

Janaina de Alcântara Buzachi Garcia Thomazi

Modelagem da Probabilidade de *Default* com uso de Séries Temporais

Brasil

2022, v-X1.X9.X7

Janaina de Alcântara Buzachi Garcia Thomazi

Modelagem da Probabilidade de *Default* com uso de Séries Temporais

Dissertação apresentada ao Curso de Mestrado Profissional em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Universidade de Brasília - UnB

Faculdade de Administração Contabilidade e Economia - FACE

Departamento de Economia - ECO

Programa de Pós-Graduação

Orientador: Marina Delmondes de Carvalho Rossi

Brasil

2022, v-X1.X9.X7

Janaina de Alcântara Buzachi Garcia Thomazi

Modelagem da Probabilidade de *Default* com uso de Séries Temporais/ Janaina de Alcântara Buzachi Garcia Thomazi. – Brasil, 2022, v-X1.X9.X7-78p. : il. (algumas color.) ; 30 cm.

Orientador: Marina Delmondes de Carvalho Rossi

Dissertação (Mestrado) – Universidade de Brasília - UnB
Faculdade de Administração Contabilidade e Economia - FACE
Departamento de Economia - ECO
Programa de Pós-Graduação, 2022, v-X1.X9.X7.

1. Probabilidade de *default*. 2. Macroeconomia. 3. Risco de crédito. 4. Crédito imobiliário. 5. Perda esperada. 6. Inferência com SARIMAX. II. Universidade de Brasília. III. Faculdade de Administração, Contabilidade e Economia - FACE. IV. Departamento de Economia IV. Modelagem da Probabilidade de *Default* com uso de Séries Temporais

Janaina de Alcântara Buzachi Garcia Thomazi

Modelagem da Probabilidade de *Default* com uso de Séries Temporais

Dissertação apresentada ao Curso de Mestrado Profissional em Economia, Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Economia

Trabalho aprovado. Brasil, 25 de fevereiro de 2022:

Marina Delmondes de Carvalho Rossi
Orientador

Professor
Daniel Oliveira Cajueiro

Professor
Francisco Helano de Oliveira Farias

Brasil
2022, v-X1.X9.X7

*Este trabalho é dedicado às mulheres, mães e trabalhadoras que
sonham em se tornar cientistas.*

Agradecimentos

À minha família, esposo, filhos, pai e mãe, irmão e irmã, tias, primo, sogro e sogra, cunhados e cunhadas, sobrinhos, sobrinhas e afilhados que em algum momento entenderam minha ausência, ansiedade e principalmente, a necessidade de me dedicar ao estudo e à produção científica.

Aos meus avós que antes de partirem deixaram a importante lição da educação como tesouro maior.

Às amigas, amigos e vizinhos queridos que cuidaram dos meus filhos quando precisavam da atenção que eu não podia dar.

Ao grupo de estudos do mestrado que transformava minhas preocupações em entregas, em especial à Sabrina Alencar Ferreira, Camila Silva Pinto, Leonardo Rodrigo Caetano Dizaró Caetano e Ana Paula Lara Rabelo.

À minha orientadora Marina Delmondes de Carvalho Rossi, que com incrível generosidade e sororidade compreendeu e auxiliou prontamente em todas as minhas necessidades.

Aos professores da banca examinadora, Daniel Oliveira Cajueiro e Francisco Helano de Oliveira Farias, por toda contribuição e apoio na execução do trabalho.

Ao corpo docente e aos monitores da FACE/UNB com quem tive a oportunidade de me aventurar em terras, por mim, nunca antes exploradas. Em especial aos professores Herbert Kimura, José Guilherme de Lara Resende, Pablo de Carvalho, Ricardo Vieira Barroso e ao monitor Pedro Watuhã dos Santos Guerra.

Aos meus colegas de profissão pela oportunidade, incentivo, assistência, reconhecimento e aprendizado.

À minha professora e colegas de Yoga, pelo incentivo em manter corpo e mente em equilíbrio.

Ao divino poder do universo, por me trazer até o mestrado, por manter a motivação necessária e por apresentar todas as respostas as quais precisava.

*“If they don’t give you a seat
at the table,
bring a folding chair.
(Shirley Chisholm)*

Resumo

O presente trabalho tem por objetivo identificar variáveis macroeconômicas que possam inferir a estimação da probabilidade de *default* de uma carteira de crédito imobiliário de um banco brasileiro do Segmento 1(S1). As instituições financeiras tem por prática avaliar a probabilidade de *default* a partir de critérios de escolha de perfil de riscos dos tomadores de crédito. No entanto, modelos que observam variáveis macroeconômicas têm produzido melhores preditores para o risco de crédito. Utilizando a metodologia SARIMAX desenvolvemos um modelo de inferência que, além de considerar, os valores e os erros passados da probabilidade de *default*, por meio da análise da autocorrelação e da média móvel, adiciona variáveis exógenas e componentes sazonais. Nossos resultados demonstraram coeficientes estatisticamente significantes para 9 variáveis macroeconômicas, além de correlação serial da variável dependente no Lag 1 e 2. As variáveis explicativas (1) valores dos insumos e serviços necessários às obras e serviços de engenharia (SINAPI), renda nacional disponível bruta das famílias, comprometimento de renda das famílias com o serviço da dívida e o endividamento das famílias com o Sistema Financeiro Nacional se mostraram significantes para explicar que quando maior seu crescimento menor a probabilidade de *default*. A relação pode ser entendida pela resposta da probabilidade de *default* à postura dos Bancos quanto à oferta de crédito diante do grau de incerteza do cenário econômico. As variáveis explicativas (2) índice nacional de preços ao consumidor-amplo (IPCA), inadimplência da carteira de crédito com recursos direcionados, taxa média mensal de juros das operações de crédito com recursos direcionados, *spread* médio das operações de crédito com recursos direcionados, e o índice de valor de garantias reais (IVG-R) se mostraram estatisticamente significantes para explicar que quando maior seu crescimento maior a probabilidade de *default*. A relação pode ser justificada pela resposta da probabilidade de *default* aos períodos de crescimento, choques e de recessão econômica.

Palavras-chave: Probabilidade de *default*, macroeconomia, risco de crédito, crédito imobiliário, perda esperada, inferência com SARIMAX.

Abstract

The present work aims to identify macroeconomic variables that can infer the estimation of the probability of default of a mortgage portfolio of a Brazilian bank in Segment 1(S1). It is the practice of financial institutions to assess the default probability based on criteria for choosing the risk profile of borrowers. However, models that observe macroeconomic variables have been producing better predictors for credit risk. Using the SARIMAX methodology, we developed an inference model that, in addition to considering the values and past errors of the default probability, through the analysis of autocorrelation and the moving average, adds exogenous variables and seasonal components. Our results showed statistically significant coefficients for 9 macroeconomic variables, in addition to a serial correlation of the dependent variable in Lag 1 and 2. The explanatory variables (1) values of inputs and services required for engineering works and services (SINAPI), Households gross disposable national income, Household debt service ratio and Household debt to income, proved to be significant to explain that the higher its growth, the lower the probability of default. The relationship can be understood by the response of the default probability to the attitude of banks regarding the offer of credit in face of the degree of uncertainty of the economic scenario. The explanatory variables (2) Broad National Consumer Price Index (IPCA), Percent of 90 days past due loans of earmarked credit operations outstanding (Households/Real estate financing), Month average interest rate of earmarked new credit operations (Households/Real estate financing), Average spread of earmarked new credit operations(Households/Total), and the Residential real estate collateral value index (IVG-R) proved to be statistically significant to explain that the greater their growth, the greater the probability of default. The relationship can be justified by the response of default probability to periods of economic growth, shocks and recession.

Keywords: Default probability, macroeconomics, credit risk, real estate credit, expected loss, inference with SARIMAX.

Lista de ilustrações

Figura 1 – Distribuição Gaussiana da PD em log	28
Figura 2 – Série Probabilidade de <i>default</i>	29
Figura 3 – Evolução Temporal: Probabilidade de <i>default</i> e IPCA	30
Figura 4 – Evolução Temporal: Probabilidade de <i>default</i> e Taxa de Juros Real (Selic)	30
Figura 5 – Evolução Temporal: Probabilidade de <i>Default</i> e Endividamento das Famílias com o Sistema Financeiro Nacional em Relação à Renda Acumulada dos Últimos Doze Meses (RNDBF)	31
Figura 6 – Evolução Temporal: Probabilidade de <i>Default</i> e Concessões de Crédito com Recursos Direcionados - Pessoas físicas - Financiamento imobiliário Total	31
Figura 7 – Evolução Temporal: Probabilidade de <i>Default</i> e <i>Spread</i> Médio das Operações de Crédito com Recursos Direcionados - Pessoas físicas - Total	32
Figura 8 – Evolução Temporal: Probabilidade de <i>Default</i> e Inadimplência da Carteira de Crédito com Recursos Direcionados - Pessoas físicas - Financiamento Imobiliário Total	32
Figura 9 – Evolução Temporal: Probabilidade de <i>Default</i> e Renda Nacional Disponível Bruta das Famílias Restrita- em valores constantes, deflacionada pelo IPCA, com ajuste sazonal (média móvel trimestral)	33
Figura 10 – Evolução Temporal: Probabilidade de <i>default</i> e Comprometimento de renda das Famílias com o Serviço da Dívida com o Sistema Financeiro Nacional - Com ajuste sazonal (RNDBF)	34
Figura 11 – Evolução Temporal: Probabilidade de <i>Default</i> e Pessoas ocupadas - PNADC	34
Figura 12 – Evolução Temporal: Probabilidade de <i>Default</i> e Índice da taxa de câmbio real (IPCA) - Dólar Americano	36
Figura 13 – Evolução Temporal: Probabilidade de <i>Default</i> e PIB a Preços de Mercado com Variação Real Trimestral sobre mesmo trimestre do ano anterior	36
Figura 14 – Evolução Temporal: Probabilidade de <i>Default</i> e Remuneração de Empregados - mensal - líquido	37
Figura 15 – Evolução Temporal: Probabilidade de <i>Default</i> e Rendimento Médio Real Habitual das Pessoas Ocupadas -Setor privado e público - PNADC	37
Figura 16 – Evolução Temporal PD: Índice de Valores de Garantias de Imóveis Residenciais Financiados (IVGR)	38
Figura 17 – Parcelas da função de autocorrelação (ACF) e autocorrelação parcial (PACF) depois da primeira diferenciação	44
Figura 18 – Autocorrelação da variável dependente	45

Figura 19 – Análise dos resíduos SARIMAX (0,1,3) (0,0,0) [12]	49
Figura 20 – Análise de resíduos SARIMAX (0,2,2) (0,0,0) [12]	52
Figura 21 – Previsão na amostra	52

Lista de quadros

Quadro 1 – Teste de raiz unitária.	43
Quadro 2 – Causalidade de Granger	46
Quadro 3 – SARIMAX (0,1,3) (0,0,0) [12]	48
Quadro 4 – Revisão de suposições SARIMAX (0,1,3) (0,0,0) [12]	49
Quadro 5 – SARIMAX (1,1,1) (0,0,0) [12]	50
Quadro 6 – SARIMAX (0,2,2) (0,0,0) [12]	51
Quadro 7 – Revisão de suposições SARIMAX (0,2,2) (0,0,0) [12]	51
Quadro 8 – Variáveis Econômicas	65
Quadro 9 – Pesquisa mínimo AIC Modelo SARIMAX(0,1,3)(0,0,0)[12]	67
Quadro 10 – Pesquisa mínimo AIC Modelo SARIMAX(0,2,2)(0,0,0)[12]	68

Lista de símbolos

θ	Letra grega minúscula theta
Θ	Letra grega maiúscula theta
ϕ	Letra grega minúscula phi
Φ	Letra grega maiúscula phi
ε	Letra grega minúscula epsilon
Δ	Letra grega Delta
Σ	Letra grega minúscula sigma

Sumário

1	INTRODUÇÃO	23
2	DADOS	27
3	METODOLOGIA	39
4	RESULTADOS	47
5	CONCLUSÕES	57
	REFERÊNCIAS	59
	APÊNDICES	63
	APÊNDICE A – VARIÁVEIS ECONÔMICAS	65
	APÊNDICE B – MÍNIMO AIC	67
	APÊNDICE C – CÓDIGO PYTHON	69

1 Introdução

Os valores de perda esperada são insumos para compor o preço dos produtos de crédito, além de ser o conceito adotado para as demonstrações contábeis, em substituição ao de perda incorrida, conforme estabelecido pela International Financial Reporting Standards (IFRS) 9, correlacionado às normas brasileiras pelo Comitê de Pronunciamento Contábeis (CPC) pelo pronunciamento técnico CPC 48 (BRASIL, 2016). Caso o modelo preditivo de perda esperada não estime os impactos das alterações macroeconômicas, no curto prazo, pode submeter a Instituição Financeira à eventos de risco de crédito. Soma-se o aumento da sensibilidade aos riscos de forma integrada, considerando o atual cenário econômico brasileiro, o qual ainda oscila em razão do crescimento dos índices de inflação e desemprego como consequência das ações de contenção de disseminação do vírus SARS-CoV-2 nos anos de 2020 e 2021.

No segundo trimestre de 2020 a atividade econômica vivenciou um colapso histórico em decorrência da pandemia. Países da América Latina como o Brasil, Chile, Colômbia, México e Peru experimentaram contrações trimestrais no Produto Interno Bruto (PIB) maiores do que em qualquer outra recessão já registrada (IMF, 2020).

Entre 2019 e 2020, ainda que em período de recessão, o crédito imobiliário continuou a crescer, o impulso deu-se especialmente em razão das baixas taxas de juros e novos indexadores (BCB, 2021). O mercado de crédito imobiliário brasileiro apresentou significativo crescimento entre 2003 e 2018, passando de 1,5 % para quase 10 % do PIB. A sua participação na carteira de crédito total passou de 6 % para 20 % nesse período (IPEA, 2020).

A presente pesquisa investiga variáveis macroeconômicas que possam inferir a estimação da probabilidade de *default* de uma carteira de crédito imobiliário de um banco brasileiro do Segmento 1(S1). A partir de dados da probabilidade de *default* (PD) do portfólio de crédito imobiliário de um banco brasileiro do Segmento 1 (S1) desenvolvemos um modelo de inferência baseado na metodologia SARIMAX de Box e Tiao (1975) com variáveis exógenas e componentes sazonais. Nossos resultados demonstraram coeficientes estatisticamente significantes para 9 variáveis macroeconômicas, relacionadas à inflação, renda, endividamento, inadimplência, taxa de juros e *spread* das operações de crédito, bem como com a avaliação de imóveis e preços da construção civil e preços da construção civil.

O risco de crédito é um dos principais riscos de um Banco. Define-se o crédito como o risco de perda potencial para o banco se um tomador ou contraparte, deixar de cumprir suas obrigações nos termos acordados.(APOSTOLIK; CHRISTOPHER; PETER, 2009).

A taxa de inadimplência é um importante indicador da qualidade da carteira de

empréstimos dos Bancos. Uma alta taxa de inadimplência significa elevado risco de crédito. Órgãos de supervisão utilizam a taxa de inadimplência, junto a outros indicadores, para medir a solidez do sistema financeiro. Isso porque pode indicar a iminência de falências bancárias, além de significar uma redução da lucratividade dos bancos e, conseqüentemente, redução da atividade econômica em virtude da restrição à concessão de novos empréstimos. (TABAK et al., 2007).

Em 1988, o Basel Committee on Banking Supervision (BCBS) instituiu o primeiro Acordo de Basiléia, o qual propôs orientações sobre o nível de capital seguro para a atividade bancária considerando os riscos. Definiram-se os conceitos de capital regulatório, de fatores de ponderação de riscos de ativos e de índice mínimo de Capital para Cobertura do Risco de Crédito (índice de Basiléia). Em 1996, um aditivo trouxe a ampliação dos controles sobre riscos incorridos pelos bancos, a extensão dos requisitos para a definição do capital mínimo (ou regulatório), adicionando o Risco de Mercado e a utilização de modelos internos na mensuração de riscos.

Em 2004 o BCBS revisa os acordos e publica o Basiléia II, adicionando 3 pilares aos critérios de Basiléia I. O Pilar 1, trata dos requerimentos de capital mínimo para risco de crédito, mercado e operacional; o Pilar 2, objetiva o fortalecimento da supervisão bancária; e o Pilar 3, abarca a auto-regulação ou disciplina do mercado. O Acordo de Basiléia III, de 2012, traz como alterações, o aumento dos requerimentos de capital e maior incidência da supervisão bancária.

O Pilar 1 do acordo de Basiléia II possibilita o uso de modelos internos pelas Instituições Financeiras para o cálculo dos requisitos mínimos de capital, com base em fatores de riscos conforme a estrutura de capital. Dentre os principais elementos estimados para o risco de crédito estão a perda esperada (EL - Expected Loss), a probabilidade de inadimplência (PD - Probability of Default), a perda dada a inadimplência (LGD - Loss Given Default) e a exposição à inadimplência (EAD - Exposure at Default) (BCBS 2005a; 2005b).

Pesquisas dos últimos anos, buscam elencar quais os fatores fundamentais para os modelos de risco de crédito. Busca-se compreender se o risco de crédito é ocasionado por fatores idiossincráticos, por fatores sistemáticos, ou por ambos. É sabido que fatores idiossincráticos são determinantes na decisão de um empréstimo bancário, porém variáveis macroeconômicas também podem explicar o risco do tomador.

CARLING et al. (2007) desenvolvem um modelo para estimar o risco de *default* da carteira de empréstimos de um banco da Suécia. Os autores medem características específicas da firma e avaliam as condições macroeconômicas. Como conclusão indicam que variáveis macroeconômicas como o hiato do produto, curva de juros e expectativa das famílias quanto à atividade econômica futura ajudam a prever o risco de *default* das firmas. Ao comparar o modelo que utilizava variáveis macroeconômicas com o modelo

condicionado somente às informações específicas da firma, concluem que o primeiro seria mais apto a captar o risco de *default* do período.

AMATO e FURFINE (2003a) analisam firmas dos EUA classificadas pela Standard & Poor's entre 1981 e 2001. Os autores entendem que, mesmo as classificações de risco das firmas tendo poucas alterações no período analisado, classificar novas firmas e mudanças de classificação demonstra alta sensibilidade às condições do ciclo econômico. O estudo conclui que a utilização de modelos de risco de crédito que não considerem o ambiente macroeconômico pode acentuar o comportamento pró-cíclico dos bancos, com potenciais consequências macroeconômicas.

ALLEN e SAUNDERS (2003b) pesquisam modelos acadêmicos e proprietários para examinar como são incorporados os efeitos do risco macroeconômico e sistemático às medidas de exposição ao risco de crédito. Os autores concluem que muitos modelos consideram a correlação entre a probabilidade de inadimplência e fatores cíclicos. Poucos modelos ajustam as taxas de perda (LGD) para refletir os efeitos cíclicos. A possibilidade de correlação sistemática entre PD e LGD também está negligenciada. Durante uma expansão econômica estimativas otimistas dos modelos para o risco de *default*, reforçam o comportamento pró-cíclico dos empréstimos bancários. Em períodos de recessão, se as estimativas forem pessimistas, uma política monetária expansionista do Banco Central pode não encorajar os bancos a emprestar mesmo para devedores com baixo nível de risco de crédito.

Segundo BORIO et al. (2001a), durante uma expansão, subestimar riscos contribui para crescimento do crédito, inflação dos preços de ativos, redução dos *spreads* dos empréstimos, e favorece que os Bancos mantenham menos capital e provisões. Na recessão, momento em que o risco e a inadimplência são elevados, o contrário pode acontecer.

O crescimento rápido do crédito é um fator definitivo para a ocorrência de crises bancárias. A maioria das crises bancárias é precedida por expansão dos empréstimos (*lending booms*), embora o contrário não seja verdadeiro, ou melhor, uma expansão dos empréstimos não é seguida necessariamente por uma crise bancária. (GOURINCHAS; VALDÉS; LANDERRECHE, 2001b).

Ao considerar os efeitos do ciclo econômico sobre o risco de crédito, estudos propõem relacioná-lo às variáveis macroeconômicas utilizando modelos econométricos. KOOPMAN e LUCAS (2005c) estudam o comportamento da taxa de *default* e o *spread* de crédito, e sua relação com o ciclo de negócios, medido pelo PIB real. Os autores mostram a existência de um co-movimento cíclico positivo entre os *spreads* e as taxas de *default* e negativo entre os *spreads* e o PIB real. Comentam ainda as implicações das descobertas para o gerenciamento de risco de crédito.

Poucos estudos no Brasil relacionam o risco de crédito ao ambiente macroeconômico.

SCHECHTMAN et al. (2004) e SCHECHTMAN (2006) verificam a adequação dos níveis de provisão e capital regulamentar exigidos pelo Banco Central do Brasil (BCB) para cobrir a exposição ao risco de crédito. CHU (2001) analisa a sensibilidade da inadimplência a aumentos nas variáveis explicativas, a partir de fatores macroeconômicos que explicam a inadimplência bancária. O trabalho utilizou um modelo de correção de erro com variáveis de *spread* bancário, índice de produção industrial, índice de desemprego, taxa de juros Selic e taxa média de compulsório.

Conforme estabelecido na Resolução do Conselho Monetário Nacional nº 4.557/2017, artigo 23, a estrutura de gerenciamento de riscos deve prever para o risco de crédito, dentre outros, a estimação das perdas esperadas associadas ao risco de crédito, sendo que a estimativa da perda esperada deve considerar o ambiente macroeconômico corrente e alterações previstas no curto prazo. (BRASIL, 2017)

O presente trabalho visa contribuir com os estudos sobre o risco de crédito ao apresentar uma análise de como o ambiente macroeconômico influencia a probabilidade de *default* e por consequência a perda esperada dos empréstimos imobiliários.

O trabalho possui a seguinte estrutura. No capítulo 2, apresentamos o conjunto de dados e analisamos o conjunto de variáveis dentro do contexto de acontecimentos relevantes que causaram alterações no cenário macroeconômico brasileiro no período da amostra. No capítulo 3, apresentamos o modelo, as especificações e os procedimentos adotados para a escolha dos parâmetros. No capítulo 4, mostramos os resultados dos modelos, escolhemos um modelo e apresentamos as variáveis estatisticamente significantes. No último capítulo, apresentamos uma breve conclusão sobre os achados da pesquisa.

2 Dados

Utilizamos dados mensais de probabilidade de *default* de uma carteira habitacional de um banco brasileiro do Segmento 1 (S1) para um período de 5 anos. Nosso *dataset* começa em janeiro de 2015 e termina em agosto de 2020. Nossa amostra consiste em 68 meses.

O cálculo da perda esperada (Expected Loss, EL), dá-se por meio da seguinte equação:

$$EL = PD * EAD * LGD \quad (2.1)$$

A *Probability of default* (PD) é a probabilidade de o devedor entrar em situação de *default*, e não cumprir com as suas obrigações. A PD é dada por nota de *rating*, e consiste na porcentagem média de devedores que inadimplam neste grau de *rating* no decorrer de um ano (BCBS, 2005a).

A *Exposure at default* (EAD) é o valor que o credor pode vir a perder se o devedor entrar em *default* sem o credor ter qualquer recuperação do montante investido. Fornece uma estimativa do valor em aberto (valores sacados mais saques futuros prováveis de linhas ainda não sacadas) caso ocorra a inadimplência do mutuário (BCBS, 2005a).

A *Loss Given default* (LGD) trata da porcentagem do montante que o credor irá perder. Em geral, quando uma contraparte entra em *default*, uma parte do valor investido é reavido pela contraparte credora. Essas perdas geralmente são mostradas como uma porcentagem da EAD, e dependem, entre outros, do tipo e montante das garantias, bem como o tipo de mutuário e os rendimentos esperados da liquidação dos ativos. A porcentagem que é recuperada define-se por *Recovery Rate* (RR). O cálculo da RR consubstancia-se em $RR = 1 - LGD$ (BCBS, 2005a).

Para obter o valor da perda esperada da contraparte credora, é preciso, portanto, estimar os *inputs* mencionados. A EAD e a LGD são fáceis de conhecer, visto que a EAD é apenas o valor que o credor pode vir a perder no caso de *default* da contraparte devedora, e sabendo a *Recovery Rate* a LGD fica simples de calcular. Por outro lado, calcular a PD é mais complexo e com diferentes métodos de obtenção.

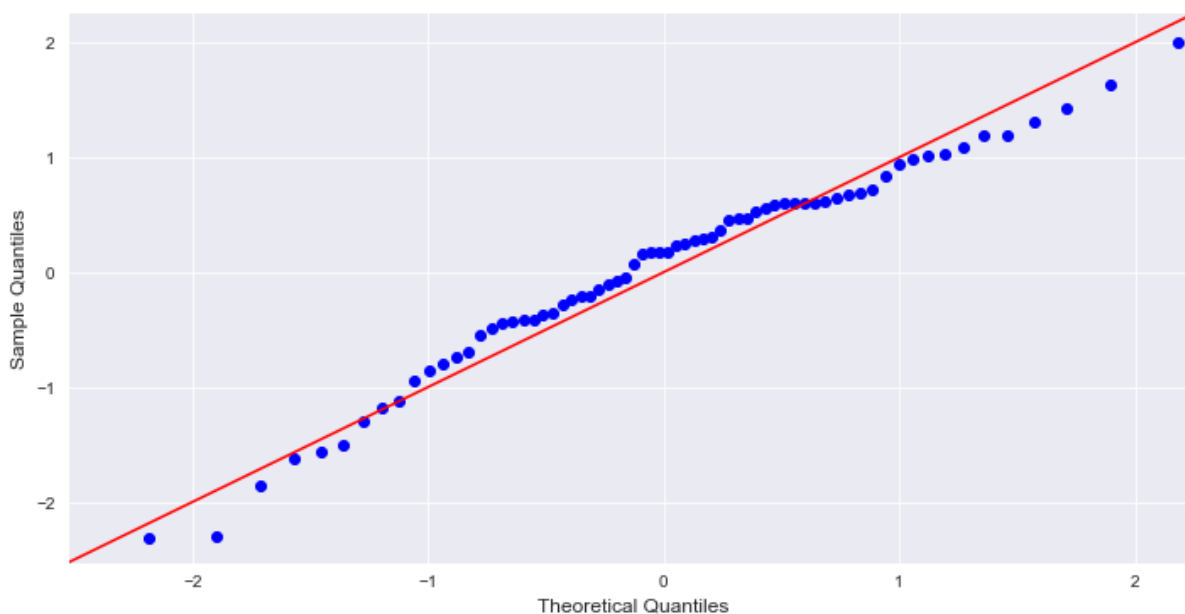
As tabelas de desempenho da PD são rotineiramente desenvolvidas usando algoritmos de classificação. Muitos estudos têm examinado a precisão de classificadores alternativos. É prática comum usar um pequeno número de conjuntos de dados, muitos dos quais contêm apenas alguns casos e/ou variáveis independentes. (LESSMANN et al., 2015).

Nosso *dataset* contém os valores de EL, PD, EAD e LGD em formato de séries de tempo. O cálculo da PD não considera variáveis macroeconômicas. A Probabilidade de *default* informada no *dataset* trata-se da inadimplência observada no período de janeiro de 2015 a agosto de 2020. A partir do ano de 2018 foi inserido o conceito de ativo problemático presente no artigo 24 da Resolução do Conselho Monetário Nacional nº 4.557/2017.

Referido artigo amplia o conceito de *default*, ao estabelecer que para o gerenciamento do risco de crédito, a exposição deve ser caracterizada como ativo problemático quando verificado pelo menos um dos seguintes eventos, quais sejam, atraso superior há mais de noventa dias e indicativos de que a obrigação fará uso de garantias ou colaterais para ser honrada (BRASIL, 2017).

Utilizamos a série de tempo da Probabilidade de *default* como variável dependente. Por se tratar de variável censurada entre 0 e 100% aplicamos a função logarítmica (\log) na variável dependente, com o objetivo de melhorar a distribuição, cujo resultado está conforme Figura 1.

Figura 1 – Distribuição Gaussiana da PD em log



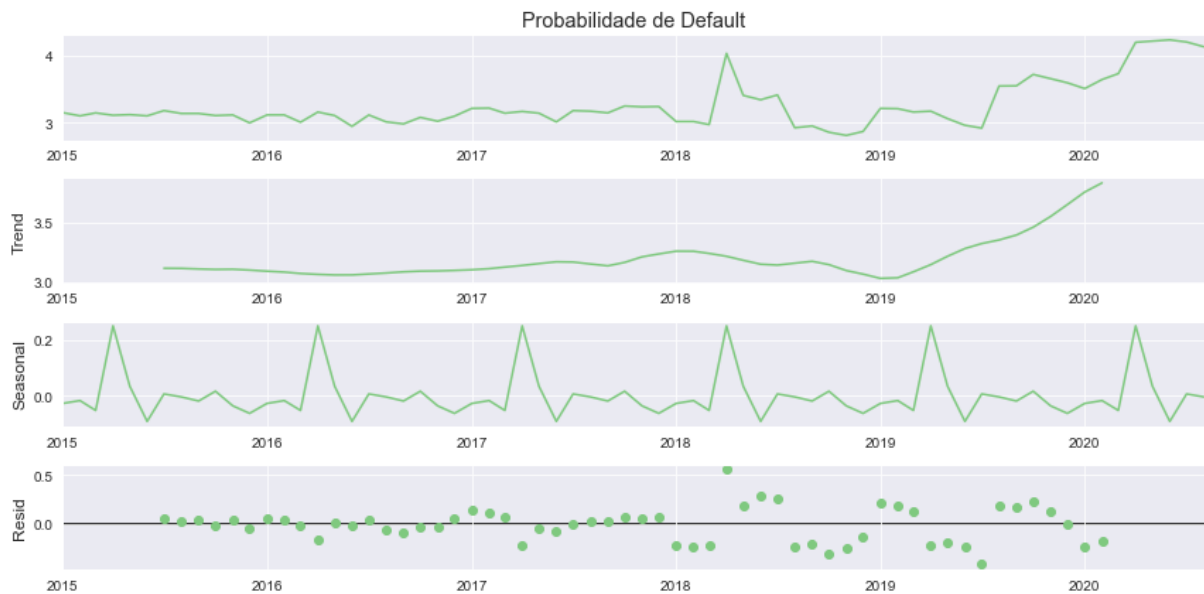
Fonte: Elaborada pela autora (2022)

Na Figura 2, mostramos a tendência, sazonalidade e resíduos da PD. Podemos ver que a PD aparenta possuir sazonalidade, o que indica a influencia por fatores externos. Realizamos o teste Canova Hansen (CH) usado principalmente para testar diferenças sazonais e para validar a hipótese nula de que o padrão sazonal é estável durante um período de amostra ou está mudando ao longo do tempo, cujo resultado foi 0. Aplicamos também o teste Osborn, Chui, Smith e Birchenhall (OCSB) usado para determinar se os dados precisam de diferenciação sazonal (componente D de P, D, Q, m), sendo m igual a

12, considerando que os dados são mensais, o resultado para o teste foi 0, revelando que a série é constante.

Identificamos o valor da Probabilidade de *default* pelos fatores de variáveis fora do modelo utilizado para estimá-la, por meio das variáveis macroeconômicas listadas no Quadro 8, Apêndice A.

Figura 2 – Série Probabilidade de *default*



Fonte: Elaborada pela autora (2022)

No que se refere às séries macroeconômicas utilizamos as séries disponíveis nos sites do Banco Central do Brasil (BCB), do Instituto de Pesquisas Econômicas Aplicadas (IPEA) e da Fundação Getúlio Vargas (FGV), com indicadores de emprego, dados de inflação, produto interno bruto (PIB), taxa de juros, renda, endividamento, concessões e inadimplência de crédito, *spread* das operações de crédito, índice de preços do mercado imobiliário e valores dos insumos e serviços necessários às obras e serviços de engenharia. A primeira seleção conta com 77 variáveis, das quais selecionamos 20, após análise gráfica, listadas no apêndice A quadro 8. A revisão bibliográfica é utilizada para apoiar a seleção.

A seleção partiu da análise do comportamento das variáveis explicativas com relação à variável dependente ou resposta. Demonstramos adiante os gráficos da evolução no tempo das variáveis explicativas e da probabilidade de *default*.

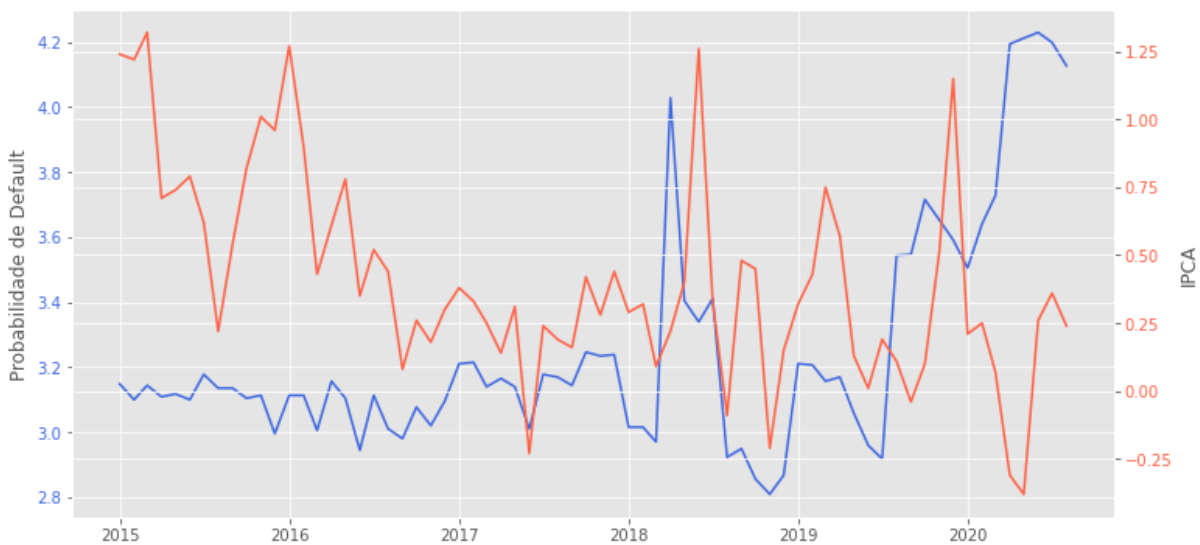
Selecionam-se as variáveis a partir de janeiro de 2015 até agosto de 2020, considerando o período de abrangência da variável dependente. O período abrange a recuperação da economia após a crise de 2015.

Em 2015, o aumento das expectativas inflacionárias e a queda no valor da demanda agregada, em especial no investimento privado e no consumo das famílias, levou o Brasil

à profunda recessão. No último trimestre de 2016, a economia ainda apresentava lenta recuperação e registrava níveis de desemprego elevados (NASSIF; FEIJO; ARAUJO, 2020).

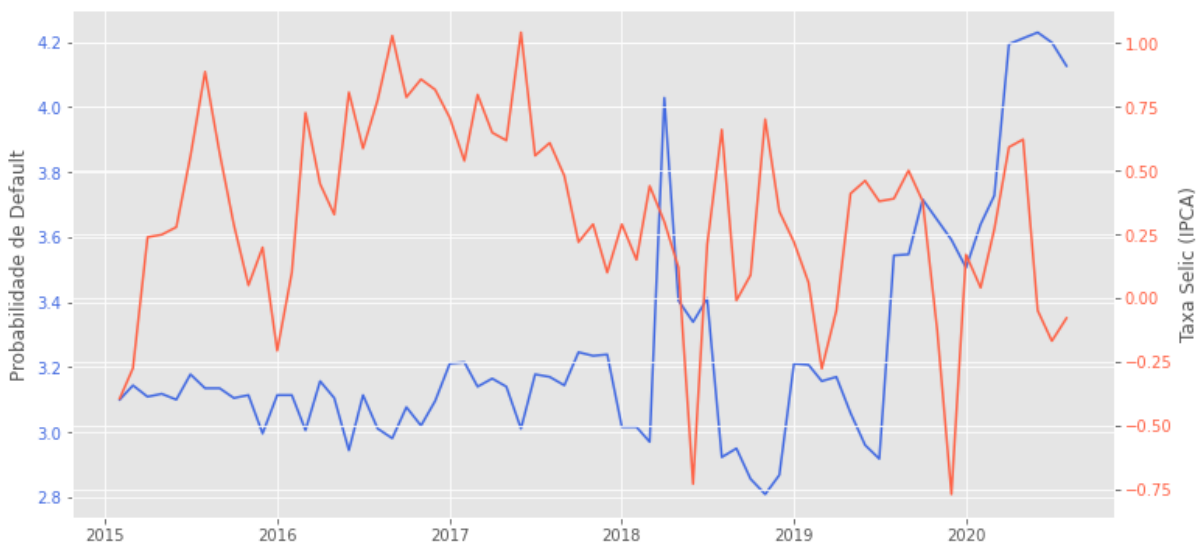
O que explica, nos anos de 2015 a 2016, o registro de alta do índice de inflação ao consumidor (IPCA), da taxa de juros real e do endividamento das famílias com o Sistema Financeiro Nacional. No crédito imobiliário para pessoa física, há queda na concessão com recursos direcionados, crescimento na taxa média de juros e no *spread* médio e tendência de crescimento na inadimplência da carteira de crédito. Como mostra respectivamente as Figuras 3, 4, 5, 6, 7 e 8.

Figura 3 – Evolução Temporal: Probabilidade de *default* e IPCA



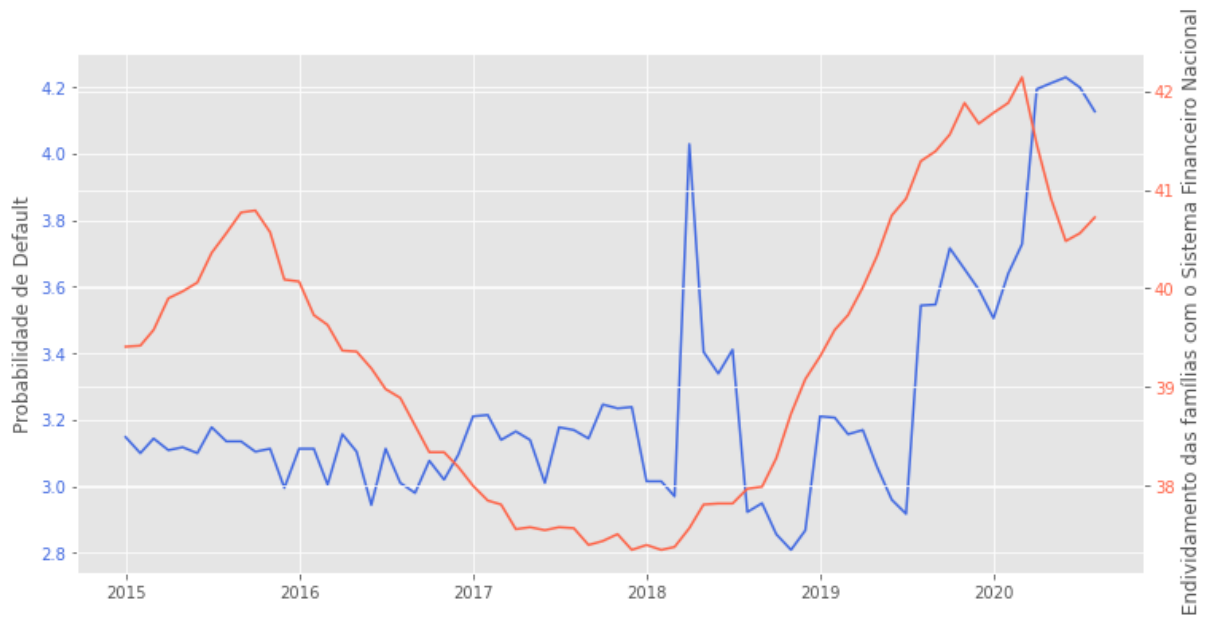
Fonte: Elaborada pela autora (2022)

Figura 4 – Evolução Temporal: Probabilidade de *default* e Taxa de Juros Real (Selic)



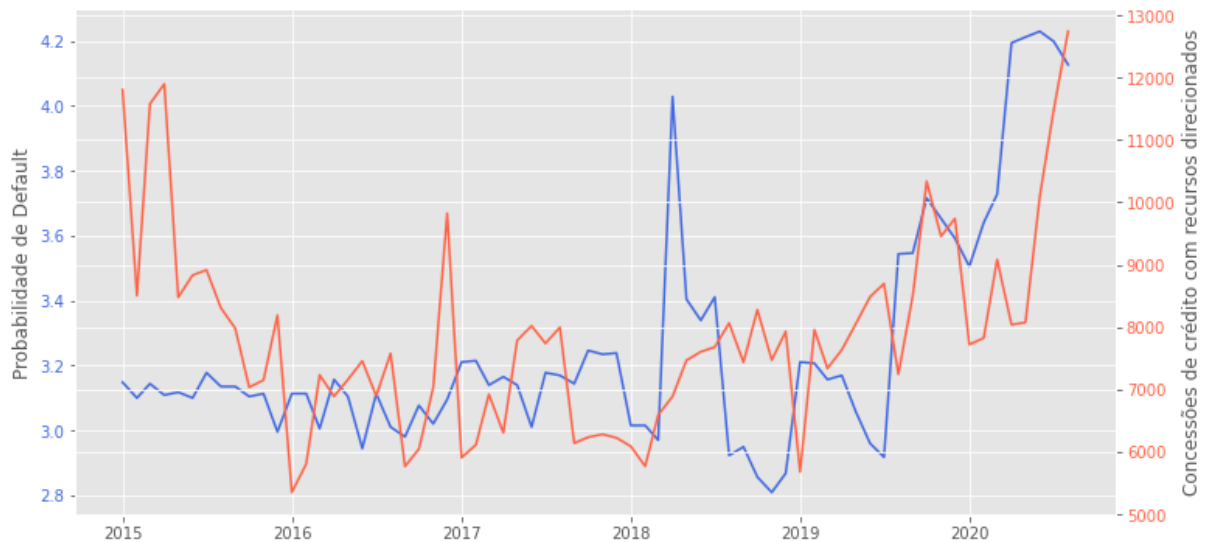
Fonte: Elaborada pela autora (2022)

Figura 5 – Evolução Temporal: Probabilidade de *Default* e Endividamento das Famílias com o Sistema Financeiro Nacional em Relação à Renda Acumulada dos Últimos Doze Meses (RNDBF)



Fonte: Elaborada pela autora (2022)

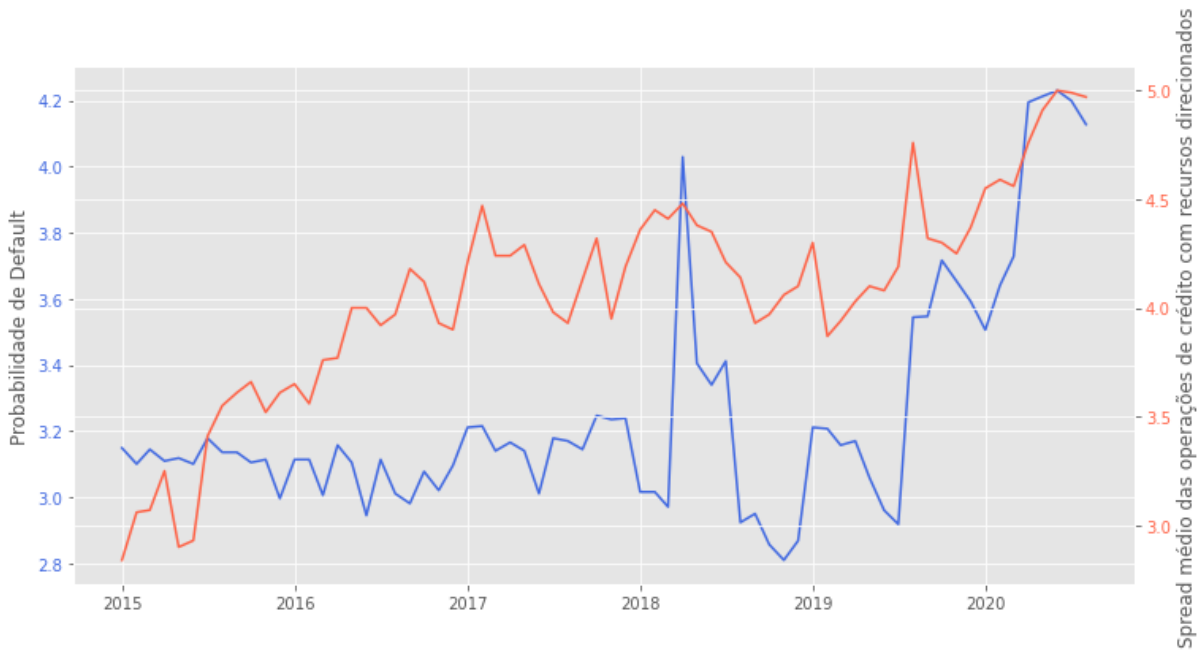
Figura 6 – Evolução Temporal: Probabilidade de *Default* e Concessões de Crédito com Recursos Direcionados - Pessoas físicas - Financiamento imobiliário Total



Fonte: Elaborada pela autora (2022)

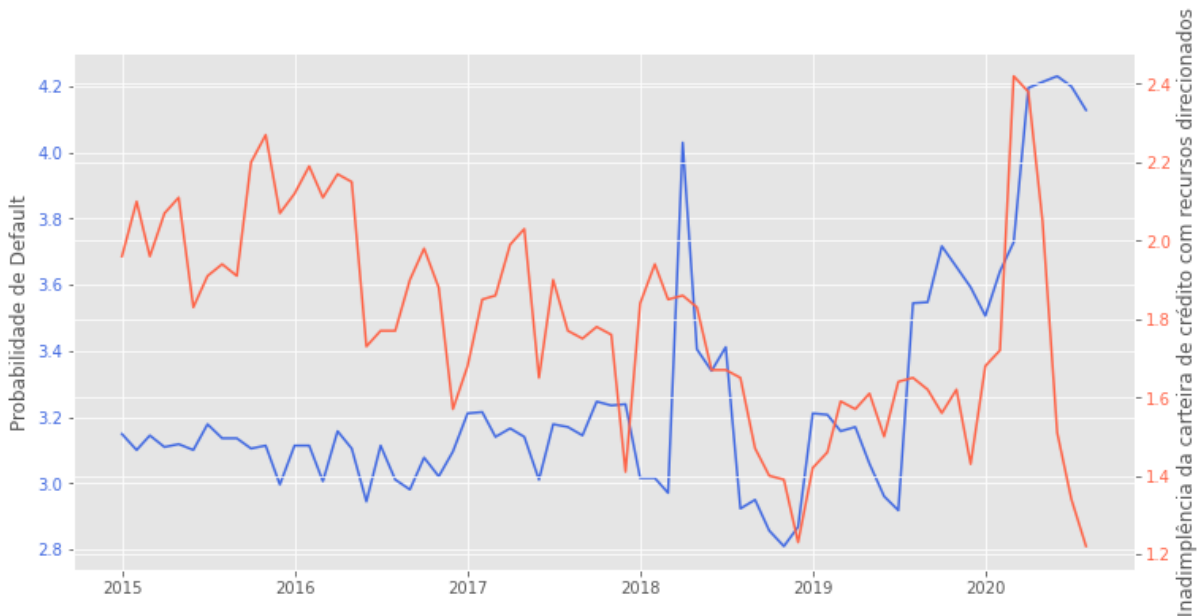
Ao final do ano de 2016, o Congresso Nacional aprova a Emenda Constitucional 95/2016 que define uma regra fiscal para o teto de gasto, o que fica conhecido como Novo Regime Fiscal. A causa raiz, defendida pelo Governo de Michel Temer, seria o quadro fiscal deteriorado em razão do acelerado crescimento da despesa, o que implica elevados prêmios de risco, perda de confiança, juros altos e efeito negativos na atividade. A implementação

Figura 7 – Evolução Temporal: Probabilidade de *Default* e *Spread* Médio das Operações de Crédito com Recursos Direcionados - Pessoas físicas - Total



Fonte: Elaborada pela autora (2022)

Figura 8 – Evolução Temporal: Probabilidade de *Default* e Inadimplência da Carteira de Crédito com Recursos Direcionados - Pessoas físicas - Financiamento Imobiliário Total



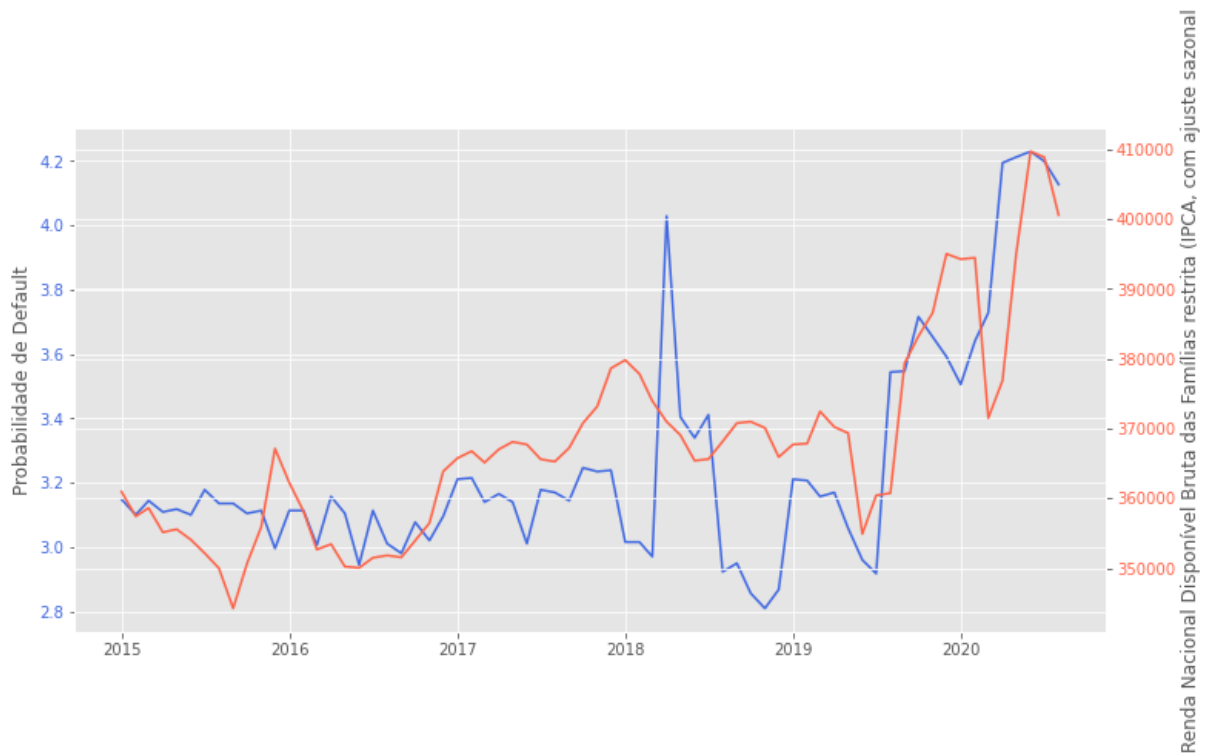
Fonte: Elaborada pela autora (2022)

do teto de gastos colabora para a percepção de menor risco fiscal, com impactos nas taxa de juros (TINOCO; GIAMBIAGI, 2018).

No ano de 2017 o Brasil obteve rápida redução da inflação, com a convergência do IPCA e expectativas de inflação para níveis abaixo da meta (NASSIF; FEIJO; ARAUJO, 2020).

Conforme registra-se na Figura 9 a renda nacional disponível bruta das famílias apresenta elevação a partir de 2017 e em 2018 um grande crescimento do endividamento das famílias com do Sistema Financeiro Nacional (SFN), conforme Figura 5. No crédito imobiliário para pessoa física, registra-se queda na taxa média de juros das operações de crédito com recursos direcionados até o período final da série e na inadimplência até o final de 2018, conforme Figura 8.

Figura 9 – Evolução Temporal: Probabilidade de *Default* e Renda Nacional Disponível Bruta das Famílias Restrita- em valores constantes, deflacionada pelo IPCA, com ajuste sazonal (média móvel trimestral)

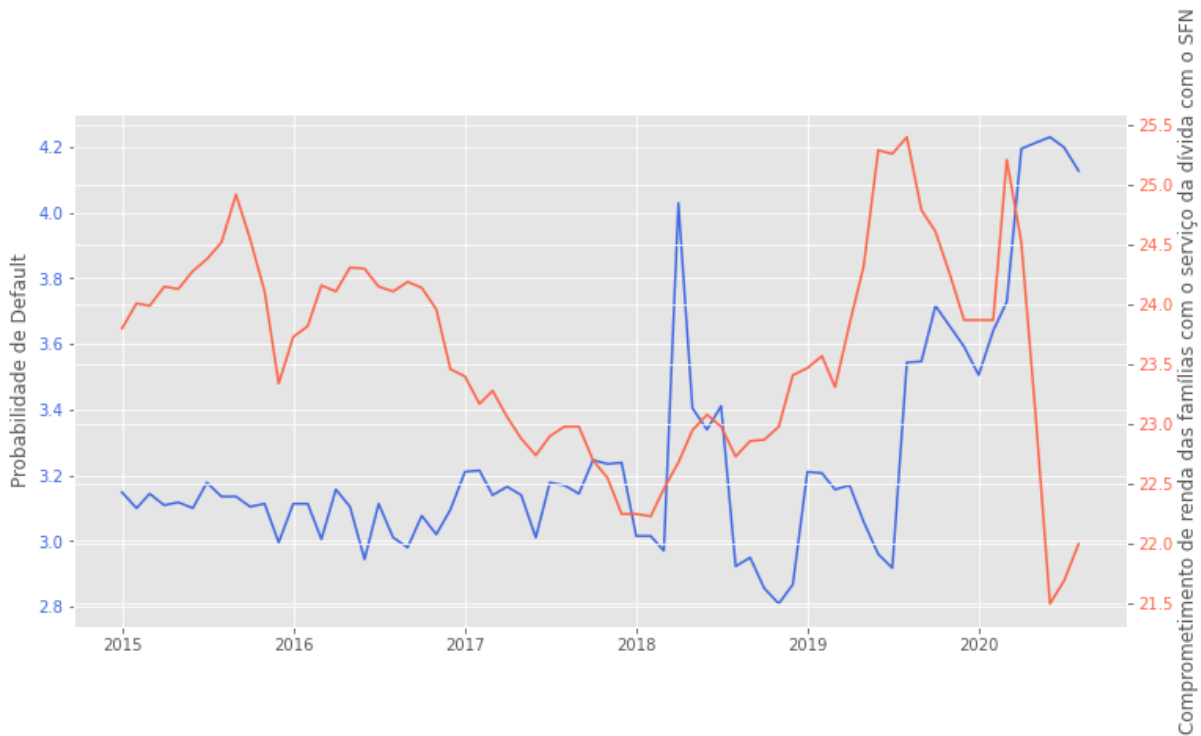


Fonte: Elaborada pela autora (2022)

Dos dados de emprego observa-se crescimento no número de pessoas ocupadas e a partir de 2017 até início de 2020. O rendimento médio real das pessoas ocupadas, no setor público e privado, apresenta tendência de crescimento entre os anos de 2017 a 2020 (Figuras 11 e 15).

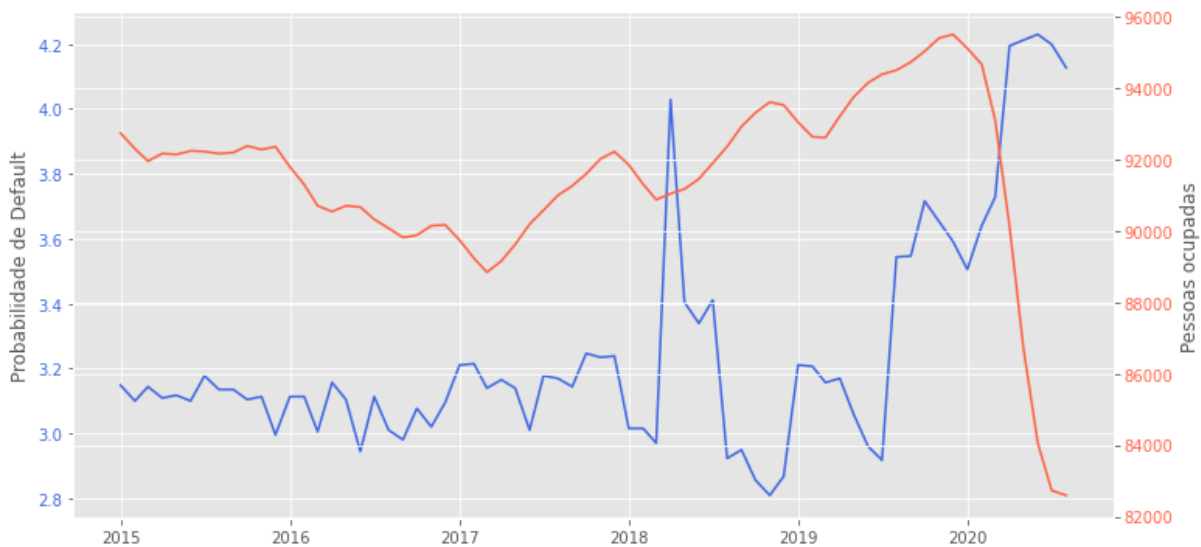
O ano de 2020 representa uma ruptura temporal com a política macroeconômica adotada pelo Brasil, desde 1999, relacionada ao tripé macroeconômico que combina metas de inflação, metas de superavit fiscal primário e regime cambial (NASSIF; FEIJO; ARAUJO, 2020). A pandemia SARS-CoV-2 (Covid-19) fez com que o Brasil introduzisse medidas para

Figura 10 – Evolução Temporal: Probabilidade de *default* e Comprometimento de renda das Famílias com o Serviço da Dívida com o Sistema Financeiro Nacional - Com ajuste sazonal (RNDBF)



Fonte: Elaborada pela autora (2022)

Figura 11 – Evolução Temporal: Probabilidade de *Default* e Pessoas ocupadas - PNADC



Fonte: Elaborada pela autora (2022)

mitigar o choque na economia. O que permitiu ao Banco Central comprar ações públicas e títulos privados no mercado secundário e ao Tesouro Nacional assumir investimentos para o combate à pandemia, conforme autorizado pela Emenda Constitucional 20/10 de

03/03/2020.

Em abril de 2020, publica-se a lei nº 13.982 e o decreto nº 10.316, que estabelecem medidas excepcionais de proteção social a serem adotadas durante o período de enfrentamento da emergência de saúde pública de importância internacional decorrente do coronavírus (Covid-19). Os regramentos definem, dentre outros, o pagamento de auxílio emergencial, no valor de R\$ 600,00, concedido pelo período de três meses, contado da data de publicação da Lei nº 13.982, de 2020.

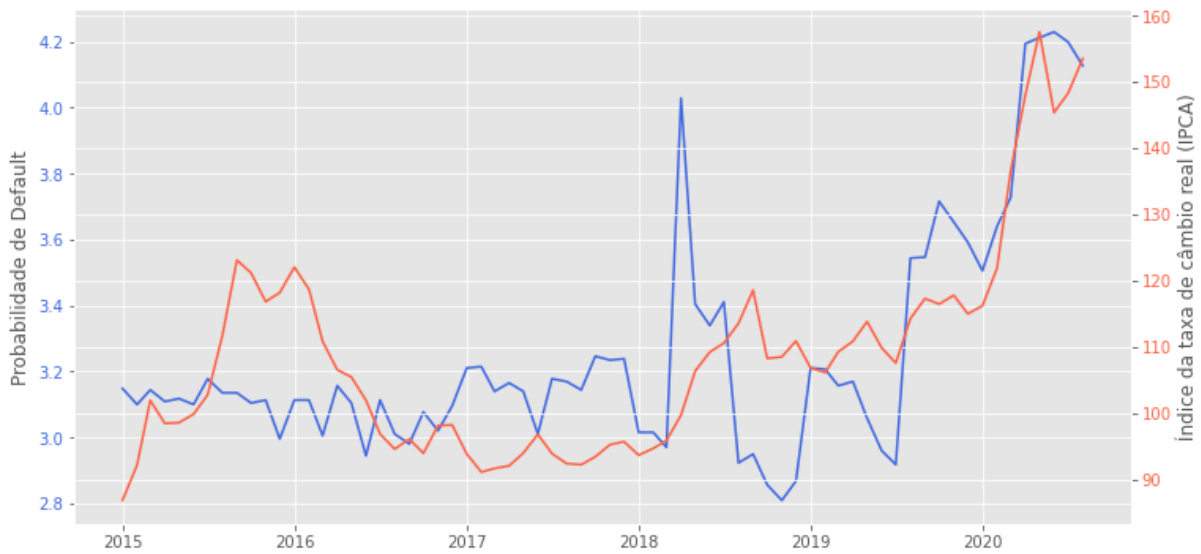
Houve prorrogação do auxílio emergencial para 7 meses, de setembro a dezembro de 2020, o Auxílio Emergencial Extensão pagou mais quatro parcelas com a metade do valor, R\$ 300,00 (R\$ 600,00 para mães solteiras). O programa foi retomado em abril de 2021 em resposta à segunda onda da pandemia de Covid-19, com parcelas entre R\$ 150,00 e R\$ 375,00. A princípio seriam cinco parcelas, mas a lei que autorizou o auxílio emergencial em 2021 permitia a prorrogação por mais três parcelas de igual valor, o que acabou acontecendo (MÁXIMO, 2021).

Em 16 de março de 2020, os cinco maiores bancos atuantes no país, BANCO DO BRASIL S.A., BRADESCO S.A, Caixa Econômica Federal, ITAÚ UNIBANCO S.A e SANTANDER S.A, anunciaram a possibilidade de prorrogação, por 60 dias, dos vencimentos de dívidas de clientes pessoas físicas e micro e pequenas empresas para os contratos vigentes em dia, limitados aos valores já utilizados. Em contratos habitacionais, as prorrogações e/ou pausas nos pagamentos chegaram à seis meses (DIEESE, 2020).

A crise financeira em 2020 pode ser observada no aumento expressivo do índice da taxa de câmbio real (Figura 12). As medidas de mitigação do choque na economia refletem-se nas análises gráficas, observa-se que a partir do primeiro trimestre de 2020, no crédito imobiliário para pessoa física, ocorre uma queda expressiva da inadimplência da carteira de crédito com recursos direcionados, depois de apresentar forte crescimento a partir do último trimestre de 2019 (Figura 8).

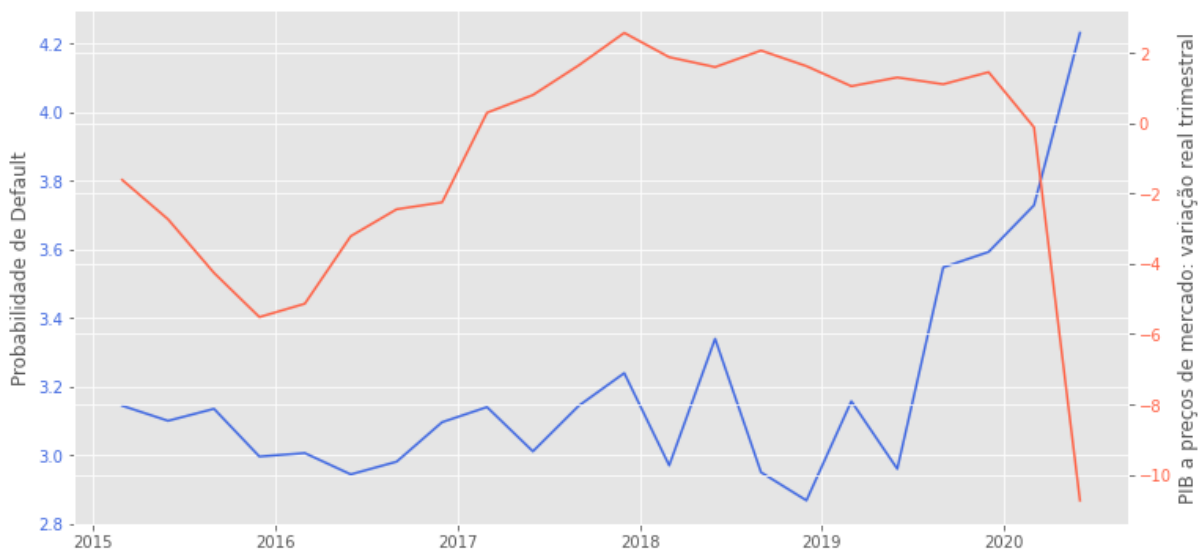
Quanto aos dados de emprego, ainda que o número de pessoas ocupadas tenha apresentado queda significativa no primeiro trimestre de 2020, há aumento no rendimento médio habitual, o que pode ser explicado pelo início do pagamento do auxílio emergencial e maior disponibilidade de recursos na economia, mantendo e gerando empregos (Figura 14 e 15). Registra-se ainda o choque no produto, no primeiro trimestre de 2020 (Figura 13).

Figura 12 – Evolução Temporal: Probabilidade de *Default* e Índice da taxa de câmbio real (IPCA) - Dólar Americano



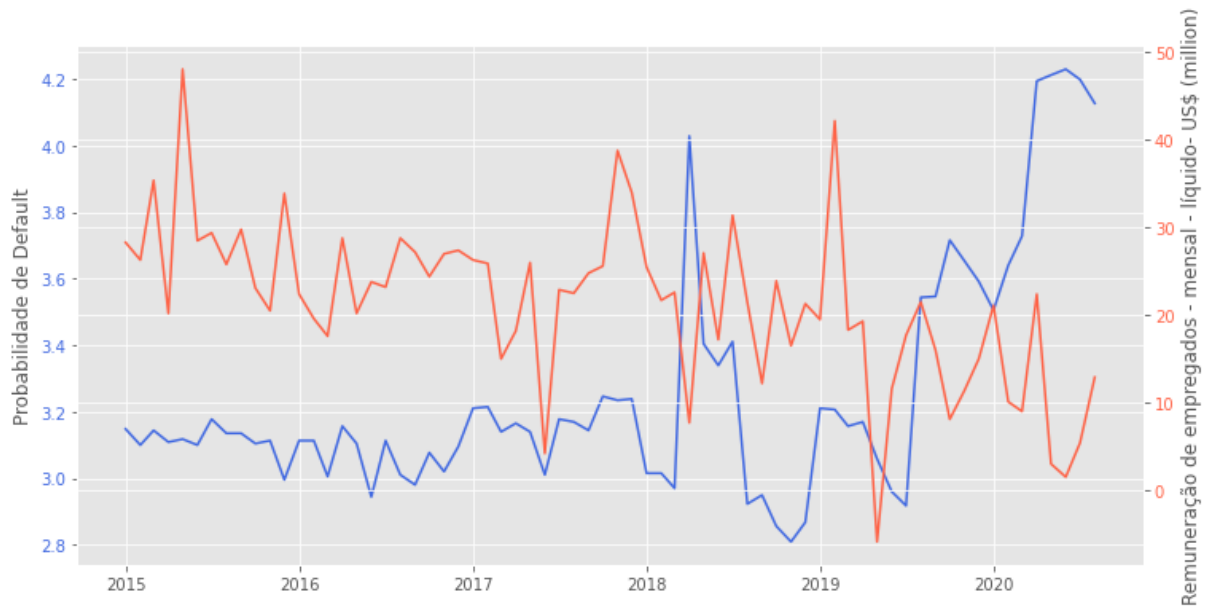
Fonte: Elaborada pela autora (2022)

Figura 13 – Evolução Temporal: Probabilidade de *Default* e PIB a Preços de Mercado com Variação Real Trimestral sobre mesmo trimestre do ano anterior



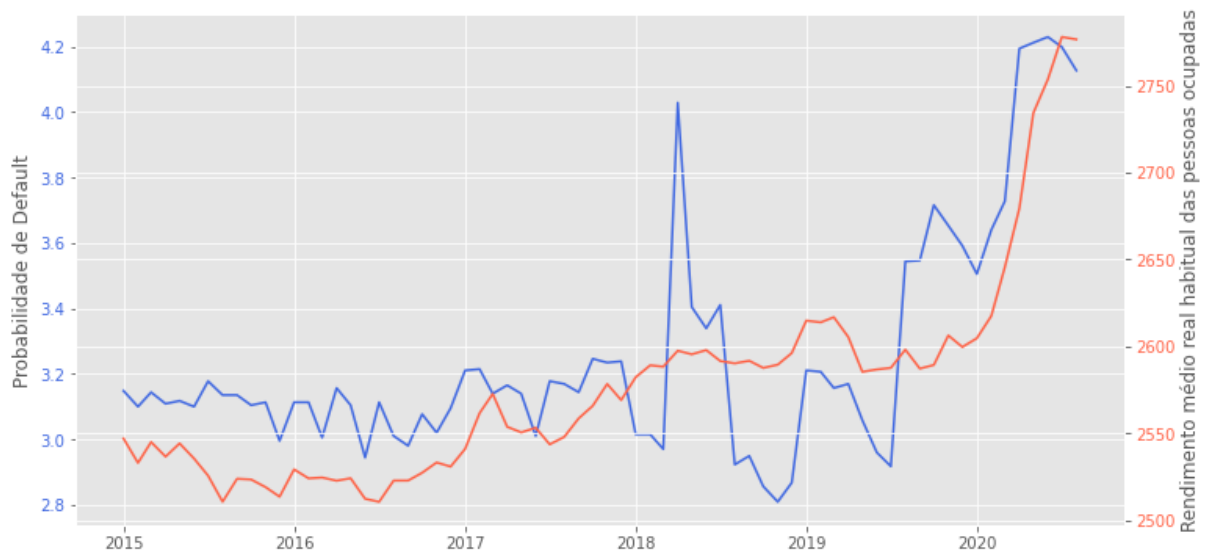
Fonte: Elaborada pela autora (2022)

Figura 14 – Evolução Temporal: Probabilidade de *Default* e Remuneração de Empregados - mensal - líquido



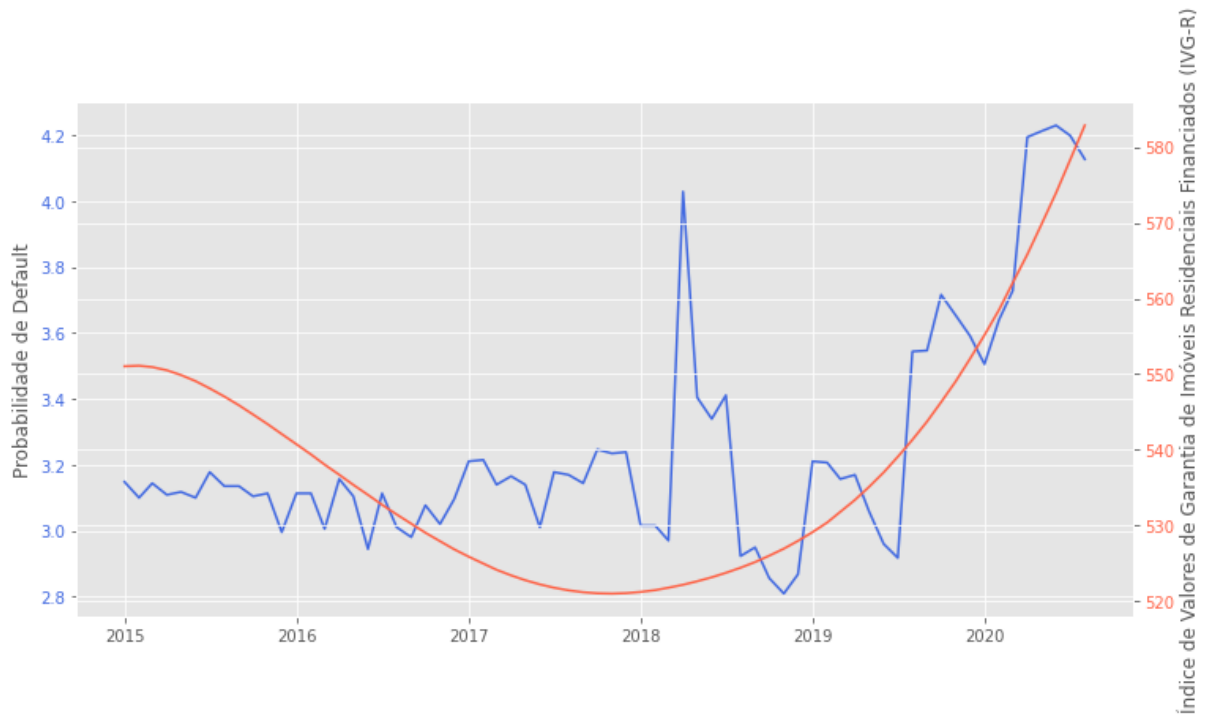
Fonte: Elaborada pela autora (2022)

Figura 15 – Evolução Temporal: Probabilidade de *Default* e Rendimento Médio Real Habitual das Pessoas Ocupadas -Setor privado e público - PNADC



Fonte: Elaborada pela autora (2022)

Figura 16 – Evolução Temporal PD: Índice de Valores de Garantias de Imóveis Residenciais Financiados (IVGR)



Fonte: Elaborada pela autora (2022)

3 Metodologia

Nessa seção, mostramos o método utilizado para fazer a inferência das variáveis macroeconômicas na variável dependente Probabilidade de *default*. Apresentamos o modelo e então descrevemos os procedimentos utilizados para validá-lo. Faz-se uso da linguagem de programação Python, utilizando os pacotes Pandas, Numpy, Statsmodel e Pmdarima.

A análise da série temporal com variáveis explicativas abrange métodos para modelar e prever dados correlacionados levando em conta informações adicionais, conhecidas como variáveis exógenas.

Estudos de Box e Jenkins (2008) introduzem os modelos de séries temporais integradas auto-regressivas (ARIMA) para caracterizar e prever observações de séries temporais em períodos igualmente espaçados. O modelo criado em 1970 foi revisado em 1976.

Existem três métodos diferentes derivados do modelo ARIMA, quais sejam *Auto-Regressive* (AR), *Moving Average* (MA) e *Auto-Regressive Moving Average* (ARMA). Os modelos AR são aplicados a uma série temporal que pode ser representada por seus próprios valores passados, enquanto os modelos MA são usados para representar uma série temporal onde os erros passados (distúrbios) determinam seus valores futuros. Os modelos ARMA são apropriados quando uma série é uma função de erros não observados, bem como seu próprio comportamento passado.

O modelo *Auto-Regressive* (AR), visa capturar a autocorrelação da série. O modelo *Auto-Regressive* é definido como AR(p), onde p se refere a ordem do componente AR. Conforme Pal e Prakash (2017):

A primeira ordem do modelo AR é denotado por AR(1):

$$\hat{X}_t = \phi\varepsilon_{t-1} + \varepsilon_t \quad (3.1)$$

A segunda ordem do modelo AR é denotado por AR(2):

$$\hat{X}_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \varepsilon_t \quad (3.2)$$

A ordem p do modelo AR é denotado por AR(p):

$$\hat{X}_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \dots + \phi_p\varepsilon_{t-p} + \varepsilon_t \quad (3.3)$$

Sendo ϕ o coeficiente e p a ordem de autoregressão (PAL; PRAKASH, 2017), e $\hat{X}_t = X_t - \mu$ (BOX; TIAO, 1975).

O modelo *moving average* (MA) usa a dependência entre os erros residuais para prever valores no próximo período de tempo. Conforme Pal e Prakash (2017):

A primeira ordem do modelo MA é denotado por MA(1):

$$\hat{X}_t = \alpha - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (3.4)$$

A segunda ordem do modelo MA é denotado por MA(2):

$$\hat{X}_t = \alpha - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \varepsilon_t \quad (3.5)$$

A ordem q do modelo MA é denotado por AR(q):

$$\hat{X}_t = \alpha - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.6)$$

Aqui, ε_t é o erro identicamente distribuído de forma independente no tempo t e segue distribuição normal $N(0, \sigma_\varepsilon^2)$ com media zero e variância σ_ε^2 . O componente ε_t representa o erro no tempo t , sendo a notação α o intercepto médio e ε o coeficiente dos erros.

Segundo Pal e Prakash (2017), o modelo ARMA (p, q) incorpora a ordem p AR e a ordem q MA.

O modelo ARMA (1,1) é representado como segue:

$$\chi_t = \alpha + \phi_1 \chi_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (3.7)$$

O modelo ARMA (1,2) é representado como segue:

$$\chi_t = \alpha + \phi_1 \chi_{t-1} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \varepsilon_t \quad (3.8)$$

O modelo ARMA (p, q) é representado como segue:

$$\chi_t = \alpha + \phi_1 \chi_{t-1} + \phi_2 \chi_{t-2} + \dots + \phi_p \chi_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.9)$$

Aqui, ϕ e θ representam o coeficiente AR e MA, respectivamente. O α e ε_t capturam o intercepto e o erro no tempo t (PAL; PRAKASH, 2017).

O modelo ARIMA também conhecido como modelo Box-Jenkins, é a generalização do modelo ARMA com a inclusão de componentes de integração. É aplicado à séries que não estão estacionárias ao longo do tempo, ou seja, com alguma tendência, e por isso precisam ser diferenciadas para ficarem estacionárias, com média, variância e autocorrelação constantes. O ARIMA (p, q, d) representa a ordem AR, MA, e o componente de diferenciação. Para diferentes valores de d , a resposta da série muda da seguinte forma (PAL; PRAKASH, 2017):

Para $d=0$: $\chi_t = \chi_t$

Para $d=1$: $\chi_t = \chi_t - \chi_{t-1}$

Para $d=2$: $\chi_t = (\chi_t - \chi_{t-1}) - (\chi_{t-1} - \chi_{t-2}) = \chi_t - 2\chi_{t-1} + \chi_{t-2}$

O modelo ARIMA (p, q, d) é representado como segue:

$$\hat{\chi}_t = \phi_1 \hat{\chi}_{t-1} + \phi_2 \hat{\chi}_{t-2} + \dots + \phi_p \hat{\chi}_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.10)$$

Esta equação pode ser feita de forma mais concisa através do uso do operador lag, L.

$$L^n \hat{\chi}_t = \hat{\chi}_{t-n} \quad (3.11)$$

Para ajudar a lidar com dados não estacionários, introduz-se um operador de integração Δ^d , definido da seguinte forma (CHATFIELD, 2004):

$$\hat{\chi}_t^{[d]} = \Delta^d \hat{\chi}_t = \hat{\chi}_t^{[d-1]} - \hat{\chi}_{t-1}^{[d-1]} \quad (3.12)$$

onde $\hat{\chi}_t^{[d]} = \hat{\chi}_t$ e d é a ordem da diferenciação utilizada.

O que é equivalente ao modelo ARIMA p, d, q , tomando $\Theta(L)^p$ como uma função polinomial de ordem p de L, θ é definido de forma análoga a Θ (CHATFIELD, 2004):

$$\Delta^d \hat{\chi}_t = \Theta(L)^p \Delta^d \hat{\chi}_t + \phi(L)^q \Delta^d \varepsilon_t + \Delta^d \varepsilon_t \quad (3.13)$$

Podemos reorganizar a equação e absorver constantes nos polinômios θ é definido de forma análoga a Θ :

$$\Theta(L)^p \Delta^d \hat{\chi}_t = \theta(L)^q \Delta^d \varepsilon_t \quad (3.14)$$

A metodologia ARIMAX, desenvolvida por Box e Tiao (1975), caracteriza-se por uma expansão da metodologia Box-Jenkins. O modelo ARIMAX acrescenta um componente linear em função das observações das covariáveis ou variáveis exógenas. A principal diferença entre os dois modelos é que o ARIMAX possui, além dos parâmetros auto-regressivos e de médias móveis, a entrada exógena (BENNETT; STEWART; LU, 2014). O modelo ARIMAX pode ser entendido como a combinação dos modelos Auto-Regressivo AR (p), Integrado (d), Média-Móvel MA (q) e Exógeno X(r), e que pode então ser simbolizado por ARIMAX (p, d, q, r) .

Os modelos ARIMAX levam em consideração variáveis exógenas, variáveis medidas no tempo t que influenciam o valor da série temporal no tempo t , mas que não são autorregressadas. Para fazer isso, adiciona-se os termos no lado direito da equações ARIMA e SARIMA. Para n variáveis exógenas definidas em cada passo de tempo t , denotadas por x_t^i para $i \leq n$, com coeficientes β_i , o modelo ARIMAX (p, d, q, r) é definido por:

$$\Theta(L)^p \Delta^d \hat{\chi}_t = \theta(L)^q \Delta^d \varepsilon_t + \sum_{i=1}^n \beta_i x_t^i \quad (3.15)$$

O modelo SARIMAX é um modelo ARIMA com adição de uma ou mais variáveis exógenas e de mais três parâmetros, P, D e Q, referentes à autoregressão, diferenciação e ordem da média móvel da componente sazonal, sendo S o intervalo de tempo do padrão sazonal de repetição. Um modelo ARIMA com regressores externos e componente sazonal é referido como SARIMAX (p,d,q)(P,D,Q), onde p é a ordem de autocorrelação AR, d, o grau de diferença, e q a ordem da parte média móvel, sendo P,D e Q, a parte sazonal do modelo.

Segundo Box Jenkins et al. 2008 quando uma série exibe comportamento sazonal com conhecida periodicidade S, em dados periódicos não são um, mas dois intervalos de tempo de importância. Para o exemplo de sazonalidade anual, esses intervalos correspondem a meses e anos. Espera-se que os relacionamentos ocorram entre as observações para meses sucessivos em um determinado ano e entre as observações para o mesmo mês em anos sucessivos. O efeito sazonal implica que uma observação para um determinado mês, digamos abril, está relacionada com as observações para abril anterior.

Agora as componentes de erro $\varepsilon_t, \varepsilon_{t-1}$, nestes modelos não seriam em geral não correlacionados. No exemplo citado por Box Jenkins et al. (2008), o total de abril, enquanto relacionados aos totais de abril anteriores, também estariam relacionados aos totais de março, fevereiro, janeiro e assim por diante.

Assim, esperaríamos que ε_t estaria relacionado a ε_{t-1} e a ε_{t-2} , e assim por diante. Sendo ε_t um processo de ruído branco $\theta(L)$ e $\Theta(L)$ são polinômios em L, de graus p e q, respectivamente, e satisfazendo as condições de estacionariedade e invertibilidade, e $\Delta = \Delta_1 = 1L$.

Resultando no modelo SARIMA (p,d,q) (P,D,Q)S (BOX et al., 2008) introduzindo o componente exógeno X:

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D \hat{\chi}_t = \theta(L)^q \phi(L^s)^Q \varepsilon_t + \sum_{i=1}^n \beta_i x_t^i \quad (3.16)$$

O modelo implica que devemos tomar “d” diferenças simples e “D” diferenças sazonais da série $\hat{\chi}_t$.

O trabalho de Maçaira et al. (2018), que resulta na identificação, em trabalhos acadêmicos, dos principais métodos de séries temporais que incorporam variáveis de insumo, por área de conhecimento e metodologia, identifica 30 tipos diferentes de métodos presentes nos documentos principais. Os métodos aplicados com maior frequência são os modelos de regressão, seguidos por redes neurais artificiais, a metodologia Box e Jenkins’s para modelo

ARIMA com a incorporação de variável explicativa (ARIMAX), máquinas vetoriais de suporte (SVMs) e modelos estruturais, que juntos estão presentes em 70% dos trabalhos.

O SARIMAX exige especificação mais detalhada da composição do modelo. Sendo assim, antes de determinar a estimativa dos modelos da série temporal, realizamos o teste dickey-fuller aumentado (ADF) (DICKEY; FULLER, 1979), para determinar se as séries temporais do *dataset* seguem um projeto estacionário estocástico ao longo do tempo. A hipótese nula do teste ADF propõe que $H_0 : \rho = 0$ contra a hipótese alternativa $H_0 : \rho < 0$. A hipótese nula é a presença de raiz unitária ou não estacionária enquanto a hipótese alternativa sugere estacionariedade da série.

Os resultados dos testes ADF indicam que nem todos os dados são originalmente estacionários. Por esse motivo diferenciamos nossas variáveis, incluindo as exógenas. Após a diferenciação, concluímos que a hipótese nula é rejeitada.

Quadro 1 – Teste de raiz unitária.

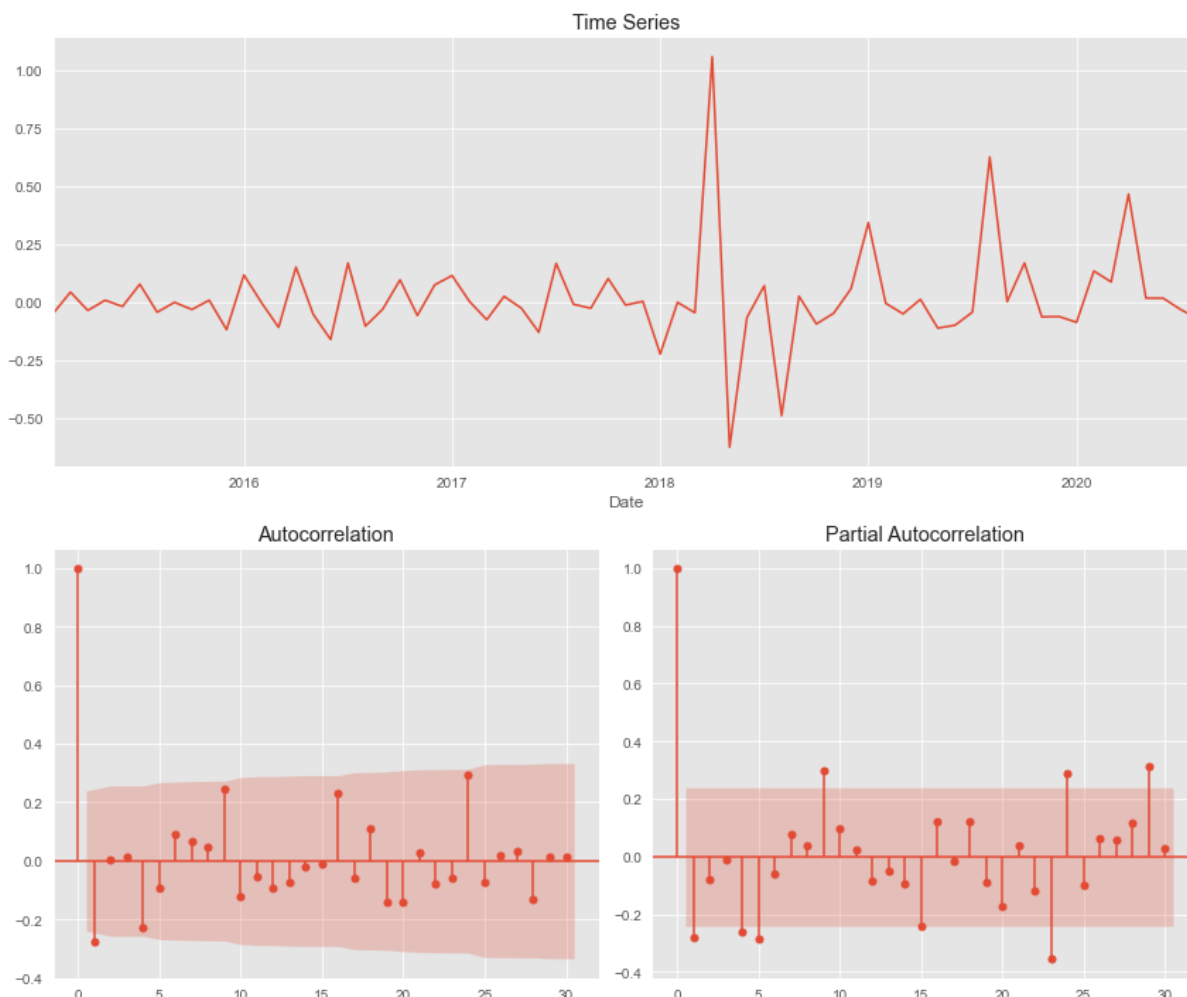
Teste ADF	Correlação	IC	p-value	IC Lag 2	p-value
PD	1,0	<90%	0.724234	>99%	0,000001
433	-0,2	<99%	0.000260		
4380	0,4	<90%	0,898270	>99%	0,000000
11753	0,7	<90%	0,898113	>99%	0,000000
4390	-0,6	<99%	0,000441		
29028	0,7	<90%	0.988543	>99%	0,000000(Lag3)
29034	-0,1	<90%	0,168383	> 99%	0,000000 (Lag3)
29037	0,4	>95%	0,037399	>95%	0,022514
20704	0,4	>99%	0,004331	>99%	0,000000
21151	-0,1	<90%	0,613859	>99%	0,000859
25499	-0,5	<90%	0,937398	>99%	0,000000
21340	0,6	<90%	0.741316	<90%	0,167323(Lag3)
20837	0,6	<90%	0,250574	>99%	0,000000
22803	-0,2	>95%	0,021583	>99%	0,000000
24377	-0,5	>90%	0,376393	>99%	0,000011 (Lag3)
28763	-0,2	>90%	0,668833	>99%	0,000056 (Lag3)
24379	-0,4	<90%	0,123288	>99%	0,000098
24380	0,3	<90%	0.606587	>99%	0.000000 (Lag3)
24399	0,8	<90%	0,996205	>99%	0,000001
21803	0,6	<90%	0,992340	>99%	0,000000
7495	-0,1	>99%	0,000007		

Fonte: Elaborada pela autora (2022)

O método de determinação dos termos sazonais (P, D, Q) e não-sazonais (p, d, q) baseia-se principalmente no critério de informação de Akaike, ou AIC (Akaike's Information Criterion). Além disso, também observamos as parcelas da função de autocorrelação (ACF) e da função de correlação automática parcial (PACF) da variável dependente, que nos

ajudam a determinar se as ordens identificadas pelo modelo estão condizentes com as parcelas de autocorrelação.

Figura 17 – Parcelas da função de autocorrelação (ACF) e autocorrelação parcial (PACF) depois da primeira diferenciação



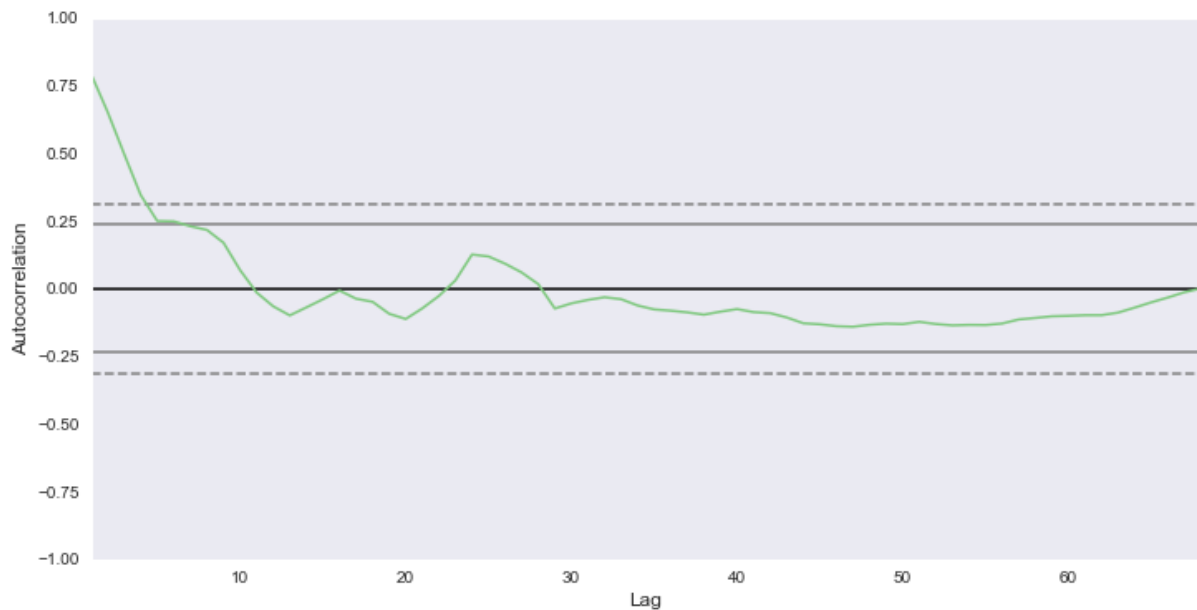
Fonte: Elaborada pela autora (2022)

Para determinar o termo de diferença sazonal analisamos o gráfico de autocorrelação da série dependente antes da diferenciação (Figura 18), onde atribui o resultado de 12 lags quando a série atinge autocorrelação 0 pela primeira vez.

Para a seleção das variáveis realizamos o teste de causalidade de Granger. Referido teste tem o objetivo de determinar o sentido causal de duas variáveis, podendo reforçar ou refutar a primeira correlação encontrada. Há causalidade se valores passados de uma variável X ajudam a prever o valor presente de outra variável Y.

Consideramos o p-valor da variável independente com a Probabilidade de *Default*. O p-valor menor que 0,05, rejeita-se a Hipótese Nula, ou seja, é verdade que variável independente influencia a PD (Quadro 2), que resultou nas variáveis taxa de juros reais, Renda Nacional Disponível Bruta das Famílias, Taxa média mensal de juros das operações

Figura 18 – Autocorrelação da variável dependente



Fonte: Elaborada pela autora (2022)

de crédito, Pessoas desocupadas, Rendimento médio real habitual, RWA para risco de crédito e SINAPI.

Utilizamos no modelos as variáveis exógenas que apresentaram análise gráfica condizente com a Probabilidade de *default*. Para avaliar se, de fato, o uso dessas variáveis poderiam acrescentar melhorias no cálculo da Probabilidade de *default*, utilizamos o método SARIMAX para verificar a significância dos coeficientes. A seleção das variáveis exógenas foi realizada com base em uma análise para discernir a resposta e uma abordagem de regressão.

As variáveis são adicionadas ao modelo e a significância estatística é calculada, por meio do teste de hipótese bicaudal. Se a variável tem a probabilidade de significância (valor-p) menor que o nível de significância (α) 0,05, rejeitamos a hipótese nula e a consideramos como estatisticamente significativa.

A hipótese nula (H_0), indica que a variável macroeconômica não tem impacto na variação da Probabilidade de *default*. Enquanto a hipótese alternativa (H_1) fornecerá a hipótese de impacto da variável macroeconômica na Probabilidade de *default*.

Consideramos o valor crítico do z igual a 1,645 para 5% bicaudal, que significa que a probabilidade de estar abaixo de 1,645, desde os padrões acima da média vai ser de 95%, o que vale para qualquer distribuição normal, quando o valor de z estiver acima de 1,645 terá probabilidade menor que 5%.

Para o diagnóstico da análise utilizamos os testes de autocorrelação dos resíduos, resíduos padronizados ao longo do tempo, histograma mais densidade estimada de resíduos

Quadro 2 – Causalidade de Granger

Variável	Lag1	Lag2	Lag3
433 Índice nacional de preços ao consumidor-amplo (IPCA)	0.2465	0.1629	0.1831
4380 PIB mensal real	0.5654	0.7507	0.4366
11753 Índice da taxa de câmbio real (IPCA)	0.7337	0.1232	0.3211
4390 Taxa de juros reais - Selic	0.2149	0.0213	0.1244
29028 Renda Nacional Disponível Bruta das Famílias restrita	0.0465	0.1191	0.3480
29034 Comprometimento de renda das famílias com o serviço da dívida com o Sistema Financeiro Nacional	0.1408	0.7374	0.6045
29037 Endividamento das famílias com o Sistema Financeiro Nacional	0.2414	0.3970	0.4516
20704 Concessões de crédito com recursos direcionados	0.5013	0.2460	0.0973
21151 Inadimplência da carteira de crédito com recursos direcionados	0.4996	0.5110	0.5343
25499 Taxa média mensal de juros das operações de crédito com recursos direcionados	0.8621	0.0605	0.0150
21340 Índice de Valores de Garantia de Imóveis Residenciais Financiados (IVG-R)	0.9373	0.7320	0.4929
20837 <i>Spread</i> médio das operações de crédito com recursos direcionados	0.5394	0.2839	0.3085
22803 Remuneração de empregados	0.4858	0.4768	0.4699
24377 Empregados no setor privado	0.8046	0.5288	0.6126
28763 Estoque de empregos formais	0.9578	0.4192	0.4277
24379 Pessoas ocupadas	0.9602	0.4146	0.5835
24380 Pessoas desocupadas	0.0559	0.2615	0.1697
24399 Rendimento médio real habitual das pessoas ocupadas -Setor privado e público	0.1205	0.4261	0.0000
21803 RWA para risco de crédito	0.8935	0.0043	0.0000
7495 SINAPI	0.5122	0.0417	0.0045

Fonte: Elaborada pela autora (2022)

padronizados, juntamente com uma densidade Normal (0,1) traçada para referência e o gráfico QQ normal, com linha de referência normal.

4 Resultados

A função *stepwise* auto arima do programa Python ajusta o melhor modelo ARIMA para uma série temporal univariada de acordo com um critério de informação fornecido (AIC, AICc, BIC ou HQIC). A função realiza uma pesquisa sobre possíveis pedidos de modelo e sazonal dentro das restrições fornecidas e seleciona os parâmetros que minimizam a métrica fornecida. Conforme Quadro 9 o modelo com menor AIC retornou $p=0$, $d=1$, $q=3$ e $P=0$, $D=0$, $Q=0$.

A resposta do modelo representa os parâmetros encontrados ao realizar o teste dickey-fuller aumentado (ADF), sendo $d=1$ e para o teste Canova Hansen (CH) e teste Osborn, Chui, Smith e Birchenhall (OCSB), sendo $D=0$. Os valores de p e q divergem, no entanto, com os valores encontrados quando analisada a autocorrelação parcial e a autocorrelação, onde a análise gráfica resulta em 1 para os valores de p e q .

Sendo assim apresenta-se os resultados para os dois modelos.

O modelo SARIMAX $p=0$, $d=1$, $q=3$ e $P=0$, $D=0$, $Q=0$ apresenta coeficientes e níveis descritivos para as variáveis estatisticamente significantes conforme Quadro 3:

Para avaliar se o modelo assume a suposição de que os resíduos são independentes analisamos os resultados dos testes estatísticos. Os quais resultaram no que segue no Quadro 4 e interpretamos adiante.

O Ljung-Box (L1) (Q) é a estatística de teste LBQ na defasagem 1, sendo o Prob(Q) 1,00 e o valor-p 0,00. Como a probabilidade está acima de 0,05, não podemos rejeitar a hipótese nula de que os resíduos são não correlacionados, então os resíduos são não correlacionados. A heterocedasticidade testa se os resíduos de erro são homocedásticos ou têm a mesma variância. O resumo realiza o teste White. Nossas estatísticas de resumo mostram uma estatística de teste de 4,69 e um valor-p de 0,00, o que significa que rejeitamos a hipótese nula e nossos resíduos mostram variância. Para a normalidade dos erros os testes Jarque-Bera testam o nulo de que os dados são normalmente distribuídos em relação a uma alternativa de outra distribuição. Vemos uma estatística de teste de 295,11 com probabilidade 0,00, o que significa que rejeitamos a hipótese nula e os dados não são normalmente distribuídos. Além disso, como parte do teste Jarque-Bera, vemos que a distribuição tem uma inclinação positiva e uma grande curtose.

Não há padrões óbvios para os resíduos (standardized residuals). O histograma com a estimativa de densidade mostra a distribuição dos resíduos, enquanto a curva KDE, a versão suavizada do histograma, a linha $N(0,1)$ mostra uma distribuição normal. O correlograma apresenta que 95% das correlações para defasagem maior que um não é

Quadro 3 – SARIMAX (0,1,3) (0,0,0) [12]

Variável	coeficiente	P>[z]
433 Índice nacional de preços ao consumidor-amplo (IPCA)	0.0771	0.000*
4380 PIB mensal real	5.18E-07	0.863
11753 Índice da taxa de câmbio real (IPCA)	0.0061	0.639
4390 Taxa de juros reais - Selic	-0.0010	0.947
29028 Renda Nacional Disponível Bruta das Famílias restrita	-2.59E-03	0.701
29034 Comprometimento de renda das famílias com o serviço da dívida com o Sistema Financeiro Nacional	0.0157	0.000*
29037 Endividamento das famílias com o Sistema Financeiro Nacional	-0.0486	0.000*
20704 Concessões de crédito com recursos direcionados	-1.70E-02	0.639
21151 Inadimplência da carteira de crédito com recursos direcionados	0.1238	0.000*
25499 Taxa média mensal de juros das operações de crédito com recursos direcionados	0.5529	0.000*
21340 Índice de Valores de Garantia de Imóveis Residenciais Financiados (IVG-R)	0.0187	0.228
20837 Spread médio das operações de crédito com recursos direcionados	0.1568	0.000*
22803 Remuneração de empregados	-0.0001	0.979
24377 Empregados no setor privado	1.29E-02	0.971
28763 Estoque de empregos formais	-6.50E-05	0.742
24379 Pessoas ocupadas	9.47E-03	0.954
24380 Pessoas desocupadas	-2.72E-02	0.829
24399 Rendimento médio real habitual das pessoas ocupadas -Setor privado e público	0.0032	0.301
21803 RWA para risco de crédito	-6.29e-07	0.550
7495 Sistema Nacional de Pesquisa de Custos e Índices (SINAPI)	-0.1364	0.000*
ma.L1 Correlação serial usando relação de erro no lag 1	-0.4631	0.000*
ma.L2 Correlação serial usando relação de erro no lag 2	-0.1856	0.000*
ma.L3 Correlação serial usando relação de erro no lag 3	-0.1279	0.000*
sigma2 Termo de erro	0.0333	0.000*

Fonte: Elaborada pela autora (2022)

significativa. E a normal Q-Q apresenta parte dos pontos dos dados na linha reta, indicando uma distribuição dos resíduos próxima da normal.

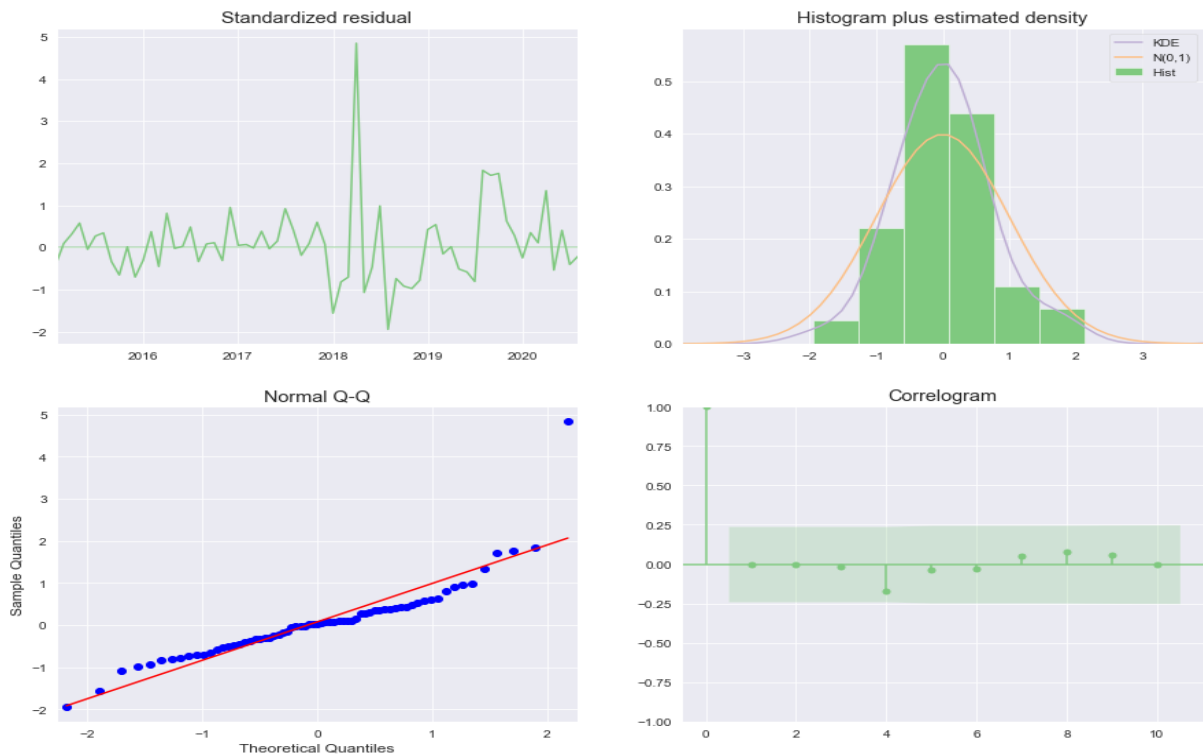
O modelo SARIMAX $p=1, d=1, q=1$ e $P=0, D=0, Q=0$ apresenta coeficientes e níveis descritivos para as variáveis estatisticamente significantes conforme Quadro 5. Observa-se que não há alteração na definição das variáveis significantes, havendo somente o acréscimo do termo auto-regressivo.

Quadro 4 – Revisão de suposições SARIMAX (0,1,3) (0,0,0) [12]

Ljung-Box (L1) (Q): 0.00	Jarque-Bera (JB): 295.11
Prob(Q): 1.00	Prob(JB): 0.00
Heteroskedasticity (H): 4.69	Skew: 2.08
Prob(H) (two-sided): 0.00	Kurtosis: 12.40

Fonte: Elaborada pela autora (2022)

Figura 19 – Análise dos resíduos SARIMAX (0,1,3) (0,0,0) [12]



Fonte: Elaborada pela autora (2022)

Considerando que os resultados do teste de Granger apresentam maior causalidade a partir do terceiro lag, testamos um terceiro modelo, que na segunda diferença apresenta melhor AIC (Quadro 10) no modelo SARIMAX $p=0, d=2, q=2$ e $P=0, D=0, Q=0$.

O resultado da análise dos resíduos para o modelo SARIMAX (2,2,0) (0,1,1) [12] se dá conforme Quadro 7 e interpretamos adiante.

A Prob(Q) de 0,42 maior que 0,05 nós não rejeitamos a hipótese nula de que os resíduos são não correlacionados, então os resíduos são não correlacionados. A Prob(JB) de 0,00 menor 0,05 rejeita-se a hipótese nula de que os resíduos são normalmente distribuídos, então eles não são normalmente distribuídos. Como parte do teste Jarque-Bera, vemos que a distribuição tem uma inclinação positiva e uma grande curtose. Os resultados da heterocedasticidade mostram uma estatística de teste de 2,44 e um valor- p de 0,04, o que significa que rejeitamos a hipótese nula e nossos resíduos mostram variância.

Quadro 5 – SARIMAX (1,1,1) (0,0,0) [12]

Variável	coeficiente	P>[z]
433 Índice nacional de preços ao consumidor-amplo (IPCA)	0.0771	0.000*
4380 PIB mensal real	5,18E-07	0.848
11753 Índice da taxa de câmbio real (IPCA)	0.0061	0.648
4390 Taxa de juros reais - Selic	-0.0010	0.953
29028 Renda Nacional Disponível Bruta das Famílias restrita	-2.59E-03	0.695
29034 Comprometimento de renda das famílias com o serviço da dívida com o Sistema Financeiro Nacional	0.0157	0.000*
29037 Endividamento das famílias com o Sistema Financeiro Nacional	-0.0486	0.000*
20704 Concessões de crédito com recursos direcionados	-1,70E-02	0.599
21151 Inadimplência da carteira de crédito com recursos direcionados	0.1238	0.000*
25499 Taxa média mensal de juros das operações de crédito com recursos direcionados	0.5529	0.000*
21340 Índice de Valores de Garantia de Imóveis Residenciais Financiados (IVG-R)	0.0187	0.409
20837 <i>Spread</i> médio das operações de crédito com recursos direcionados	0.1568	0.000*
22803 Remuneração de empregados	-0.0001	0.976
24377 Empregados no setor privado	1.29E-02	0.975
28763 Estoque de empregos formais	-6.50E-05	0.763
24379 Pessoas ocupadas	9.47E-03	0.966
24380 Pessoas desocupadas	-2.72E-02	0.855
24399 Rendimento médio real habitual das pessoas ocupadas -Setor privado e público	0.0032	0.464
21803 RWA para risco de crédito	-6.29e-07	0.528
7495 Sistema Nacional de Pesquisa de Custos e Índices (SINAPI)	-0.1364	0.000*
ar.L1 Autocorrelação da série no lag 1	0.3555	0.000*
ma.L1 Correlação serial usando relação de erro no lag 1	-0.6730	0.000*
sigma2 Termo de erro	0.0333	0.000*

Fonte: Elaborada pela autora (2022)

Não há padrões óbvios para os resíduos (standardized residuals). O histograma com a estimativa de densidade mostra a distribuição dos resíduos, enquanto a curva KDE, a versão suavizada do histograma, a linha $N(0,1)$ mostra uma distribuição normal. Sendo que as curvas das linhas KDE e a linha da distribuição normal são semelhantes. O correlograma apresenta que 95% das correlações para defasagem maior que um não é significativa. E a normal Q-Q apresenta a maioria dos pontos dos dados na linha reta, indicando uma distribuição dos resíduos próxima da normal.

Quadro 6 – SARIMAX (0,2,2) (0,0,0) [12]

Variável	coeficiente	P>[z]
433 Índice nacional de preços ao consumidor-amplio (IPCA)	0.0807	0.000*
4380 PIB mensal real	2.57E-08	0.993
11753 Índice da taxa de câmbio real (IPCA)	0.0098	0.509
4390 Taxa de juros reais - Selic	-0.0011	0.790
29028 Renda Nacional Disponível Bruta das Famílias restrita	-3.71E-02	0.000*
29034 Comprometimento de renda das famílias com o serviço da dívida com o Sistema Financeiro Nacional	-0.3564	0.000*
29037 Endividamento das famílias com o Sistema Financeiro Nacional	-0.6865	0.000*
20704 Concessões de crédito com recursos direcionados	-7,75E-03	0.855
21151 Inadimplência da carteira de crédito com recursos direcionados	0.2519	0.000*
25499 Taxa média mensal de juros das operações de crédito com recursos direcionados	1.5116	0.000*
21340 Índice de Valores de Garantia de Imóveis Residenciais Financiados (IVG-R)	0.0192	0.000*
20837 <i>Spread</i> médio das operações de crédito com recursos direcionados	0.0806	0.000*
22803 Remuneração de empregados	-0.0017	0.694
24377 Empregados no setor privado	0.0002	0.799
28763 Estoque de empregos formais	1.36E-04	0.633
24379 Pessoas ocupadas	5.57e-05	0.888
24380 Pessoas desocupadas	-0.0002	0.233
24399 Rendimento médio real habitual das pessoas ocupadas -Setor privado e público	0.0049	0.439
21803 RWA para risco de crédito	-1.08E-03	0.224
7495 Sistema Nacional de Pesquisa de Custos e Índices (SINAPI)	-0.1319	0.000*
ma.L1 Correlação serial usando relação de erro no lag 1	-0.5261	0.000*
ma.L2 Correlação serial usando relação de erro no lag 2	-0.3203	0.000*
sigma2 Termo de erro	0.0601	0.000*

Fonte: Elaborada pela autora (2022)

Quadro 7 – Revisão de suposições SARIMAX (0,2,2) (0,0,0) [12]

Ljung-Box (L1) (Q): 0.66	Jarque-Bera (JB): 84.36
Prob(Q): 0.42	Prob(JB): 0.00
Heteroskedasticity (H): 2.44	Skew: 0.42
Prob(H) (two-sided): 0.04	Kurtosis: 8.48

Fonte: Elaborada pela autora (2022)

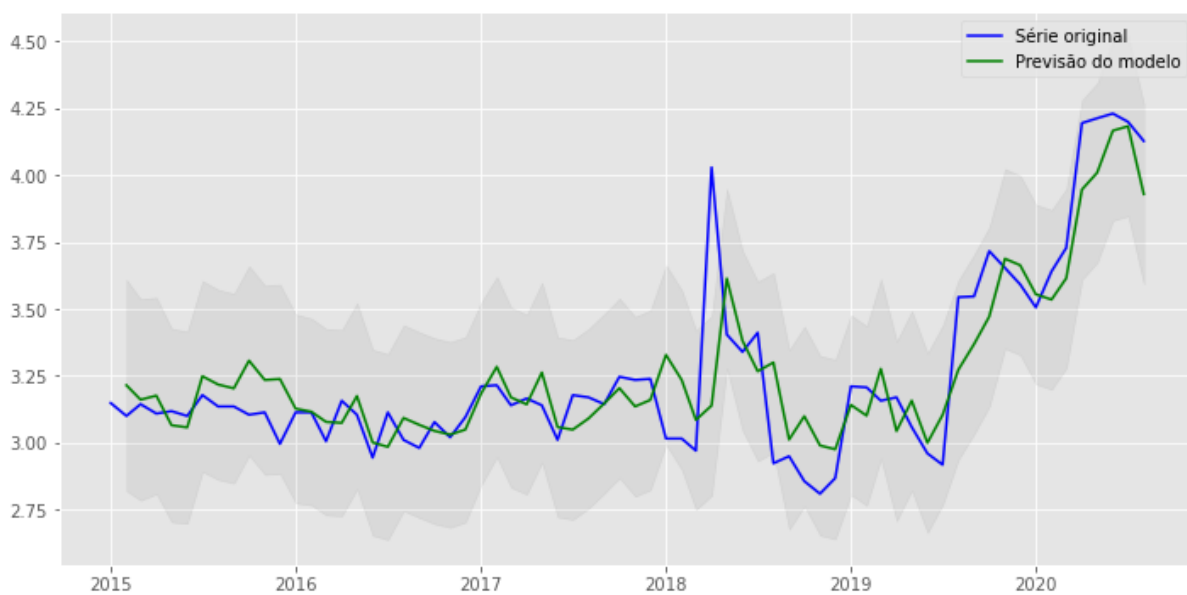
Figura 20 – Análise de resíduos SARIMAX (0,2,2) (0,0,0) [12]



Fonte: Elaborada pela autora (2022)

A análise gráfica da previsão na amostra apresenta bom ajuste à série original, conforme Figura 21.

Figura 21 – Previsão na amostra



Fonte: Elaborada pela autora (2022)

Sendo assim, escolhe-se o modelo SARIMAX (0,2,2) (0,0,0) [12] para análise da

significância estatística da variáveis macroeconômicas.

Define-se as variáveis explicativas estatisticamente significantes da seguinte forma:

- Índice nacional de preços ao consumidor-Amplio (IPCA) - aponta a variação do custo de vida médio de famílias com renda mensal de 1 e 40 salários mínimos. Os principais indicadores coletados são: IPC-Fipe, IPCA, IPCA-15, INPC, ICV e IPCA-E (BCB, 2021).
- A Renda Nacional Disponível Bruta das Famílias - RNDBF - soma os rendimentos obtidos pelo uso de fatores de produção (trabalho e capital) e as transferências recebidas descontadas das transferências pagas (BCB, 2021).
- Comprometimento de renda - Relação entre o valor correspondente aos pagamentos esperados para o serviço da dívida com o Sistema Financeiro Nacional e a renda mensal das famílias, em média móvel trimestral, ajustado sazonalmente (BCB, 2021).
- Endividamento - Relação entre o valor atual das dívidas das famílias com o Sistema Financeiro Nacional e a renda das famílias acumulada nos últimos doze meses (BCB, 2021).
- Inadimplência da carteira de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário total - Percentual da carteira de crédito regulamentado pelo Conselho Monetário Nacional (CMN) ou vinculados a recursos orçamentários com pelo menos uma parcela com atraso superior a 90 dias (BCB, 2021).
- Taxa média mensal de juros das operações de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário total - Taxa média de juros das novas operações de crédito regulamentado pelo Conselho Monetário Nacional (CMN) ou vinculado a recursos orçamentários, contratadas no período de referência. Taxa ponderada pelo valor das concessões (BCB, 2021).
- O Índice de Valor de Garantias Reais (IVG-R) estima a tendência de preço de longo prazo dos valores de imóveis residenciais no Brasil utilizando informações do Sistema de Informações de Crédito (SCR) do Banco Central do Brasil. Para isso, utiliza os valores de avaliação dos imóveis dados em garantia a financiamentos imobiliários residenciais para pessoas físicas nas modalidades de alienação fiduciária e hipoteca residencial (BCB,2021).
- *Spread* médio das operações de crédito com recursos direcionados - Pessoas físicas - Total - Diferença entre a taxa média de juros das novas operações de crédito regulamentado pelo Conselho Monetário Nacional (CMN) ou vinculado a recursos orçamentários, contratadas no período de referência e o custo de captação referencial médio (BCB, 2021).

- Sistema Nacional de Pesquisa de Custos e Índices (SINAPI) - Séries mensais de custos e índices para o setor habitacional, de séries mensais de salários medianos de mão de obra e preços medianos de materiais, máquinas e equipamentos e serviços da construção para os setores de saneamento básico, infraestrutura e habitação (BCB, 2021).

No que se refere às variáveis estatisticamente significantes identificadas no modelo SARI-MAX, as variáveis explicativas com sinal negativo mostraram-se estatisticamente significativas para explicar que menor a probabilidade de *default* quando maior a renda nacional disponível bruta das famílias, o comprometimento de renda das famílias com o serviço da dívida, o endividamento das famílias com o Sistema Financeiro Nacional e os valores dos insumos e serviços necessários às obras e serviços de engenharia, o que inclui salários (SINAPI).

A redução da probabilidade de *default* encontra respaldo no comportamento pró-cíclico das Instituições Financeiras. Em Souza e Feijó (2007) os Bancos são mais seletivos nas operações de crédito em períodos de incertezas econômicas, enquanto o contrário é verdadeiro. Ou melhor, os Bancos, alteram a postura quanto à seleção e oferta das operações de crédito como prevenção de mudanças no nível de inadimplência diante do grau de incerteza do cenário econômico. Sendo assim, o aumento do saldo das operações de crédito refletem momentos de crescimento econômico e redução de incertezas macroeconômicas e portanto, diminuição da probabilidade de *default*. Momentos de crescimento econômico aliam-se à baixas taxas de juros, sendo esse um estímulo ao crédito, aumentando os prazos médios e reduzindo a inadimplência. O que pode ser uma evidência de que menor probabilidade de *default* quando maior o comprometimento de renda e o endividamento das famílias com o Sistema Financeiro Nacional.

Pela análise gráfica da variável saldo de concessões de crédito com recursos direcionados (Figura 6) pode-se observar até meados de 2019 o comportamento de redução da probabilidade de *default* quando há crescimento do saldo de concessões. Entendemos como sendo o período posterior ao segundo semestre de 2019 como responsável pelo não entendimento pelo modelo SARIMAX da significância estatística da variável, apesar de capturar esse resultado cíclico nas variáveis comprometimento de renda e o endividamento das famílias com o Sistema Financeiro Nacional. Também não foi possível avaliar se haveria um limite máximo ao endividamento que invertesse a resposta da probabilidade de *default*, como observa-se a partir de 2019. Uma análise com uma base com um número maior de observações poderia melhor capturar esse comportamento e com a metodologia adequada explicar a relação entre o saldo de concessões, endividamento das famílias e o ciclo econômico.

O aumento do SINAPI seria uma resposta ao crescimento econômico e portanto à demanda por insumos e serviços necessários às obras e serviços de engenharia.

As variáveis explicativas com sinal positivo mostraram-se estatisticamente significativas para explicar que maior a probabilidade de *default*, quando maior o índice nacional de preços ao consumidor-amplo (IPCA), a inadimplência da carteira de crédito com recursos direcionados, a taxa média mensal de juros das operações de crédito com recursos direcionados, o *spread* médio das operações de crédito com recursos direcionados, bem como o índice de valor de garantias reais (IVG-R).

O aumento do nível esperado dos preços, como ocorre com o SINAPI em momentos de crescimento econômico, leva a um aumento no nível de preço efetivo em um segundo momento, portanto um aumento da inflação. Blanchard (2011) ao tratar da taxa natural de desemprego e a curva de Phillips explica com a espiral de preços e salários, que um salário nominal mais elevado leva a um nível de preços mais alto, uma inflação maior.

A inflação maior, reduz o estoque real da dívida pública em moeda nacional, deixando mais recursos para cobrir as obrigações da dívida externa. Por outro lado, é sintomático de problemas no nível macroeconômico (AFONSO; GOMES; ROTHER, 2010).

Como resposta os agentes da política monetária nacional, para atingir a meta da inflação, aumentam a taxa de juros, reduzindo o investimento, o que aumenta o desemprego e reduz a inflação. Segundo a regra da taxa de juros ou regra de Taylor, recorrentemente utilizada pelo Comitê de Política Monetária (COPOM) (NASSIF; FEIJO; ARAUJO, 2020), se a inflação for maior que a meta estabelecida pelo Conselho Monetário Nacional (CMN) aumenta-se a taxa nominal de juros. Essa taxa de juros maior aumentará o desemprego o que levará a uma diminuição da inflação (BLANCHARD, 2011). O aumento da taxa de juros corresponde a uma dívida pública mais elevada e deve corresponder a um risco maior de inadimplência (AFONSO; GOMES; ROTHER, 2010).

O *spread* médio das operações de crédito com recursos direcionados refletem a expectativa do resultado da atividade econômica, se diminui, os riscos de crédito e, portanto, os *spreads* de crédito parecem estar positivamente correlacionados com a probabilidade de *default*. A resposta do modelo está de acordo com a conclusão de Koopman e Lucas (2005c) que ao refazerem algumas experiências empíricas sobre a relação do risco de crédito com a macroeconomia, entenderam que haveria forte co-ciclicidade entre *spreads* e *defaults*.

No que se refere a taxa média de juros das operações de crédito com recursos direcionados é intuitivo que o aumento do risco de crédito, gere uma reação de alta de taxa de juros do financiamento seguindo no mesmo caminho da probabilidade de *default*. Apesar de ser esse o resultado do modelo, a análise gráfica das duas séries parecem dizer o contrário. Uma explicação estaria no fato de que a mudança na taxa de juros de financiamento imobiliário refletem a expectativa de mercado no longo prazo, considerando tratar-se de operações com prazo de até 35 anos, enquanto a série de probabilidade de *default* utilizada refere-se ao *default* observado no período analisado.

Por fim, a resposta positiva do índice de valor de garantias reais (IVG-R) ao aumento da probabilidade de *default* pode indicar que os mutuários têm maior incentivo para deturpar a renda em mercados de alto preço, levando ao aumento de fraudes e por consequência das taxas de inadimplência (BASIRI; MAHMOUDI, 2021). Nesse sentido, Blackburn e Vermilyea (2012), avaliaram que a tendência a distorcer a renda estaria associada a mercados com grandes aumentos de preços de imóveis durante o *boom*, a sobre declaração de renda seria mais comum em mercados com maior crescimento dos preços das casas. Na outra ponta as expectativas de crescimento dos preços da habitação mais altas diminuem a perda estimada em caso de inadimplência, permitindo assim que os credores tenham como alvo clientes mais arriscados (MIAN; SUFI, 2009).

Outro ponto de análise estaria no fato de que o crédito mais barato alimentaria temporariamente os preços das casas mais altos, vendas mais rápidas e menos execuções de hipotecas, mas o aumento dos empréstimos facilita o excesso de dívidas e o calote no futuro. Assim concluiu Hedlund (2016) ao avaliar os efeitos macroeconômicos da iliquidez da habitação usando modelo de busca direcionada de habitação com dívida de longo prazo e inadimplência.

5 Conclusões

As instituições financeiras tem por prática avaliar a probabilidade de *default* a partir de critérios de escolha de perfil de riscos dos tomadores de crédito. Segundo Souza e Feijó (2007) o controle parcial sobre o perfil da carteira de crédito possibilita às instituições bancárias manterem o nível de risco dentro de um intervalo preestabelecido por elas como um forma de contrabalancear os efeitos provindos do ambiente macroeconômico. No entanto, modelos que observam variáveis macroeconômicas têm demonstrado produzir melhores preditores para o risco de crédito, além disso, a Resolução CMN 4.557/2007 define que para prever o risco de crédito deve-se fazer uso da estimação das perdas esperadas associadas ao risco de crédito, considerando o ambiente macroeconômico corrente e alterações previstas no curto prazo (BRASIL, 2017).

Considerando a constituição do cálculo da perda esperada, tornou-se necessária a escolha da variável da probabilidade de *default*, componente do produto de cálculo da perda esperada, por apresentar melhor sensibilidade as mudanças na economia. Entendemos que uma vez apresentadas variáveis macroeconômicas estatisticamente significantes para explicar a probabilidade de *default* contribui-se para que novos estudos encontrem o melhor modelo de previsão para a probabilidade de *default* considerando o ambiente macroeconômico corrente e alterações previstas no curto prazo.

No que se refere às variáveis estatisticamente significantes identificadas no modelo SARIMAX, as variáveis explicativas renda nacional disponível bruta das famílias, comprometimento de renda das famílias com o serviço da dívida, o endividamento das famílias com o Sistema Financeiro Nacional e os valores dos insumos e serviços necessários às obras e serviços de engenharia (SINAPI) se mostraram significantes para explicar que quando maior seu crescimento menor a probabilidade de *default*. A relação pode ser entendida pela resposta da probabilidade de *default* à postura dos Bancos quanto à oferta de crédito diante do grau de incerteza do cenário econômico.

As variáveis explicativas índice nacional de preços ao consumidor-amplo (IPCA), inadimplência da carteira de crédito com recursos direcionados, taxa média mensal de juros das operações de crédito com recursos direcionados, *spread* médio das operações de crédito com recursos direcionados, e o índice de valor de garantias reais (IVG-R) se mostraram estatisticamente significantes para explicar que quando maior seu crescimento maior a probabilidade de *default*. A relação pode ser justificada pela resposta da probabilidade de *default* aos períodos de crescimento, de choques e de recessão econômica.

A contribuição do trabalho para a literatura do risco de crédito está na descrição e mensuração das variáveis macroeconômicas dentro do modelo estatístico SARIMAX, com

uso da linguagem de programação PYTHON. As conclusões podem ser pontos de partida para o desenvolvimento de modelos de previsão da probabilidade de *default* baseados nas variáveis macroeconômicas identificadas pelo modelo.

A revisão de literatura foi abrangente, considerando a apresentação de estudos na área de risco de crédito, taxa de inadimplência, inadimplência e variáveis econômicas, acordo de Basileia, efeitos dos ciclos econômicos sobre o risco de crédito. A respeito das limitações do trabalho, percebe-se que a pouca quantidade de observações da variável dependente probabilidade de *default*, no total de 68, limitou a escolha ao modelo SARIMAX. Em trabalhos futuros sugere-se a utilização de outra variável, observada por maior período que componha a probabilidade de *default*, como por exemplo as estimativas de inadimplência, de forma que possa fazer uso de algoritmos de redes neurais, dentre outros.

Referências

- AFONSO, A.; GOMES, P.; ROTHER, P. Short- and long-run determinants of sovereign debt credit ratings. *International Journal of Finance and Economics*, v. 16, p. 1–15, 2010. Citado na página 55.
- ALLEN, L.; SAUNDERS, A. A survey of cyclical effects in credit risk measurement models. *Basel Bank for International Settlements*, n. 126, 2003b. Citado na página 25.
- AMATO, J. D.; H, F. C. Are credit ratings procyclical? *Basel Bank for International Settlements*, n. 129, 2003a. Citado na página 25.
- APOSTOLIK, R.; CHRISTOPHER, D.; PETER, W. *Foundations of banking risk: An overview of banking, banking risks, and risk-based banking regulation*. John Wiley & Sons, p. 79, 2009. Citado na página 23.
- BASIRI, K.; MAHMOUDI, B. Possible income misstatement on mortgage loan applications: Evidence from the canadian housing market. *Real estate economics*, Wiley Subscription Services, Inc, v. 49, n. 3, p. 917–935, 2021. Citado na página 56.
- BCB - BANCO CENTRAL DO BRASIL. *Relatório de Estabilidade Financeira*. [S.l.], 2021. v. 20, n. 1, 6 p. Disponível em: <<https://www.bcb.gov.br/content/publicacoes/ref/202110/RELESTAB202110-refPub.pdf>>. Citado na página 23.
- BCBS - BASEL BANK FOR INTERNATIONAL SETTLEMENTS. *An Explanatory Note on the Basel 2 IRB Risk Weight Functions*. [S.l.], 2005. Disponível em: <<https://www.bis.org/bcbs/irbriskweight.htm>>. Citado 2 vezes nas páginas 24 e 27.
- BCBS - BASEL BANK FOR INTERNATIONAL SETTLEMENTS. *Guidance on Paragraph 468 of the Framework Document*. [S.l.], 2005. Disponível em: <<https://www.bis.org/publ/bcbs115.pdf>>. Citado na página 24.
- BENNETT, C.; STEWART, R.; LU, J. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies*, v. 7, n. 5, p. 2938–2960, 2014. Citado na página 41.
- BLACKBURN, M. L.; VERMILYEA, T. The prevalence and impact of misstated incomes on mortgage loan applications. *Journal of housing economics*, Elsevier Inc, v. 21, n. 2, p. 151–168, 2012. Citado na página 56.
- BLANCHARD, O. *Macroeconomia*. [S.l.]: Pearson Prentice Hall, 2011. 23,150,512 p. Citado na página 55.
- BORIO, C.; H, F. C.; LOWE, P. Procyclicality of the financial system and financial stability: issues and policy options. *Basel Bank for International Settlements*, p. 57, 2001a. Citado na página 25.
- BOX, G. E. P. et al. *Time Series Analysis: Forecasting and Control*. [S.l.]: John Wiley & Sons, 2008. v. 4. Citado 2 vezes nas páginas 39 e 42.

- BOX, G. E. P.; TIAO, G. C. Intervantion analysis with applications to economic and environmental problems. *Journal of American Statistical Association*, v. 70, p. 70–79, 1975. Citado 3 vezes nas páginas 23, 39 e 41.
- BRASIL. Pronunciamento técnico cpc 48. 2016. Disponível em: <<http://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=106>>. Citado na página 23.
- BRASIL. Resolução do conselho monetário nacional nº 4.557. *Diário Oficial da União*, p. 41–46, 2017. Disponível em: <https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50344/Res_4557_v1_O.pdf>. Citado 3 vezes nas páginas 26, 28 e 57.
- CARLING, K. et al. Corporate credit risk modeling and the macroeconomy. *Journal of Banking and Finance*, n. 31, p. 845–868, 2007. Citado na página 24.
- CHATFIELD, C. *The Analysis of Time Series: An Introduction*. [S.l.]: Chapman & HALL/CRC, 2004. 59-84 p. Citado na página 41.
- CHU, V. Juros e *Spread* bancário no brasil avaliação de 2 anos do projeto. Banco Central do Brasil, p. 41–45, 2001. Citado na página 26.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association*, v. 74, p. 427–431, 1979. Citado na página 43.
- DIEESE - DEPARTAMENTO INTERSINDICAL DE ESTATÍSTICAS E ESTUDOS SOCIOECONÔMICOS. *Nota Técnica: Medidas emergenciais adotadas pelo Sistema Financeiro Nacional devido à pandemia do coronavírus - Covid 19*. [S.l.], 2020. 11-12 p. Disponível em: <<https://www.dieese.org.br/notatecnica/2020/notaTec234MedidasSistemaFinanceiro.pdf>>. Citado na página 35.
- GOURINCHAS, P.-O.; VALDÉS, R.; LANDERRECHE, O. Lending booms: Latin america and the world. NBER Working Paper Series, p. 61, 2001b. Citado na página 25.
- HEDLUND, A. Illiquidity and its discontents: Trading delays and foreclosures in the housing market. *Journal of monetary economics*, Elsevier B.V, v. 83, p. 1–13, 2016. Citado na página 56.
- IMF - INTERNATIONAL MONETARY FUND. *Regional Economic Outlook. Western Hemisphere : pandemic persistence clouds the recovery*. [S.l.], 2020. 1-42 p. Disponível em: <<https://www.elibrary.imf.org/view/books/086/29365-9781513558370-pt/29365-9781513558370-pt-book.xml>>. Citado na página 23.
- JARDIM, E.; LORRANY, L. *Evolução Recente do Mercado de Crédito Imobiliário no Brasil*. [S.l.], 2020. 12 p. Disponível em: <https://www.ipea.gov.br/portal/images/stories/PDFs/radar/200818_radar_63_artigo_02.pdf>. Citado na página 23.
- KOOPMAN, S. J.; LUCAS, A. Business and default cycles for credit risk. *Journal of Applied Econometrics*, v. 20, n. 2, p. 311–323, 2005c. Citado 2 vezes nas páginas 25 e 55.
- LESSMANN, S. et al. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, n. 247, p. 124–136, 2015. Citado na página 27.

- MAÇAIRA, P. M. et al. Time series analysis with explanatory variables: A systematic literature review. *Environmental Modelling and Software*, v. 107, p. 199–209, 2018. ISSN 1364-8152. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S136481521730542X>>. Citado na página 42.
- MIAN, A.; SUFI, A. The consequences of mortgage credit expansion: Evidence from the u.s. mortgage default crisis. *The Quarterly journal of economics*, MIT Press, v. 124, n. 4, p. 1449–1496, 2009. Citado na página 56.
- MÁXIMO, W. *Caixa encerra pagamento do auxílio emergencial após sete meses*. 2021. Agência Brasil - Brasília. Disponível em: <<https://agenciabrasil.ebc.com.br/economia/noticia/2021-10/caixa-encerra-pagamento-do-auxilio-emergencial-apos-sete-meses>>. Citado na página 35.
- NASSIF, A.; FEIJO, C.; ARAUJO, E. Macroeconomic policies in brazil before and after the 2008 global financial crisis: Brazilian policy-makers still trapped in the new macroeconomic consensus guidelines. *Cambridge journal of economics*, Oxford University Press, v. 44, n. 4, p. 749–779, 2020. Citado 3 vezes nas páginas 30, 33 e 55.
- PAL, A.; PRAKASH, P. K. S. *Practical Time Series Analysis: Master Time Series Data Processing, Visualization, and Modeling using Python*. [S.l.]: Packt Publishing, 2017. v. 1. 113-159 p. Citado 2 vezes nas páginas 39 e 40.
- SCHECHTMAN, R. *Uma Investigação Baseada em Reamostragem sobre Requerimentos de Capital para Risco de Crédito no Brasil*. [S.l.], 2006. Disponível em: <<https://EconPapers.repec.org/RePEc:bcb:wpaper:127>>. Citado na página 26.
- SCHECHTMAN, R. et al. *Credit Risk Measurement and the Regulation of Bank Capital and Provision Requirements in Brazil - A Corporate Analysis*. [S.l.], 2004. Disponível em: <<https://ideas.repec.org/p/bcb/wpaper/91.html>>. Citado na página 26.
- SOUZA, G. J. de G.; FEIJÓ, C. A. V. C. O processo interativo entre a dinâmica macroeconômica e os bancos: uma perspectiva acerca do risco de crédito além do spread. In: ANAIS ELETRÔNICOS SÃO PAULO: ANPEC. 2007. p. 14–15. Disponível em: <<http://www.anpec.org.br/encontro2007/artigos/A07A099.pdf>>. Citado 2 vezes nas páginas 54 e 57.
- TABAK, B. M. et al. *The Stability-Concentration Relationship in the Brazilian Banking System*. [S.l.], 2007. Disponível em: <<https://ideas.repec.org/p/bcb/wpaper/145.html>>. Citado na página 24.
- TINOCO, G.; GIAMBIAGI, F. *Perspectivas DEPEC 2018: O Crescimento da Economia Brasileira 2018-2023*. [S.l.], 2018. Disponível em: <https://web.bndes.gov.br/bib/jspui/bitstream/1408/14760/1/Perspectivas202018-2023_P.pdf>. Citado na página 32.
- VISUALIZAÇÃO dos dados básicos e metadados das séries BCB. 2021. BCB - Banco Central do Brasil. Disponível em: <<https://www3.bcb.gov.br/sgspub>>. Citado 2 vezes nas páginas 53 e 54.

Apêndices

APÊNDICE A – Variáveis econômicas

Quadro 8 – Variáveis Econômicas

Index	Variável
433	Índice nacional de preços ao consumidor-amplo (IPCA)
4380	PIB mensal real - Valores correntes deflacionado pelo IGPM (R\$ milhões)
11753	Índice da taxa de câmbio real (IPCA)
4390	Taxa de juros reais - Selic acumulada no mês deflacionada pelo IPCA
29028	Renda Nacional Disponível Bruta das Famílias restrita - em valores constantes, deflacionada pelo IPCA, com ajuste sazonal (média móvel trimestral)
29034	Comprometimento de renda das famílias com o serviço da dívida com o Sistema Financeiro Nacional - Com ajuste sazonal (RNDBF)
29037	Endividamento das famílias com o Sistema Financeiro Nacional em relação à renda acumulada dos últimos doze meses (RNDBF)
20704	Concessões de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário total
21151	Inadimplência da carteira de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário total
25499	Taxa média mensal de juros das operações de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário total
21340	Índice de Valores de Garantia de Imóveis Residenciais Financiados (IVG-R)
20837	<i>Spread</i> médio das operações de crédito com recursos direcionados - Pessoas físicas - Total
22803	Remuneração de empregados - mensal - líquido
24377	Empregados no setor privado - Total - PNADC
28763	Estoque de empregos formais - Total
24379	Pessoas ocupadas - PNADC
24380	Pessoas desocupadas - PNADC
24399	Rendimento médio real habitual das pessoas ocupadas - Setor privado e público - PNADC
21803	RWA para risco de crédito por abordagem padronizada
7495	Sistema Nacional de Pesquisa de Custos e Índices (SI-NAPI)

Fonte: Elaborada pela autora (2022)

APÊNDICE B – Mínimo AIC

Quadro 9 – Pesquisa mínimo AIC Modelo SARIMAX(0,1,3)(0,0,0)[12]

Modelo	AIC	Tempo
ARIMA(1,1,1)(1,0,1)[12] intercept	AIC=10.645	Time=0.45 sec
ARIMA(0,1,0)(0,0,0)[12] intercept	AIC=10.131	Time=0.19 sec
ARIMA(1,1,0)(1,0,0)[12] intercept	AIC=9.189	Time=0.34 sec
ARIMA(0,1,1)(0,0,1)[12] intercept	AIC=8.662	Time=0.35 sec
ARIMA(0,1,0)(0,0,0)[12]	AIC=223.857	Time=0.18 sec
ARIMA(0,1,1)(0,0,0)[12] intercept	AIC=5.633	Time=0.24 sec
ARIMA(0,1,1)(1,0,0)[12] intercept	AIC=7.171	Time=0.34 sec
ARIMA(0,1,1)(1,0,1)[12] intercept	AIC=11.919	Time=0.40 sec
ARIMA(1,1,1)(0,0,0)[12] intercept	AIC=3.892	Time=0.23 sec
ARIMA(1,1,1)(1,0,0)[12] intercept	AIC=5.499	Time=0.36 sec
ARIMA(1,1,1)(0,0,1)[12] intercept	AIC=7.263	Time=0.37 sec
ARIMA(1,1,0)(0,0,0)[12] intercept	AIC=7.576	Time=0.23 sec
ARIMA(2,1,1)(0,0,0)[12] intercept	AIC=5.769	Time=0.24 sec
ARIMA(1,1,2)(0,0,0)[12] intercept	AIC=3.362	Time=0.24 sec
ARIMA(1,1,2)(1,0,0)[12] intercept	AIC=4.903	Time=0.37 sec
ARIMA(1,1,2)(0,0,1)[12] intercept	AIC=6.511	Time=0.36 sec
ARIMA(1,1,2)(1,0,1)[12] intercept	AIC=10.135	Time=0.40 sec
ARIMA(0,1,2)(0,0,0)[12] intercept	AIC=2.887	Time=0.23 sec
ARIMA(0,1,2)(1,0,0)[12] intercept	AIC=4.379	Time=0.35 sec
ARIMA(0,1,2)(0,0,1)[12] intercept	AIC=5.575	Time=0.37 sec
ARIMA(0,1,2)(1,0,1)[12] intercept	AIC=8.848	Time=0.45 sec
ARIMA(0,1,3)(0,0,0)[12] intercept	AIC=4.320	Time=0.25 sec
ARIMA(1,1,3)(0,0,0)[12] intercept	AIC=13.737	Time=0.29 sec
ARIMA(0,1,2)(0,0,0)[12]	AIC=1.013	Time=0.25 sec
ARIMA(0,1,2)(1,0,0)[12]	AIC=2.516	Time=0.36 sec
ARIMA(0,1,2)(0,0,1)[12]	AIC=3.700	Time=0.46 sec
ARIMA(0,1,2)(1,0,1)[12]	AIC=6.903	Time=0.38 sec
ARIMA(0,1,1)(0,0,0)[12]	AIC=3.671	Time=0.19 sec
ARIMA(1,1,2)(0,0,0)[12]	AIC=1.249	Time=0.22 sec
ARIMA(0,1,3)(0,0,0)[12]	AIC=0.315	Time=0.24 sec
ARIMA(0,1,3)(1,0,0)[12]	AIC=1.832	Time=0.35 sec
ARIMA(0,1,3)(0,0,1)[12]	AIC=2.882	Time=0.40 sec
ARIMA(0,1,3)(1,0,1)[12]	AIC=6.001	Time=0.59 sec
ARIMA(1,1,3)(0,0,0)[12]	AIC=1.501	Time=0.29 sec
ARIMA(0,1,4)(0,0,0)[12]	AIC=1.731	Time=0.29 sec
ARIMA(1,1,4)(0,0,0)[12]	AIC=3.813	Time=0.43 sec

Fonte: Elaborada pela autora (2022)

Quadro 10 – Pesquisa mínimo AIC Modelo SARIMAX(0,2,2)(0,0,0)[12]

Modelo	AIC	Tempo
ARIMA(1,2,1)(0,0,0)[12]	AIC=44.339,	Time=0.25 sec
ARIMA(0,2,0)(0,0,0)[12]	AIC=270.960,	Time=0.18 sec
ARIMA(1,2,0)(1,0,0)[12]	AIC=57.066,	Time=0.36 sec
ARIMA(0,2,1)(0,0,1)[12]	AIC=51.216,	Time=0.35 sec
ARIMA(1,2,1)(1,0,0)[12]	AIC=45.982,	Time=0.39 sec
ARIMA(1,2,1)(0,0,1)[12]	AIC=46.308,	Time=0.38 sec
ARIMA(1,2,1)(1,0,1)[12]	AIC=48.731,	Time=0.40 sec
ARIMA(0,2,1)(0,0,0)[12]	AIC=49.256,	Time=0.23 sec
ARIMA(1,2,0)(0,0,0)[12]	AIC=55.470,	Time=0.22 sec
ARIMA(2,2,1)(0,0,0)[12]	AIC=47.872,	Time=0.22 sec
ARIMA(1,2,2)(0,0,0)[12]	AIC=42.612,	Time=0.28 sec
ARIMA(1,2,2)(1,0,0)[12]	AIC=44.246,	Time=0.50 sec
ARIMA(1,2,2)(0,0,1)[12]	AIC=44.580,	Time=0.48 sec
ARIMA(1,2,2)(1,0,1)[12]	AIC=47.022,	Time=0.50 sec
ARIMA(0,2,2)(0,0,0)[12]	AIC=39.938,	Time=0.25 sec
ARIMA(0,2,2)(1,0,0)[12]	AIC=41.622,	Time=0.37 sec
ARIMA(0,2,2)(0,0,1)[12]	AIC=41.909,	Time=0.42 sec
ARIMA(0,2,2)(1,0,1)[12]	AIC=44.323,	Time=0.53 sec
ARIMA(0,2,2)(0,0,0)[12] intercept	AIC=42.841,	Time=0.32 sec

Fonte: Elaborada pela autora (2022)

APÊNDICE C – Código Python

```
#!/usr/bin/env python
# coding: utf-8

# In[1]:

# Import libraries
## Basic libs
import pandas as pd
import numpy as np
import warnings
## Building Model
from scipy import stats
import statsmodels
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.stats.api as sms
from statsmodels.compat import lzip
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import pmdarima as pm
## Data Visualization
import seaborn as sns
import matplotlib.pyplot as plt
get_ipython().run_line_magic('matplotlib', 'inline')
from mpl_toolkits.mplot3d import Axes3D
plt.rcParams.update({'figure.figsize':(12,6)})
warnings.filterwarnings('ignore')
plt.style.use('ggplot')

# In[2]:

# Import Data
df = pd.read_csv('Dados/DFMod4.csv', sep = ';')
df['Date'] = pd.to_datetime(df['Date'])
```

```
df.set_index('Date', inplace=True)
```

```
# In[3]:
```

```
# PD Log
```

```
df ['PDlog'] = np.log(df.PD.dropna())
```

```
# In[4]:
```

```
# Vizualize
```

```
import IPython.display
```

```
IPython.display.HTML(df.to_html())
```

```
# In[5]:
```

```
# Augmented DickeyFuller (ADCF)
```

```
from statsmodels.tsa.stattools import adfuller
```

```
def summary_adfuller(data):
```

```
    result = adfuller(data)
```

```
    print('ADF Statistic: %f' % result[0])
```

```
    print('p-value: %f' % result[1])
```

```
    print('Critical Values:')
```

```
    for key, value in result[4].items():
```

```
        print('\t%s: %.3f' % (key, value))
```

```
# In[6]:
```

```
summary_adfuller(df.PDlog.dropna())
```

```
# In[7]:
```

```
# Differentiation
DF0 = df.copy(deep=True)

# In[8]:

DF0['PDlog'] = DF0['PDlog'].diff().dropna()
summary_adfuller(DF0['PDlog'].dropna())

# In[9]:

DF0['PDlogDiff2'] = (DF0['PDlog'] - 2*DF0['PDlog'].shift(1) +
                    DF0['PDlog'].shift(2))

# In[10]:

# Differentiation Plot
first_order_diff =df['PDlog'].diff(1).dropna()
fig, ax = plt.subplots(2, sharex=True)
fig.set_size_inches(5.5, 5.5)
df['PDlog'].plot(ax=ax[0], color='b')
ax[0].set_title('PD Jan 2015-Ago 2020')
first_order_diff.plot(ax=ax[1], color='r')
ax[1].set_title('PD First differencing Jan 2015-Ago 2020')

# In[11]:

# Function to plot signal, ACF and PACF
def plotds(xt, nlag=30, fig_size=(12, 10)):
    if not isinstance(xt, pd.Series):
        xt = pd.Series(xt)
    plt.figure(figsize=fig_size)
    layout = (2, 2)

    # Assign axes
```

```
ax_xt = plt.subplot2grid(layout, (0, 0), colspan=2)
ax_acf= plt.subplot2grid(layout, (1, 0))
ax_pacf = plt.subplot2grid(layout, (1, 1))
```

```
# Plot graphs
xt.plot(ax=ax_xt)
ax_xt.set_title('Time Series')
plot_acf(xt, lags=30, ax=ax_acf)
plot_pacf(xt, lags=30, ax=ax_pacf)
plt.tight_layout()
return None
```

```
# In[12]:
```

```
# Plot ACF and PACF
PD=df['PDlog'].dropna()
plotds(df['PDlog'], nlag=33)
```

```
# In[13]:
```

```
# Plot signal
plotds(first_order_diff, nlag=50)
adf_result= adfuller(first_order_diff)
print('ADF Statistic: %f' % adf_result[0])
print('p-value: %f' % adf_result[1])
```

```
# In[14]:
```

```
# Granger Causality test
from statsmodels.tsa.stattools import grangercausalitytests
maxlag=1

def grangers_causation_matrix(data, variables, test='ssr_chi2test',
    verbose=False):
    df = pd.DataFrame(np.zeros((len(variables), len(variables))),
        columns=variables, index=variables)
```

```
for c in df.columns:
    for r in df.index:
        test_result = grangercausalitytests(data[[r, c]], maxlag=maxlag,
            verbose=False)
        p_values = [round(test_result[i+1][0][test][1],4) for i in
            range(maxlag)]
        if verbose: print(f'Y = {r}, X = {c}, P Values = {p_values}')
        min_p_value = np.min(p_values)
        df.loc[r, c] = min_p_value
df.columns = [var + '_x' for var in variables]
df.index = [var + '_y' for var in variables]
return df

def ImprimeGT(data):
    variables=data.columns
    df = grangers_causation_matrix(data, variables, test='ssr_chi2test',
        verbose=False)
    print(grangers_causation_matrix(df, variables=df.columns))
    return

ImprimeGT(df)

# In[15]:

# Correlation
correlation = df.corr().round(4)
plot = sns.heatmap(correlation, annot = True, fmt=".1f", linewidths=.6)
plot

# In[16]:

df.corr()

# In[17]:

#Seasonal Decompose
```

```
from statsmodels.tsa.seasonal import seasonal_decompose
result= seasonal_decompose(df.PDlog)
result.plot()
plt.show()
```

```
# In[18]:
```

```
# Calculate and plot autocorrelation
from pandas.plotting import autocorrelation_plot
autocorrelation_plot(df.PD)
```

```
# In[19]:
```

```
# OCSB test of seasonality
resultsOCSB = pm.arma.OCSBTest(m
    =12).estimate_seasonal_differencing_term(df['PDlog'])
print (resultsOCSB)
```

```
# In[20]:
```

```
#Canova-hansen test of seasonality
resultsCH =
    pm.arma.CHTest(m=12).estimate_seasonal_differencing_term(df['PDlog'])
print(resultsCH)
```

```
# In[21]:
```

```
#Performing stepwise search to minimize aic
stepwise=pm.auto_arma(df[['PDlog']],
    exogenous=df[['433', '4380R', '11753', '4390R', '29028',
        '29034', '29037', '20704', '21151', '25499',
        '21340', '20837', '22803',
        '24377', '28763',
```

```

        '24379',
        '24380', '24399', '21803', '7495']],

start_p=1,
start_q=1,
d=1,
max_p=6,
max_q=6,
max_d=2,
start_P=1,
start_Q=1,
D=0,
max_P=2, max_D=1, max_Q=2, max_order=5,
m=12,
seasonal=True,
trace=True,
stationary = False,
error_action='ignore', suppress_warnings=True,
stepwise=True)

# In[22]:

#SARIMAX (0,1,3)(0,0,0)[12]
model =pm.auto_arima(df[['PDlog']],
    exogenous=df[['433', '4380R', '11753', '4390R', '29028',
        '29034', '29037', '20704', '21151', '25499',
        '21340', '20837', '22803',
        '24377', '28763',
        '24379',
        '24380', '24399', '21803', '7495']],

    start_p=1, start_q=1,
    test='adf',      # use adftest to find optimal 'd'
    max_p=3, max_q=3, # maximum p and q
    m=12,           # frequency of series
    d=1,
    seasonal= True,
    start_P=0,
    D=0,
    trace=True,

```

```
error_action='trace',  
suppress_warnings=True,  
stepwise=True)
```

```
print(model.summary())
```

```
# In[23]:
```

```
model.plot_diagnostics(figsize=(15, 12))
```

```
# In[24]:
```

```
# SARIMAX (0,2,2)(0,0,0)[12]
```

```
sxmodelJ = pm.auto_arima(df[['PDlog']],  
    exogenous=df[['433', '4380R', '11753', '4390R', '29028',  
    '29034', '29037', '20704', '21151', '25499',  
    '21340', '20837', '22803',  
    '24377', '28763',  
    '24379',  
    '24380', '24399', '21803', '7495']],  
    start_p=1, d=2, start_q=1,  
    max_p=2, max_d=2, max_q=2,  
    start_P=0, D=0, start_Q=0,  
    max_P=3, max_D=3, max_Q=3,  
    max_order = 6, m=12,seasonal = True,  
    stationary = False,  
    trace=True,  
    error_action='ignore',  
    suppress_warnings=True,  
    stepwise=True)
```

```
sxmodelJ.summary()
```

```
# In[25]:
```

```
sxmodelJ.plot_diagnostics(figsize=(15, 12))
```

```
# In[26]:
```

```
#SARIMAX (1, 1, 1)(0,0,0)
from statsmodels.tsa.api import SARIMAX
y=df[['PDlog']]
sarimax_exog_res = SARIMAX(y, exog= df[['433', '4380R', '11753', '4390R', '29028',
                                         '29034', '29037', '20704', '21151', '25499',
                                         '21340', '20837', '22803', '24377', '28763',
                                         '24379', '24380', '24399', '21803', '7495']],
                           order=(1, 1, 1), trend="n", seasonal =
                               True, m=12).fit()

print(sarimax_exog_res.summary())
```

```
# In[27]:
```

```
sarimax_exog_res.plot_diagnostics(figsize=(15, 12))
```

```
# In[28]:
```

```
import abc
from abc import ABCMeta
from sklearn.base import BaseEstimator
```

```
# In[29]:
```

```
#PREDICT IN SAMPLE
fitted, conf_int =
    model.predict_in_sample(df[['433', '4380R', '11753', '4390R', '29028',
                                '29034', '29037', '20704', '20774', '21151',
                                '25499', '1636', '21340', '20837', '22803',
                                '28763', '24379', '24399', '21803', '7495']],
                            return_conf_int=True, alpha=0.05)

plt.plot(df[['PDlog']], color = 'blue')
```

```
hd_series = pd.Series(fitted[1:], index=df[['PDlog']].index[1:])
hdlower_series = pd.Series(conf_int[1:, 0], index=df[['PDlog']].index[1:])
hdupper_series = pd.Series(conf_int[1:, 1], index=df[['PDlog']].index[1:])

plt.plot(hd_series,color='green')
plt.fill_between(hdlower_series.index,
hdlower_series,
hdupper_series,
color='k', alpha=.15)
plt.show()
```
