Jackson da Silva Torres

2020

Table of Contents

1.50cm

No processo histórico, ao longo dos séculos, os instrumentos financeiros passaram por profundas modificações e evoluções, assumindo papel determinante na geração e acúmulo de riqueza para as famílias e nações. Nesse contexto surgiram e se consolidaram as instituições bancárias, atuando essencialmente na segurança de depósitos e na oferta de crédito.

Na contrapartida destas atividades, os bancos são remunerados basicamente de duas formas. A primeira delas é através de taxas sobre os serviços prestados. A segunda se dá pelo resultado da diferença entre a taxa cobrada no oferecimento de crédito e a taxa que remunera os recursos captados e utilizados para empréstimo, caracterizada como o *spread* bancário.

A medida que a oferta de crédito desponta como um fator fundamental para o crescimento econômico de longo prazo, incentivando empreendimentos produtivos, contribuindo assim com a geração de emprego, renda e lucros, o *spread* bancário passa a ser um indicador estratégico para determinação do nível de desenvolvimento dos países e regiões.

A primeira via da importância do *spread* bancário está relacionada com a solidez do sistema financeiro. O nível deste indicador deve ser suficiente para garantir lucros atrativos, fazendo que as instituições mantenham suas atividades e que novas tenham interesse em entrar no mercado, resultando em um setor forte, com segurança e liquidez.

A segunda via remete a relação entre o *spread* e o nível de atividade econômica. Segundo a teoria, um elevado nível de *spread* bancário desfavoreceria o crédito produtivo e consequentemente o nível de atividade econômica, impactando no crescimento e desenvolvimento do país ou região.

Tais premissas são sustentadas pelo Fundo Monetário Internacional (FMI) e pelo Banco Mundial (BM), que realiza e incentivam estudos sobre o indicador a nível mundial. A grande maioria dos estudos demonstram a relação inversa entre a taxa de *spread* bancário e o indicadores que indicam o desenvolvimento dos países e regiões .

Dentre os países estudados, desde a década de 1990, é evidenciado o caso brasileiro, com elevados níveis de *spread* bancário, baixa relação entre crédito e PIB e cenários de crescimento econômicos instáveis e considerados baixos. E mesmo com recentes modificações no cenário, o mercado brasileiro é um caso que desperta constantes iniciativas de estudos que buscam sua compreensão.

Diversos estudos constataram que a América Latina possui as maiores taxas de juros, bancos mais ineficientes implicando níveis de *spread* elevados. O caso brasileiro é destacado com um cenário ainda mais crítico, como níveis de lucro considerados muito elevado, afetando principalmente o setor produtivo .

Durante a década de 1990, o *spread* bancário brasileiro esteve superior a 50%a.a., enquanto na América Latina o observado ficou foi entre 10% e 15% a.a. A relação crédito/PIB no Brasil, em 2003, era de 23%, considerado muito baixo em comparação ao Chile com 68,5%, Uruguai com 64,3%, Estados Unidos com 60,8%, Japão com 64,3%, Coréia com 98,9% e Europa com 140,6%

Esta pesquisa parte da indagação: Quais variáveis exógenas e endógenas influenciam os componentes explícitos e implícitos do *spread* bancário e como a variação destes componentes afetam a rentabilidade dos bancos?. Para atingir os objetivos este estudo está dividido em dois capítulos além desta introdução.

Este estudo buscará verificar quais as variáveis microeconômicos macroeconômicos influenciam os componentes do *spread* *ex-post* e como estes afetaram a rentabilidade das instituições bancárias brasileiras entre os anos de 1999 e 2020.

1. Realizar levantamento bibliográfico acerca do setor bancário e *spread* no Brasil;
2. identificar e testar variáveis macroeconômicas e microeconômicas enquanto componentes implícitas e explícitas do *spread* bancário *ex-post* e como suas variações afetam a rentabilidade dos bancos;
3. Analisar e explanar os resultados obtidos nos capítulos 3 e 4 diante as premissas e pesquisas levantadas.

A medida em que a economia se torna cada vez mais dinâmica, com modificações em espaço de tempo mais curtos, o setor financeiro deve acompanhar essas perspectivas de dinamismo e evolução, o que vem ampliar sua importância no processo econômico, no oferecimento de novos produtos, ampliação de crédito, oportunidades para investidores, solidez e liquidez do sistema.

E nesse cenário dinâmico e evolutivo, o *spread* bancário desponta como o indicador que capta o nível de desenvolvimento do sistema financeiro, no sentido de equacionar a relação entre remuneração dos superavitários e o juros cobrados dos deficitários de capital, com maior relevância os destinados a empreendimentos produtivos.

Nesse sentido os estudos acerca do *spread* bancário se tornam necessários e e importantes, diante a perspectiva do dinamismo e constantes transformações da economia. O caso brasileiro demonstra ser ainda mais relevante, por se colocar em uma posição considerada peculiar, diante os históricos baixo crescimento e desenvolvimento e um setor bancário concentrado com elevados níveis de *spread*.

Foi verificado o panorama das publicações de pesquisas relacionadas ao setor bancário no brasil, através da plataforma Capes, entre os anos 2000 e 2020. , remontando um total de 4.512 publicações, indicando a relevância do tema.

Enquanto as pesquisas relacionadas especificamente com o *spread* bancário brasileiro, através da plataforma Capes foram identificados 3.435 estudos entre os anos de 2000 e 2020 , o que também vem destacar a importância da temática na literatura acadêmica.

Mesmo com um número considerado de estudos sobre o setor bancário brasileiro e outros diversos citando o *spread* bancário, ainda se fazem necessárias novas iniciativas, diante a importância deste indicador na solidez do setor financeiro e seu papel no desenvolvimento econômico, e principalmente por ainda existirem incongluências, divergências e lacunas a serem explanadas.

O próximo capítulo irá abordar sobre aspectos conceituais, teóricos, técnicos e históricos do setor e *spread* bancário brasileiro, buscando demonstrar as principais características deste mercado, através da descrição e demonstração da evolução da estrutura e seus principais indicadores.

A presente dissertação estará estruturada em cinco capítulos. Este primeiro, apresenta a contextualização, objetivos, justificativa e a presente estrutura do trabalho. E visa abordar as principais questões sobre o tema proposto com ótica direcionada para a proposição e objetivos, bem como nortear o andamento da pesquisa.

O segundo capítulo será realizado o levantamento teórico a respeito dos conceitos, aspectos fundamentais e dados amplos do setor bancário e *spread* no Brasil, por meio de pesquisa bibliográfica e busca em banco de dados.

O terceiro capítulo terá como anseio identificar e testar a variáveis macroeconômicas e microeconômicas que se configuram como componentes implícitos e explícitos do *spread* bancário *ex-post*, e como variação destes componentes afetam a rentabilidade dos bancos. Será realizada através da coleta, análises, tratamento de dados e construção de modelo econométrico utilizando dados em painel.

O quarto capítulo almeja explanar e analisar, diante as premissas empregadas, os resultados obtidos na pesquisa, visando descrever e embasar as relações entre as variáveis, e contextualizando com as pesquisas identificadas.

Por fim serão realizadas as considerações finais, buscando contextualizar os resultados obtidos com o cenário e os estudos identificados e sugerindo estudos complementares.

Neste capítulo serão abordados os conceitos, características, composição e evolução do setor bancário brasileiro, com objetivo de identificar variáveis quantitativas e qualitativas relevantes para as análises dos componentes e determinantes do *spread* bancário.

Essa avaliação se torna relevante na concepção que o nível de desenvolvimento do sistema financeiro guarda relação direta com grau de desenvolvimento econômico de um país ou região, diante seu nível de oferta e custo e destinação do crédito.

O setor bancário exerce papel socioeconômico fundamental, atuando na intermediação financeira, promovendo a circulação do fluxo de crédito, disponibilizando meios de pagamentos e opções para alocação de recursos MAFFILI E SOUZA (2007)

O desenvolvimento do setor bancário pode ser influenciado por diversos fatores endógenos — relacionados com a gestão, tecnologia e eficiência de cada instituição — e exógenos — envolvendo a regulação, conjuntura econômica e social .

Devido a importância de um sistema financeiro sólido no desenvolvimento econômico de longo prazo, o lucro das instituições bancárias desperta constante atenção em diversos países e regiões. Estas giram em torno dos riscos que envolvem descontinuidade e insolvência (COUTO *apud* ). De acordo com Freitas e Khöler (2009) *apud* , o Brasil apresenta uma conjuntura bancária bem específica em comparação a outros países.

O setor bancário brasileiro é componente do Sistema Financeiro Nacional (SFN), sob hierarquia normativa do Conselho Monetário Nacional (CMN) e supervisão do Banco Central do Brasil (BACEN). As instituições que formam o setor bancário assumem o papel de operadoras no mercado de crédito, atuando como intermediadoras financeiras junto às pessoas físicas e jurídicas, podendo ser de caráter público ou privado .

As modalidades de instituições no setor bancário brasileiro são os Bancos Comerciais, Bancos de Investimentos, Bancos de Desenvolvimento, Bancos de Câmbio, Bancos Múltiplos e Caixas Econômicas .

Os bancos comerciais são instituições financeiras de caráter público ou privado constituídas na forma de sociedade anônima, atuando na intermediação de recursos financeiros de curto e médio prazo para financiamento de atividades comerciais, industriais, serviços, pessoas físicas e terceiros, realizando captações através de depósitos à vista de livre movimento e depósitos à prazo .

A modalidade de Bancos de Investimento, as instituições financeiras devem ter caráter privado, podendo operar participações temporárias em sociedades, financiamentos produtivos para ativo fixo e capital de giro e gestão de recursos de terceiros. Realizam captação de recursos por meio de depósitos a prazo, repasses externos e internos e comercialização de cotas de fundos de investimentos que administram .

Na categoria de Bancos de Desenvolvimento, são autorizadas instituições financeiras de caráter público, controladas por governos estaduais, com foco em financiamento de atividades que promovam o desenvolvimento econômico regional no médio e longo prazo, realizando operações passivas de depósitos a prazo, recursos externos, endossos hipotecários e operações ativas de empréstimos e financiamentos ao setor privado .

Os Bancos Múltiplos se caracterizam por instituições financeiras que, podem assumir caráter público ou privado e, são autorizadas a realizar operações ativas e passivas por meio de acumulação das carteiras comercial, investimento, desenvolvimento, crédito imobiliário, arrendamento mercantil e crédito, financiamento e investimento .

Em sua composição os Bancos Múltiplos devem assumir no mínimo duas carteiras e, de forma obrigatória, uma delas, deve ser a comercial ou a de investimento. As que optarem por carteira comercial podem realizar captação via depósito à vista. Somente os Bancos Públicos podem acumular a carteira de desenvolvimento .

No segmento de Bancos de Câmbio, as instituições financeiras possuem autorização para realizar operações compra e venda de crédito cambial. Entre as operações de crédito estão o financiamento para exportadores e importadores e antecipação mediante contratos cambiais. Podem receber depósitos em contas com movimentação restrita e sem remuneração exclusiva para as operações cambiais .

A Caixa Econômica Federal (CEF), fundada em 1861, e regulamentada pelo Decreto-Lei nº 759 de 1969 é uma empresa pública subordinada ao Ministério da Economia, com operações similares a de um Banco Comercial, priorizando projetos e programas relacionados a área social e infraestrutura .

A CEF atua com operações de crédito ao consumidor, para financiamento de bens de consumo duráveis, operações de garantia de penhor industrial e caução de títulos. Detém o monopólio sobre o penhor de bens pessoais e venda de bilhetes de loteria. É integrante do Sistema Financeiro da Habitação (SFH) e Sistema Brasileiro de Poupança e Empréstimo (SBPE), além da detenção centralizado do recolhimento e aplicação dos recursos do FGTS .

O setor bancário brasileiro passou por significativas modificações em sua estrutura no final da década de 1980 e ao longo da década de 1990. Estas modificações ocorreram em grande parte como reflexo às mudanças internacionais e ao processo de abertura comercial e financeira que se iniciou no Brasil .

Na é possível verificar a concentração — levando em consideração a quantidade de instituições — do setor bancário brasileiro na categoria de bancos múltiplos, com 76%,3 de participação, onde apenas 11,5% das instituições bancárias operam exclusivamente com carteira comercial e 6,3% exclusivamente com investimento.

Entre as principais mudanças iniciadas na década de 1980 está a reforma bancária ocorrida em 1998, através da Resolução nº 1.524 , que instituiu diversas medidas de desregulamentação, entre elas a extinção da necessidade de carta-patente para constituição de Bancos Múltiplos.

Mesmo com as limitações da Constituição de 1988 para instalação de bancos estrangeiros, não houveram restrições para que ocorresse aumento na participação de capital estrangeiro em bancos nacionais .

O índice de concentração Herfindahl–Hirschmanahl(HHI), refere-se a uma medida de concentração de mercado que mede a participação de uma determinada firma no mercado do qual participa. É obtida pela somatória quadrática da parcela de mercado a ser considerada, variando entre e .

A versão normalizada do HHO traz uma variação entre e , perdendo em seu resultado a captação diante o número de firmas:

A versão decomposta do HHI avalia assimetria da concentração de mercado inserindo um componente de variabilidade das participações das firmas, assim se verifica se as firmas possuem uma participação de mercado simétrica resultando e .

A demonstra a evolução número de instituições bancárias por segmento entre 1978 à 2019, podendo ser visualizada uma mudança na composição da estrutura, com significativo aumento de instituições aderindo a modalidades de múltiplas carteiras e redução de instituições que operam exclusivamente com carteira comercial e exclusivamente com carteira de investimento.

Alguns dos efeitos da abertura comercial-financeira e das modificações na estrutura bancária provenientes das medidas governamentais foram o aumento da participação de instituições estrangeiras no país e, um consistente processo de fusões e aquisições, de ambas as origens de capital, que resultou em considerável elevação do grau de concentração .

A observação sobre o aumento da concentração bancária no Brasil realizada por pode ser visualizada na . Entre as metades das décadas de 1980 e 1990, com redução da concentração, levando em consideração o número de instituições. Esse cenário passou se inverter a partir de 1994, chegando em 2019 a um nível aproximado ao observado no início da década de 1980.

De acordo com Strachman e Vasconcelos *apud* , o aumento da concentração bancária pode ser prejudicial ao crescimento econômico, uma vez que, com maior participação de mercado, as instituições bancárias acabam por obter a prerrogativa de determinar seus preços, comportamento este observado em .

Segundo e por outra perspectiva, o ganho de escala, onde o cenário de aumento do tamanho das instituições, das operações de crédito e redução de custos operacionais atua melhorando a remuneração dos depósitos podendo atuar na redução dos juros finais pagos pelos clientes.

Outra possível tendência para a concentração bancária seria a redução do risco das operações, implicando em redução de custos, obtida por meio expansão geográfica, setorial e de produtos financeiros. Porém os possíveis efeitos da concentração dependem de uma série de condições, principalmente em torno da eficiência e do nível de concorrência no mercado .

O aumento da participação estrangeira no setor bancário brasileiro durante a década de 1990, evidenciado por pode ser observado na . Esse aumento ocorreu principalmente através do controle acionário, com elevação acentuada na segunda metade da década de 1990 até o início da década de 2000. Ocorrendo redução em instituições nacionais, estrangeiras e nacionais com participação estrangeira.

Durante este período, a inclinação para aplicação massiva em títulos públicos se dava diante a manutenção de elevadas taxas de juros, tornando o crédito para empreendimentos privados de elevado risco, e consequentemente elevando substancialmente o *spread* bancário e reduzindo a oferta de crédito .

A expectativa com a entrada de instituições estrangeiras era que houvesse elevação da concorrência e, consequentemente, redução no *spread* bancário, aumento da concessão de crédito, melhoria da qualidade e diversificação dos produtos financeiros, avanços em tecnologias, ou seja, uma elevação na eficiência do setor .

Porém, o que se observou foi a adoção de postura conservadora por partes dos bancos estrangeiros, com estratégia de ativos inclinada para negociação de títulos públicos, e passivos direcionados para a captação de recursos advindos de grupos de rendas média e alta, com exceção dos bancos públicos que concentraram em operações de crédito .

a demonstra o comportamento da relação crédito/PIB no Brasil, que entre a segunda metade da década de 1990 até a meados da primeira metade da década de 2000 sofreu significativa queda, ficando abaixo dos 25%. Após esse período a oferta de crédito sofreu uma expansão exponencial atingindo patamares acima de 50% do PIB.

Durante o período citado, foi observado no setor bancário brasileiro os maiores níveis de *spread* praticados no mundo, associado a um quadro econômico instabilidades e baixos crescimento e desenvolvimento. Esse cenário encontra embasamento em estudos teóricos e empíricos que demonstram que um sistema financeiro desenvolvido favorece o crescimento e desenvolvimento econômico .

A demonstra a evolução do saldo da carteira de crédito anual em termos correntes entre 1990 e 2020, podendo ser visualizada uma expansão exponencial de crédito a partir do início da década de 2000, com leve recuo na até metade da década de 2010 e retoma ultrapassando máxima anterior.

Diante o levantamento, o setor bancário brasileiro durante o período avaliado passou por diversas transformações em sua estrutura no que tange a concentração de mercado, aumento da participação de capital estrangeiro por meio de controle acionário, redução da participação pública.

Em relação aos indicadores foi verificado que, entre a década de 1980 até metade da década de 1990, no cenário hiperinflacionário, mesmo com redução da concentração bancária, os indicadores de eficiência de intermediação financeiras como o *spread* bancário e a relação crédito/PIB estavam em níveis considerados ineficientes e muito destoantes em comparação a outros países e regiões.

A partir de 1995 se observou mudanças significativas no setor bancário, com nova concentração, redução de instituições nacionais devido o controle acionário por capital estrangeiro, e expressiva redução no *spread* bancário e a partir de 2004 uma mudança significativa na relação crédito/PIB.

O Sistema Financeiro como um todo, em sua organização entre agentes normatizadores, supervisores e operadores, além suas instâncias — algumas já citadas —, possui uma organização em torno da Base Monetária e dos Agregados Monetários, configurando a Oferta de Moeda de determinado País.

A Base Monetária () é configurada pelo total de cédulas e moedas em circulação no país e das Reservas Bancárias — em poder das instituições ou depositadas no Banco Central — tido como emissão primária de moeda e configurando o passivo monetário, resultado líquido de todas as operações ativas e passivas do Banco Central. .

Em 1995 foi introduzido um novo conceito de Base Monetária ampliada, que possui maior capacidade de explanar os preços da economia no Brasil, que o conceito restrito, uma vez que trazem maior percepção do fator substituição entre a moeda em seu conceito convencional e os demais ativos financeiros disponíveis, incluindo os passíveis como depósitos compulsórios e títulos públicos federais .

Entre os Agregados Monetários estão o Meios de Pagamento na forma restrita (), configurado por cédulas e moedas em poder dos agentes econômicos e seus depósitos à vista, que podem ser utilizados prontamente para pagamentos de bens e serviços. O conceito de Meios de Pagamentos Ampliado adiciona à moeda legal agregados que são considerados de elevada liquidez () e (). .

No conceito de Meio de Pagamento Ampliado o Agregado Monetário (), contempla o Agregado Monetário adicionado do resultante das emissões primárias por instituições depositárias no mercado interno, capazes de realizar a multiplicação de crédito, consistindo em depósitos especiais remunerados (), depósitos de poupança () e títulos ().

No conceito ampliado, o Agregado Monetário contempla o Agregado Monetário () adicionado das captações internas intermediadas pelos fundos de renda fixa e das carteiras de títulos públicos federais registrados no Sistema Especial de liquidação e e Custodia (Selic). Consiste em quotas de fundos de renda fixa () e operações compromissadas registradas no Selic (operações compromissadas lastreadas em títulos públicos federais) () .

Existe ainda o conceito de poupança financeira composto pelo Agregado Monetário contemplando o Agregado Monetário mais a carteira livre de títulos públicos do setor não financeiro de elevada liquidez () e o Agregado Monetário () que engloba o () mais a capacidade de aquisição de cartões de crédito () .

De acordo com a teoria quantitativa da moeda, o nível de preços () em uma economia guarda relação com a quantidade de moeda em circulação () e a velocidade () em que circula a moeda na economia — frequência média em que uma unidade monetária é consumida em um período de tempo —, diante o produto seu produto real (), com a premissa que no curto prazo o produto e a velocidade a moeda são constantes .

O gráfico demonstra a evolução da Base Monetária entre os anos de 1995 e 2020 com elevação exponencial e constante durante este período, em termos constantes.

No que tange a abordagem microeconômica, as instituições bancárias como sociedade anônimas e instituições supervisionadas pelo Banco Central, são obrigadas a divulgar seus resultados em forma de demonstrações contábeis. A partir destas demostrações podem ser observados e extraídos dados e indicadores generalizados sobre a operação das instituições.

Os dados são divulgados seguindo uma padronização para o setor, onde podem ser observados as receitas, despesas, ativos, passivos, patrimônio líquido e a partir destes , calculados índices e indicadores, para cada período de registo, buscando refletir a situação econômica e financeira, podendo ser analisado de forma evolutiva e comparativa com outras instituições.

Entre os principais indicadores para avaliação de resultados das instituições bancárias, estão os índices de Liquidez Geral e Liquidez Corrente, Endividamento e Composição do Endividamento, retorno sobre o ativo, retorno sobre o patrimônio líquido, margem ebitda, margem líquida e grau de alavancagem financeira.

O índice de Liquidez Corrente () mede a capacidade da instituição em honrar os compromissos com seus credores, definindo seu nível de solvência no curto prazo. É obtido pela razão entre o ativo circulante () e o passivo circulante (), indicando o quanto do ativo circulante está disponível para cumprir com cada unidade monetária da dívida de curto prazo .

O índice de Liquidez Geral () mede a capacidade da instituição honrar os compromissos com seus credores no longo prazo, definindo seu nível de solvência geral, é obtido pela razão entre a soma do ativo circulante () e recursos realizáveis no longo prazo () e a soma do passivo circulante () e exigível no longo prazo () .

O índice de endividamento (), mede a participação de capital de terceiros em relação aos financiamentos realizados com capital próprio. Quanto maior o indicador, maior a dependência da instituição de capital de terceiros para financiamento das suas operações, obtido pela razão entre o passivo () e o patrimônio líquido () .

A composição do endividamento () indica o percentual da dívida da instituição em relação que vence no curto prazo. Quanto maior for esse indicador, mais crítica é a situação da instituição, necessitando de melhores resultados para cumprir os compromissos com credores no curto prazo, sendo obtido pela razão entre o passivo circulante () e a soma do passivo ciculante e exigível a longo prazo () .

O Índice de Eficiência bancária () é um indicador que avalia a relação entre despesas administrativas, despesas com pessoal e resultado operacional, medindo o quanto a instituição desembolsa para gerar uma unidade de receita. É obtido por meio da razão entre a soma das despesas administrativas (), despesas com pessoal () líquidas da participação nos lucros () sobre a soma entre Margem Financeira () e receita () .

Outro indicador utilizado para avaliação da situação financeira das instituições bancárias é o obtido da relação entre as receitas de prestação de serviços () e as despesas administrativas () .

O retorno sobre o Ativo (), mede a rentabilidade da instituição diante a totalidade dos seus ativos. O quanto para cada unidade monetária investida na instituição é convertida em lucro líquido, obtida da relação entre o lucro operacional () e o ativo total () .

O Retorno sobre o Patrimônio Liquido () mensura a relação entre o lucro líquido () em o Patrimônio Líquido () da instituição, configurando o retorno dos investimentos para os sócios e acionista, para cada unidade monetária com recursos próprios aplicados na empresa .

A margem EBITDA () é obtida da relação entre o EBITDA — lucro antes dos juros, depreciação, amortização e impostos sobre a renda, configurando o lucro operacional da instituição — e a Receita Líquida (), revelando a capacidade da instituição na geração de caixa .

A Margem Líquida () é um indicador que demonstra a parte de cada unidade monetária das intermediações financeiras que foi convertida em Lucro Líquido, sendo obtida da relação entre o lucro líquido () e O resultado líquido da intermediação financeira () .

O grau de alavancagem financeira () consiste no indicador que captura o efeito da tomada de recursos de terceiros a um dado custo, alocados para ativos que possuam distintas taxas de retornos. Como se dá o aumento do lucro líquido através da estrutura de financiamento, definindo a parcela do retorno que seria melhor ou pior se estivessem financiando a operações totalmente com capital próprio .

O risco de crédito das instituições bancárias pode ser obtido por meio da relação entre o saldo da Provisão para Créditos de Liquidação Duvidosa () e do total da carteira de crédito (), obtidos através das contas 16900008 e 16000001

A participação de mercado () de cada instituição pode ser mensurada a partir da relação entre suas operações de crédito {} no total das operações de crédito do mercado, sendo obtido através da conta 1600001 .

Este capítulo levantou informações amplas sobre o setor bancário brasileiro, e identificando e macroeconômicas e microeconômicas refrentes a economia como um todo, setor financeiro, ao setor bancário e as instituições em si. No próximo capítulo serão levantados conceitos, definições e estudos sobre a evolução, decomposição e determinantes do *spread* bancário.

Este capítulo irá tratar sobre os principais aspectos e características do *spread* bancário. Na primeira parte serão abordados conceitos e definições gerais. Na segunda parte as características amplas do mercado Brasileiro. Na terceira parte sobre os estudos empíricos realizados no Brasil. O foco é identificar elementos que possam contribuir com o objeto deste estudo.

Por definição o *spread* bancário é obtido através da subtração percentual, entre a taxa de aplicação incidente nas operações de crédito, e a taxa de captação que remunera as aplicações financeiras, se configurando como a diferença entre a composição dos custos destas operações .

O *spread* bancário representa uma medida que sinaliza o desempenho dos bancos . É considerado um indicador de eficiência da economia, no sentido de favorecer o crédito e a atividade econômica. Em níveis elevados pode desfavorecer o crédito destinado para produção e consumo produtivos e estar associado com o fraco desempenho econômico .

Os estudos em torno do *spread* bancário ocorrem em três óticas: evolução, estrutura e determinantes (SOUZA (2014). Em Dick *apud* é destacada a importância de distinguir a abordagem em torno da estrutura e determinante do *spread* bancário, no sentido de complementariedade. O diagrama na ilustra as óticas de estudo do *spread* bancário.

A abordagem em torno da evolução visa analisar o comportamento ao longo do tempo, através de análises quantitativas e qualitativas, enquanto a ótica da estrutura busca identificar e analisar os componentes de resultado envolvendo receitas, despesas e provisões. Na abordagem sobre os determinantes é vislumbrado identificar as variáveis que explicam as variações do indicador ao longo dos períodos .

Vem se tornando relevantes os estudos em torno da decomposição do *spread* bancário. Entre os componentes explícitos estão a inadimplência, despesas administrativas, impostos diretos e indiretos e margem de lucro dos bancos conforme ilustrado abaixo .

* Sprd = *Spread*
* Ind = Inadimplência
* DA = Despesas Administrativas
* II = Impostos Indiretos
* ML = Margem de Lucro
* CP = Custo de Captação

Esta configuração dos componentes, contemplando a margem de lucro, despesas e riscos envolvidos nas operações de crédito vem desmistificar a comum abordagem do *spread* como o rendimento auferido pelos bancos Souza (2007) *apud* . Desta forma se configurando como a diferença entre o custos operacionais na ótica de precificação, que após descontados das receitas, remontam o lucro do banco .

Além da avaliação de seus componentes, o *spread* pode ser analisado conjuntamente por três características: enquanto a abrangência da amostra, conteúdo e origem da informação .

A abrangência da amostra consiste nas especificidades das operações de crédito das instituições e seu nível de agregação e granularidade . Uma análise agregada dessa característica pode ser dificultada pela existência de heterogeneidade do setor, ressaltando a importância de realizar análises do *spread* bancário em diferentes características e óticas .

A abordagem em torno do conteúdo está relacionada com os subcomponentes que envolvem a receita e as despesas das intermediações financeiras, podendo englobar, ou não, as tarifas e comissões sobre as taxas de captações e aplicação .

A origem da informação é analisada em dois cenários: *ex-ante* e *ex-post* . A perspectiva *ex-ante* se refere as expectativas das instituições bancárias em relação ao mercado de crédito e os riscos envolvidos, obtido por método de precificação envolvendo as taxas de captação e empréstimo .

O *spread* *ex-ante*, por se tratar de um indicador de planejamento, refletindo as expectativas das instituições bancárias em relação ao mercado, finda demonstrando-se mais volátil, não representando as taxas efetivas realizadas. As informações *ex-ante* são repassadas ao Banco Central que as divulgam .

No *spread ex-post* as margens são obtidas mediante a apuração dos resultados contábeis, através dos demonstrativos, considerando as receitas e custos efetivos, implicando nas taxas de intermediação e carteira realizadas pelas instituições financeiras . Nesse sentido, em termos médios, as taxas *ex-post* se demonstram mais estáveis .

O *spread ex-post*, sendo resultante da diferença entre as taxas de empréstimos e de captação realizadas pelas instituições, é obtida por meio das contas 71100001, 16000001, 81100008 e 41000007 (das demonstrações contábeis padronizadas). O *spread ex-post* se configura como o efetiva margem auferida pela instituição, diante seus resultados contabilizados .

SPR = *Spread ex-post*

ROP = Receitas das Operações de Crédito

OP = Operações de Crédito Média

DOC = Despesas das Operações de captação

DP = Depósitos médio

Reduções no *spread* *ex-post* não necessariamente significam aumento da eficiência da intermediação financeira, pois podem estar associadas a uma redução da inadimplência . Como observado em e o *spread* bancário é determinado de acordo com as características e os riscos envolvidos nas intermediações financeiras inerentes em cada estrutura de mercado.

No Brasil, a taxa de aplicação para crédito de recursos livres é pactuado entre instituição e tomador. Somente as operações de crédito envolvendo recursos direcionados são sujeitas à limites, não podendo exceder 12%a.a. mais a taxa referencial de juros .

No mercado bancário Brasileiro, o modelo consolidado de mensuração do *spread*, conforme , leva em consideração o saldo médio de capital emprestado, e a diferença entre as receitas de aplicação e despesas de captação, ocorrendo a classificação em *spread* bruto, direto e líquido

O Banco Central, em 1999, iniciou uma série de estudos e medidas com objetivo de reduzir a taxa de juros e o *spread* realizados no setor bancário Brasileiro, atuando na identificação e ajustes em variáveis econômicas influentes. Entre as primeiras medidas estavam a redução da taxa de compulsório para depósitos à vista e até a extinção para depósitos à prazo, redução do IOF e a redução da Selic .

A mostra a evolução do *spread* bancário Brasileiro médio entre os anos de 1994 e 2012, chegando a atingir 146.44%, com significativa queda ao longo desse período, atingindo 24.69% no final do período. Esta série foi descontinuada em 2012, passando a ser utilizada nova metodologia de cálculo.

O Banco Central, até 2007 utilizava metodologia para avaliação do spread bancário contemplando somente os recursos livres, o que não vinha a proporcionar uma avaliação mais aprofundada. Em 2008 houve uma modificação na metodologia de decomposição do *spread*, alterando o cálculo do custo médio de captação e detalhando classificações do crédito

Para o custo médio de captação passou a se utilizar a taxa média ponderada entre as taxas dos depósitos à prazo (CDB), em caderneta de poupança e à vista, a participação dos custos efetivos dos recolhimentos compulsórios em detrimento do custo de oportunidade

O BACEN mantém atualmente duas séries para o indicador: spread médio das operações de crédito (MOC) e Spread do Indicador de Custo de Crédito (ICC). As séries são disponibilizadas em termos totais e nas subdivisões de tipo de recursos, tipo de crédito e tomador.

Estas séries estatísticas representam estimativas baseadas nas informações repassadas pelas instituições bancárias das taxas de juros das operações de crédito e indicadores do mercado financeiro do custo médio do dinheiro para o custo médio de captação .

A série do Spread médio das operações de crédito é calculada a partir da diferença entre a taxa média de juros de novas operações de crédito no SFN e o custo de captação referencial médio de operações de crédito livre, direcionado e não rotativo podendo ser observados por tomador.

A mostra a visualização da evolução mensal do spread médio das novas operações de crédito contratadas entre janeiro de 2013 e julho de 2020. No período entre 2014 e 2017 se visualiza uma elevação de 10 p.p no spread total, recuando 8 p.p a patamar próximo ao início do período. É possível notar a grande disparidade entre os spread de recursos livres e direcionados.

A série do Spread do ICC, considera a diferença entre o Índice de Custo de Crédito — equivalente ao custo médio de juros das operações ativas da carteira do SFN — e o custo de captação médio ponderado, levando em consideração operações de crédito livre, direcionado e não rotativo, dividido em pessoa física e jurídica.

Na pode ser visualizada a evolução do spread do ICC, entre janeiro de 2013 e julho de 2020 com expressiva elevação entre 2014 e 2017, passando a decair até retormar a patamares similares ao início do período. Também pode ser notado a significativa diferença entre os *spreads* de recursos livres e direcionados.

O Indicador Custo de Crédito (ICC) consiste no custo médio de todas as operações de crédito abertas — independentes do período em que foram contratadas — que compõem a carteira de empréstimos, financiamentos e arrendamento mercantil das instituições do Sistema Financeiro Nacional (SFN) .

A traz a visualização da evolução do Índice de Custo de Crédito entre janeiro de 2013 e agosto de 2020, com máxima de 22.98% em 2017, com queda significativa a partir de 2020, chegando a atingir 17.27% em agosto de 2020.

Na literatura acadêmica não existe uma teoria formalizada acerca do *spread* bancário . Sendo verificados estudos empíricos que visam classificar, analisar e identificar variáveis micro e macroeconômicas influentes nesse indicador em diversas perspectivas.

A grande maioria dos estudos realizados no Brasil utilizam as medidas de *spread* bancário divulgadas pelo Banco Central, que remetem a uma perspectiva *ex-ante*, registrando as taxas planejadas na fase de concessão de crédito. E para as variáveis explicativa a grande maioria utiliza indicadores macroeconômicos

No ano de 1994, realizou estudo econométrico para verificar a influência da inflação e nível de atividade econômica no *spread* bancário ex-ante, encontrando relação direta do *spread* com a inflação e indireta com o nível de atividade econômica.

Em estudo dos determinantes macroeconômicos do *spread* bancário ex-ante, utilizou regressão múltipla para identificar as variáveis influentes (modelo abaixo). O estudo chegou ao resultado que alta volatilidade e as taxas da Selic são um dos principais determinantes desse indicador no setor bancário Brasileiro, identificando também a significância do nível de atividade industrial.

* (i= 0,…, 5) = parâmetros estimados;
* trend = tendência determinista que controla outras variáveis;
* selic = taxa Selic;
* adm = despesa administrativas;
* risk = proxy para o risco de crédito (spread do C-Bond sobre o rendimento dos títulos do Tesouro Americano de mesma maturidade;
* imp são impostos indiretos;
* comp = compulsório incidente sobre os depósitos à vista.

Em análise dos determinantes do *spread* bancário ex-post, utilizou variáveis explanatórias microeconômicas de cada instituição, por meio de dados em painel dinâmico, entre janeiro de 2000 e outubro de 2009, encontrando níveis significativos e diretos com o risco de crédito, grau de concentração e nível de atividade econômica, e indireta com a participação da instituição no mercado, não encontrando níveis significativos com origem de capital e tipo de organismo.

Outra observação em foi a forte relação do *spread* *ex-post* no momento atual com o momento anterior imediato, e que as instituições tendem a cobrar maiores taxas, quando maior o nível de concentração do mercado, não encontrando significância da Selic na determinação deste indicador.

Em foi desenvolvido modelo de dados macroeconômicos e microeconômicos em painel, de 64 instituições bancárias para avaliação de determinantes do *spread* *ex-post* no Brasil entre o primeiro trimestre de 2001 e o segundo trimestre de 2012, encontrando como relevantes as despesas administrativas, receita de serviços, índice de cobertura, PIB e o grau de concentração.

Em foi realizada análise dos fatores macroeconômicos e indicadores industriais que influenciam o *spread* bancário ex-ante, através de análise de regressão linear multivariada utilizando 18 variáveis em quatro modelos. Chegando a conclusão que o aumento da atividade industrial, a redução do desemprego e o consumo atuam na diminuição do *spread* bancário.

Os modelos desenvolvidos por demonstraram que há uma relação significativa e direta entre *spread* e: inadimplência, IPIs (bens de capital, intermediários, semiduráveis, não duráveis e consumo duráveis), Selic, PIB, desemprego e o EMBI+ (medida de taxa de risco-país). As relações indiretas com o *spread* foram encontradas: no IPI de bens de consumo e geral, IPCA, saldo da carteira de crédito e índice de vendas no varejo.

O estudo de teve foco em abordagem microeconômica, ao buscar identificar a influência das variações de indicadores financeiros-contábeis no *spread* em 26 instituições bancárias, através de regressão em dados em painel. Encontrando relações significativas diretas com a alavancagem financeira, retorno sobre o patrimônio líquido, EBITDA, Ativo Total e eficiência.

No modelo de foi encontrada relação significativa e indireta do *spread* com a participação de capital de terceiros e, não identificada relação significativa com a composição do endividamento, retorno sobre ativos e a liquidez corrente.

De acordo com , existem poucos estudos inclinados para os determinantes do *spread* *ex-post* no Brasil, onde identificaram o estudos de Guimarães (2002). Foram identificados ainda os estudos acerca do *spread* ex-pots de Fipecafi (2004) *apud* e Matias (2006) *apud*

Em foi realizado estudo de apuração de resultados, ex-post, baseado em demonstrações contábeis entre o 1º semestre de 2005 de instituições que representavam 75,8% do ativo total e 76% do total de crédito. Chegando a um resultado médio de *spread* bruto de 7,6% para pessoa física e 3,2% para pessoa jurídica, e *spread* líquido de 1,6% para pessoa física e 0,5% para pessoa jurídica.

A e a trazem o resumo dos principais estudos empíricos sobre *spread* bancário ex-ante no Brasil, com resultados obtidos através de modelagem econométrica com utilização de regressão, tomando variáveis micro e macroeconômicas como explanatórias e demonstrando a relação com o spread ex-ante.

Entre os estudos da e que avaliaram a Selic e as despesas administrativas, há um consenso que estas variáveis possuem uma relação de determinação direta com o *spread ex-ante*. Em três estudos que avaliaram impostos indiretos e receita de serviços foi encontrada relação direta com o *spread ex-ante*.

Ainda analisando a e a , dois estudos chegaram a resultados diferentes para os efeitos da volatilidade da Selic no *spread ex-ante*. Os efeitos do IPCA foram testados em três estudos, os dois mais recentes encontraram uma relação indireta com a variável dependente. Em três estudos que examinaram o IGP, dois encontram relação direta, sendo que um deles foi repetido em período anterior e encontrou relação indireta.

A traz o resumo dos estudos empíricos dos determinantes do *spread ex-post* no Brasil, por meio de modelos econométricos utilizando regressão. Destaca-se que, entre os estudos, dois encontraram significância de influência direta com o grau de concentração e o *spread* ex-post. E dois dos estudos chegaram a resultados opostos para os de posição de market share e a variável dependente.

Este capítulo verificou os principais conceitos, características e estudos acerca do *spread* bancário no Brasil, identificando as óticas de análise por evolução, composição e determinantes através da abrangência da amostra, conteúdo e origem da informação. E que as maiores limitações estão na dificuldade de desagregação de informações para uma análise mais aprofundada.

No próximo capítulo, será descrita a metodologia de trabalho com a formulação das hipóteses baseado nas informações e levantamentos dos capítulos anteriores, nos estudos pesquisados e na teoria econômica, através da coleta, tratamento e análise de dados.

Neste capítulo serão descritos os principais procedimentos metodológicos, técnicas e ferramentas que serão utilizados neste trabalho, visando organizar as etapas da pesquisa e permitir um maior nível de reproducibilidade, revisão e refutabilidade da mesma.

Este trabalho está sendo desenvolvido e editado em ambiente R Markdown com utilização de linguagem Latex para padronização de textos, figuras e tabelas, e as linguagens R e Python para coleta, limpeza, tratamento, análise, visualização e modelagem e estimação econométrica dos conjuntos de dados.

Para fins de análise de dados, este trabalho atuará no intervalo de tempo entre o primeiro trimestre de 1999 e o terceiro trimestre e 2020.

Para efeitos de identificação de variáveis macroeconômicas que atuam como componentes implícitos e explícitos do *spread* serão avaliadas a taxa Selic, Taxa de Compulsório, Base Monetária, Oferta de Crédito, Oferta de Moeda. Estes dados serão obtidos de forma secundária nos bancos de dados abertos do Banco Central, IPEA, IBGE e Receita Federal do Brasil.

Os dados de resultados, operação, indicadores e estrutura de capital das instituições bancárias serão obtidos de forma secundária nos banco de dados abertos do Banco Central e da Comissão de Valores Monetários, consistindo em demonstrações contábeis trimestrais padronizadas informadas a estas instituições supervisoras.

Para construção dos modelos econométricos a serem estimados se partirá de alguns pressupostos teóricos norteadores obtidos através da pesquisa bibliográfica e concepções desenvolvidas durante a pesquisa, com intuíto embasar a seleção das variáveis a serem testadas e incluídas no modelo final.

Será assumido que o *spread* bancário é definido diante um conjunto de fatores endógenos, definidos por questões microeconômicas envolvendo as operações de cada instituição e dos mercados financeiro e bancário, e fatores exógenos provenientes de questões macroeconômicas, afetando diretamente ou indiretamente as operações.

O *spread* () será abordado dentro de uma concepção de precificação, diante um conjunto de variáveis explícitas como custo de captação (), capital emprestado () — insumo das operações de crédito —, impostos variáveis (), despesas administrativas (), lucro líquido () e inadimplência ().

Na visão microeconômica, assume-se que o *spread* bancário não se configura na margem de lucro dos bancos, não cabendo abordagem de *spread* bruto, direto e líquido. E que o *spread* bancário se relaciona com os resultados das instituições, colaborando com a solidez do setor, não cabendo a inclusão no modelo de variáveis que rementam a resultados e calculadas a partir destas.

Na abordagem macroeconômica, o *spread* bancário é tido como um indicador fundamental e determinante para o nível de desenvolvimento econômico de determinado país ou região a medida que se relaciona com a determinação de nível e crédito produtivo capaz de gerar renda, influenciado por variáveis macroeconômicas relacionadas a regulação e políticas monetárias e e fiscais.

Nesse sentido aqui é estabelecida a compreensão que o nível de atividade econômica, industrial, produtividade, desemprego e produto interno bruto de mercados, países e regiões guardam relação com o *spread* bancário, e não o contrário, mesmo que aja a compreensão da abordagem em torno das expectativas dos agentes, será mantida a abordagem econômica, não considerando essas variáveis como determinantes do *spread ex-post*.

O primeiro modelo a ser desenvolvido buscará testar e selecionar variáveis macroeconômicas e microeconômicas que exerçam significativa influência, de forma implícita e explícita no *spread* bancário *ex-post*. Partindo da definição geral tautológica de *Spread* (), resultado da diferença entre a taxa de aplicação () e a taxa de captação ()

O *Spread Ex-post* (), será calculado a partir dos resultados contábeis, resultante diferença entre a relação entre receitas de operações de crédito () e operações de crédito médio (), e a relação entre despesas de captação () e depósitos médio ()

Diante o componentes explícitos do *Spread ex-post* é possível decompor da seguinte forma:

O termo destacado entre chaves assume a simplificação de multiplicador de aplicação (), atuando sobre o custo de captação para gerar a receita da operação de crédito (), tendo como componentes explícitos os percentuais de margem líquida desejada (), impostos indiretos (), despesas administrativas (), inadimplência sobre os juros () e inadimplência sobre a amortização () .

Multiplicador da Operação de crédito, baseado no capital emprestado e sua origem.

Onde as operações de créditos () podem ser decompostas de acordo com a origem: capital próprio () e depósitos à vista () — Líquido de Compulsório — remunerados a um custo de oportunidade ( e ) — já que não são remunerados — e depósitos à prazo () remunerados à uma taxa de captação ().

Considerando o custo de oportunidade para operações de crédito como o Juros de Capital Próprio (). A partir do entendimento que o juros de capital próprio total () é composto das expectativas da soma do juros de capital próprio das destinado as operações de crédito () e ao destinado as operações de serviços ().

Diante os pressupostos, o primeiro modelo irá verificar a influência das variações de variáveis explícitas e implícitas no *Spread Ex-post*, tendo no primeiro bloco variáveis microeconômicas e o segundo bloco as variáveis macroeconômicas, selecionando para o modelo final somente as que apresentarem significância estatística.

Na construção do primeiro modelo econométrico serão adotadas simplificações para variáveis de resultado, eliminando as que possuem caráter constante, as obtidas por meio de resultado e por não possuirem dados, utilizando uma *proxy*.

O segundo modelo econométrico testará as variáveis implícitas e explícitas com significância estatística do primeiro modelo, atuando sobre a rentabilidade bancária , conforme modelos especificados. Para a rentabilidade será considerada a razão entre o lucro líquido () e a Receita das Operações de crédito ().

Diante a definição dos modelos, seguem abaixo as hipóteses conceituais baseadas em concepções teóricas obtidas na pesquisa bibliográfica e das concepções desenvolvidas durante a pesquisa. O conjunto de hipóteses se apresenta na forma objetiva incluindo a expectativas para cada variável e contemplando os dois modelos construídos, com breve explanação sobre a mesma.

: A proporção das operações de crédito com capital próprio () guarda relação direta com o *spread ex-post* () e direta com a rentabilidade bancária ().

: A proporção das operações de crédito com depósitos à vista () mantém uma relação direta com *spread ex-post* () e direta com a rentabilidade bancária ().

: A proporção das operações de crédito com depósitos à prazo () atuam na inversa do *spread ex-post* () e inversa com a rentabilidade bancária () do período.

: O multiplicador de aplicação bancária () possui relação direta com o *spread ex-post* () e mantendo-se constantes os demais elementos da além da margem de lucro (), possui relação direta com a rentabilidade bancária ()

: O juros de capital próprio das operações de crédito () guarda relação direta com *spread ex-post* () e direta com a rentabilidade bancária ()

: O ativo total () mantém uma relação inversa com *spread ex-post* () e nula com a rentabilidade bancária ()

: A participação de mercado das instituições () guarda relação inversa com o *spread ex-post* () e nula com a rentabilidade bancária ()

: o grau de concentração de mercado mantém relação direta com *spread ex-post* () e nula com a rentabilidade bancária ()

: O tipo de instituição () exerce influência na determinação do nível de *spread ex-post* () e nível da rentabilidade bancária ()

: A origem de controle de capital excerce influência do nível *spread ex-post* () e na determinação no nível de rentabilidade bancária ()

: O caráter da instituição () atua na determinação do nível do *spread ex-post* () e no nível da rentabilidade bancária ()

: A taxa Selic mantém uma relação direta no *spread ex-post* () e nula com a rentabilidade bancária ()

: A taxa de inflação () atua de forma direta no *spread ex-post* () e inversa com a rentabilidade bancária () real.

: A taxa de compulsório () atua de forma direta no *spread ex-post* () inversa com a rentabilidade bancária ()

Ao reduzir a quantidade dos depósitos à vista que serão destinados para as operações de crédito, os recolhimentos compulsórios aumentam a necessidade de utilização de depósitos à prazo que são remunerados, aumentando o *spread* e a rentabilidade no curto prazo.

: A quantidade total dos meios de pagamentos atua de forma inversa no *spread ex-post* () e nula com a rentabilidade bancária ()

: A velocidade da moeda () atua de forma direta no *spread ex-post* () e direta com a rentabilidade bancária ()

: O *spread ex-ante* () atua de forma direta *spread ex-post* () e direto com a rentabilidade bancária ()

Para a averiguação dos efeitos dos componentes do *spread ex-post* na rentabilidade das instituições bancárias serão utilizados modelos de regressão linear multivariada. Os modelos de regressão múltipla buscam, através de técnicas estatísticas e matemáticas, prever o comportamento de uma dada variável dependente, diante um conjunto de variáveis explanatórias .

O modelo econométrico a ser utilizado será o método de dados em painel, denominado *Cross Section*, que combina séries temporais e dados em corte transversal. Este modelo busca captar diferenças individuais de comportamento, possibilitando combinar os dados para fins de estimação e inferência, posteriormente realizados testes de regressão e estimação. .

O método *Cross Section* pode ser realizado por meio de três modelos de estimação que são: i) Modelo de regressão aparentemente não relacionadas (SUR); ii) Modelo de variável binárias — efeitos fixos — e iii) modelo de componentes estocásticos — efeitos aleatórios — . Serão testados os três métodos buscando selecionar o mais adequado ao modelo econométrico e ao conjunto de dados.

No modelo de regressão de dados aparentemente não relacionados — SUR —, os parâmetros dos diferentes grupos em corte transversal diferem entre si, porém são constantes ao longo do tempo. Os modelos podem ser estimados com suas funções de forma conjunta ou separada, onde esta última é indicada quando há correlação dos erros

No modelo de variável binárias — ou efeitos fixos —, o intercepto é abordado como um parâmetro desconhecido e fixo, onde as inferências são aplicadas somente ao cojunto de dados dos grupos do corte transversal do qual está disponível .

O modelo de componentes estocásticos — ou efeitos aleatórios —, considera cada grupo do conjunto de dados como uma amostra aleatória de uma população maior, onde os interceptos são encarrados como resultados aleatórios da distribuição populacional de interceptos de grupos, realizando assim uma inferência da população de grupos .

Será aplicada a técnica de *Cross Validation k-fold*, que visa dividir de forma aleatória o conjunto de dados em k grupos, de dimensão aproximada. Onde o primeiro grupo é tratado como conjunto de validação, e o método é ajustado no k - 1 conjuntos restantes. Esse método é útil para testar variáveis, selecionar parâmetros, função preditiva e acurácia para seleção do modelo final .

O mecanismo do *k-fold* consiste em calcular o erro quadrático médio nas observações do grupo retido, repetindo esse procedimento k vezes, onde a cada rodada, um grupo diferente é tratado como um conjunto de validação. Este processo resulta em k estimativas do erro de teste, MSE1, MSE2,. . . , MSEk. A estimativa de CV *k-fold* é obtida pela média desses valores .

Serão testados métodos de aprendizado não supervisionado. O aprendizado não supervisionado tende a ser mais subjetivo para análises com foco preditivo. O aprendizado não supervisionado geralmente é empregado como uma etapa da análise exploratória de dados, não há como verificar o resultado pois não é conhecida a verdadeira resposta .

Entre os métodos de aprendizagem não supervisionada estão o *Principal Component Analysis* e o *Clustering*. No método de *Clustering* que visa dividir o conjunto de dados em subgrupos distintos que contenham homogeneidades entre as observações. Existem as técnicas *K-Means*, *Hierarchical Clustering* e *Practical Issues in Clustering*

A técnica de aprendizado não supervisionado a ser aplicada será o *K-Means*, que possui como característica o particionamento do conjunto de dados em k grupos — *clusters* —, especificados e não sobrepostos, o que implica que cada observação pertence a menos a um grupo e não pertence a nenhum outro grupo. O propósito principal é buscar a mínima variação dentro de cada *cluster* .

Este capítulo buscou demonstrar os principais conceitos e procedimentos metodológicos a serem seguidos para realização das análises dos objetivos propostos. No próximo capítulo se buscará aplicar os conceitos procedimentos propostos, de forma analítica e crítica, com flexibilidade diante observações na fase de exploração, tratamento e análise dos conjuntos de dados, podendo gerar novas abstrações a serem incorporadas.