning for forecasting economic risk: evidence from an expanded Chinese financial information set. **Finance Research Letters**, Amsterdam, v. 46, part. A, May 2022. Disponível em: https://doi.org/10.1016/j.frl.2021.102273. Acesso em: 7 maio 2025.

ENDERS, W. **Applied econometric time series.** 4th ed. New York: Wiley, 2015.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of Statistics**, Hayward, v. 29, n. 5, p. 1189-1232, Oct. 2001. Disponível em: https://www.jstor.org/stable/2699986. Acesso em: 7 maio 2025.

FRIEDMAN, M. The role of monetary policy. **American Economic Review***,* Nashville, v. 58, n. 1, p. 1-17, Mar. 1968. Disponível em: https://www.aeaweb.org/aer/top20/58.1.1-17.pdf. Acesso em: 7 maio 2025.

GEORGIOU, K.; YANNACOPOULOS, A. N. **Probability of *default* modelling with Lévy-driven Ornstein-Uhlenbeck processes and applications in credit risk under the IFRS 9**. Ithaca: Cornell University, 2023. Disponível em:

https://arxiv.org/abs/2309.12384. Acesso em: 7 maio 2025.

GOMES, R. da S.; OLIVEIRA, E. R. de; SANTOS, G. C. dos; GONÇALVES, R. R. O ambiente socioeconômico influencia o uso de crédito consignado? **Redeca**: Revista Eletrônica do Departamento de Ciências Contábeis & Departamento de Atuária e Métodos Quantitativos, São Paulo, v. 11, 2024. Disponível em: https://doi.org/10.23925/2446-9513.2024v11id66978. Acesso em: 10 maio 2025.

GREENE, W. H. **Econometric analysis.** 8th ed. New York: Pearson, 2018.

GRIGUTIS, A. **Probabilistic overview of probabilities of *default* for low *default* portfolios by K. Pluto and D. Tasche**. Ithaca: Cornell University, 2023. Disponível em: https://arxiv.org/abs/2303.06148. Acesso em: 7 maio 2025.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica.** 5. ed. Porto Alegre: AMGH, 2011.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate data analysis.** 8th ed. Hampshire: Cengage Learning, 2019.

HAMILTON, J. D. **Time series analysis.** Princeton: Princeton University Press, 1994.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**: data mining, inference, and prediction. 2nd ed. New York: Springer, 2010.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting:** principles and practice. 2. ed. Melbourne: OTexts, 2018. Disponível em: https://otexts.com/fpp2/. Acesso em: 9 janeiro 2025.

IASB. **IFRS 9 financial instruments**. London, 2014. Disponível em: https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/. Acesso em: 7 maio 2025.

IPEA. **Evolução do crédito rural nos últimos anos-safra**. Brasília, 2019. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/9286/1/cc_43_nt_evolu%C3%A7%C3%A3o%20do%20cr%C3%A9dito_rural.pdf. Acesso em: 10 maio 2025.

IPEA. **IPEAData**: sistema de séries estatísticas e socioeconômicas. Brasília, 2024. Disponível em: https://www.ipeadata.gov.br/. Acesso em: 10 maio 2025.

IPEA. **Notas técnicas sobre crédito e inadimplência no Brasil**. Brasília, 2022. Disponível em: https://www.ipeadata.gov.br. Acesso em: 10 maio 2025.

JARROW, R. A.; TURNBULL, S. M. Pricing derivatives on financial securities subject to credit risk. **The Journal of Finance**, New York, v. 50, n. 1, p. 53-85, 1995.

JOHANSEN, S. **Likelihood-based inference in cointegrated vector autoregressive models.** Oxford: Oxford University Press, 1995.

KAUFFMANN, L. H. O. **Uma abordagem forward-looking para estimar a PD segundo IFRS 9.** 2017. Dissertação (Mestrado em Ciências – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2017. Disponível em: https://doi.org/10.11606/D.55.2018.tde-06112018-182558. Acesso em: 7 julho 2024.

KAUSHANSKY, V.; LIPTON, A.; REISINGER, C. Transition probability of Brownian motion in the octant and its application to *default* modelling. **Applied Mathematical Finance**, United Kingdom, v. 25, n. 5-6, p. 434-465, 2018. Disponível em: https://doi.org/10.1080/1350486X.2018.1481439. Acesso em: 7 dezembro 2024.

KILIAN, L.; LÜTKEPOHL, H. **Structural vector autoregressive analysis**. Cambridge: Cambridge University Press, 2017.

LÜTKEPOHL, H. **New introduction to multiple time series analysis.** Berlin: Springer, 2005.

MCKINNEY, W. **Python for data analysis**: data wrangling with Pandas, NumPy, and IPython. United States: O'Reilly Media, 2012.

MIAO, H.; RAMCHANDER, S.; RYAN, P.; WANG, T. Default prediction models: the role of forward-looking measures of returns and volatility. **Journal of Empirical Finance**, Amsterdam, v. 46, p. 146-162, Mar. 2018. Disponível em: https://doi.org/10.1016/j.jempfin.2018.01.001. Acesso em: 7 maio 2025.

MISHKIN, F. S. **The economics of money, banking, and financial markets.** 8th ed. Boston: Pearson Education, 2007.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis.** 5th ed. Hoboken: Wiley, 2012.

REINHART, C. M.; ROGOFF, K. S. **This time is different**: eight centuries of financial folly. Princeton: Princeton University Press, 2009.

SANTOLIN, R.; GAMA, F. Operações de crédito, desigualdade, inadimplência e crescimento da renda: uma avaliação para os estados brasileiros no período 2001–2014. **Planejamento e Políticas Públicas**, Brasília n. 59, p. 1-28, 2021. Disponível em: https://www.ipea.gov.br/ppp/index.php/PPP/article/view/1331. Acesso em: 10 maio 2025.

SERASA EXPERIAN. **Inadimplência no campo atinge 7,7% da população rural que atua como pessoa física, revela Serasa Experian**. São Paulo, 2024. Disponível em: https://www.serasaexperian.com.br/sala-de-imprensa/agronegocios/inadimplencia-no-

campo-atinge-77-da-populacao-rural-que-atua-como-pessoa-fisica-revela-serasa-experian/. Acesso em: 10 maio 2025.

SETHEE, S. Inflation forecasting: application of SARIMAX model for the Indian monthly inflation forecasting and checking the validity of the New Keynesian Phillips Curve in the Indian context. **Medium**, *[s.l.],* May 2020. Disponível em: https://medium.com/inflation-forecasting-using-sarimax-and-nkpc/plotting-monthly-inflation-over-the-selected-time-period-to-check-if-the-time-series-has-any-35e3b1fac761. Acesso em: 3 maio 2025.

SIMÕES, R. de S. **Modelagem da PD forward looking**: uma análise de impacto dos fatores macroeconômicos na inadimplência bancária nacional. 2022. Dissertação (Mestrado em Economia) – Escola de Economia de São Paulo, Fundação Getulio Vargas, São Paulo, 2022. Disponível em: https://repositorio.fgv.br/bitstreams/8ca6d604-a9e6-4c80-ad54-e757374480e1/download. Acesso em: 10 dez. 2024.

SIMS, C. A. Macroeconomics and reality. **Econometrica**, New Haven, v. 48, n. 1, p. 1-48, 1980.

SOUZA, W. R. de. **Relação entre leis de Refis e a inadimplência do crédito rural**: análise do impacto das renegociações de dívidas sobre as operações de crédito rural securitizadas. 2019. Dissertação (Mestrado em Políticas Públicas e Desenvolvimento) – Programa de Pós-Graduação em Políticas Públicas e Desenvolvimento, Instituto de Pesquisa Econômica Aplicada, Brasília, 2019. Disponível em: https://www.ipea.gov.br/sites/images/mestrado/turma3/willer-roger-de-souza.pdf. Acesso em: 10 maio 2025.

STOCK, J. H.; WATSON, M. W. **Introduction to econometrics.** 4th ed. Boston: Pearson, 2019.

VANDERPLAS, J. **Python data science handbook**: essential tools for working with data. United States: O'Reilly Media, 2016.

WEI, W. W. S. **Time series analysis**: univariate and multivariate methods. 2nd ed. Boston: Pearson, 2006.

WOOLDRIDGE, J. M. **Introductory econometrics**: a modern approach. 7th ed. Boston: Cengage, 2020.

XU, X. **Estimating lifetime expected credit losses under IFRS 9**. *[S.l.],* 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758513. Acesso em: 3 maio 2025.

XU, Y.; LI, J.; WU, A. An XGboost algorithm based model for financial risk prediction. **Tehnički Vjesnik**, Croatia, v. 31, n. 6, p. 1898-1907, 2024. Disponível em: https://doi.org/10.17559/TV-20231021001043. Acesso em: 3 maio 2025.

YANG, B. H. Point-in-time PD term structure models for multi-period scenario loss

projection: methodologies and implementations for IFRS 9 ECL and CCAR stress testing. **Journal of Risk Model Validation**, United Kingdom, v. 11, n. 3, Jan. 2017. Disponível em: https://mpra.ub.uni-muenchen.de/76271/. Acesso em: 3 maio 2025.