XXXIII Congresso Brasileiro de Custos
17, 18 e 19 de novembro de 2026
- Porto de Galinhas / PE -

Economias de escala no setor bancário brasileiro, evidências de deseconomias nos maiores conglomerados

Cristiano Augusto Forti (UFU) - crforti@gmail.com

Resumo:

Este artigo examina a estrutura de custos do setor bancário brasileiro, com dois objetivos. O primeiro é estimar a decomposição entre custos fixos e variáveis e sua evolução. O segundo é verificar a existência de economias de escala e identificar quais instituições delas se beneficiam. Foram utilizados dados anuais do IF.data do Banco Central do Brasil, no nível de conglomerado prudencial, entre 2015 e 2024, totalizando 245 instituições e 2.341 observações. Utilizamos a função custo translog na abordagem de intermediação, com três produtos e três preços de insumos, e o modelo de custos assimétricos de Anderson, Banker e Janakiraman (2003), ambos por efeitos fixos. Os resultados indicam que a parcela fixa dos custos se situa entre 55% e 60% no conjunto do período, e que essa parcela se eleva de 68,0% no triênio de 2015 a 2017 para 87,7% no biênio de 2023 e 2024, evidenciando aumento expressivo da rigidez da estrutura de custos. Existem economias de escala na produção bancária, com medida de escala de 0,749 e rejeição da hipótese de retornos constantes, porém elas se concentram nas instituições de menor porte e convertem-se em deseconomias nos dois segmentos superiores, alcançando 1,248 no segmento S1. O comportamento assimétrico dos custos apresenta o sinal previsto pela teoria, mas significância estatística apenas marginal. Conclui-se que grandes bancos não possuem menores custos unitários, o que enfraquece o argumento de ganho de escala como justificativa para operações de fusão e aquisição entre instituições de grande porte.

Palavras-chave: *Estrutura de Custos. Economias de Escala. Custos Assimétricos. Bancos.*

Área temática: *Custos como ferramenta para o planejamento, controle e apoio a decisões*

Economias de escala no setor bancário brasileiro, evidências de deseconomias nos maiores conglomerados

RESUMO

Este artigo examina a estrutura de custos do setor bancário brasileiro, com dois objetivos. O primeiro é estimar a decomposição entre custos fixos e variáveis e sua evolução. O segundo é verificar a existência de economias de escala e identificar quais instituições delas se beneficiam. Foram utilizados dados anuais do IF.data do Banco Central do Brasil, no nível de conglomerado prudencial, entre 2015 e 2024, totalizando 245 instituições e 2.341 observações. Utilizamos a função custo translog na abordagem de intermediação, com três produtos e três preços de insumos, e o modelo de custos assimétricos de Anderson, Banker e Janakiraman (2003), ambos por efeitos fixos. Os resultados indicam que a parcela fixa dos custos se situa entre 55% e 60% no conjunto do período, e que essa parcela se eleva de 68,0% no triênio de 2015 a 2017 para 87,7% no biênio de 2023 e 2024, evidenciando aumento expressivo da rigidez da estrutura de custos. Existem economias de escala na produção bancária, com medida de escala de 0,749 e rejeição da hipótese de retornos constantes, porém elas se concentram nas instituições de menor porte e convertem-se em deseconomias nos dois segmentos superiores, alcançando 1,248 no segmento S1. O comportamento assimétrico dos custos apresenta o sinal previsto pela teoria, mas significância estatística apenas marginal. Conclui-se que grandes bancos não possuem menores custos unitários, o que enfraquece o argumento de ganho de escala como justificativa para operações de fusão e aquisição entre instituições de grande porte.

Palavras-chave: Estrutura de Custos. Economias de Escala. Custos Assimétricos. Bancos.

Área Temática: Custos como ferramenta para o planejamento, controle e apoio a decisões.

1 INTRODUÇÃO

A gestão de custos ocupa posição central nas decisões empresariais, e o setor bancário reúne características que tornam essa gestão particularmente relevante. Bancos conduzem a política monetária, operam os sistemas de pagamento e constituem fonte essencial de crédito para famílias, empresas e governos. No Brasil, o setor apresenta elevado grau de concentração, o que confere às decisões de estrutura de custos das maiores instituições consequências que extrapolam seus próprios resultados.

O período recente acrescentou complexidade a esse quadro. A implantação do Pix em novembro de 2020, a entrada de instituições de modelo integralmente digital e a substituição progressiva de rede física por plataformas tecnológicas alteraram simultaneamente a forma de produzir serviços bancários e a composição dos custos de produzi-los. A transformação sugere que a estrutura de custos do setor pode ter se modificado de maneira relevante, questão que permanece pouco documentada com evidência sistemática.

A literatura oferece três contribuições principais para o problema. A primeira estabelece o arcabouço de decomposição entre custos fixos e variáveis e suas consequências sobre alavancagem operacional e risco (Horngren, Sundem & Stratton,

2004; Martins & Rocha, 2010). A segunda demonstra que os custos não respondem de forma simétrica a aumentos e reduções de atividade, comportamento que Anderson, Banker e Janakiraman (2003) atribuíram a custos de ajustamento e a decisões deliberadas de manutenção de capacidade, e que a literatura subsequente refinou ao incorporar as expectativas dos gestores (Banker, Byzalov & Plehn-Dujowich, 2014) e as fases do ciclo econômico (Bubeck & Silva, 2025). A terceira, específica do setor financeiro, define o produto bancário e a tecnologia de produção, com a abordagem de intermediação proposta por Sealey e Lindley (1977) e a função custo translog de Christensen, Jorgenson e Lau (1973), que permitem medir economias de escala sem impor restrições prévias sobre o formato da curva de custo médio.

Essas três frentes, contudo, raramente se encontram. Estudos de comportamento de custos concentram-se em companhias abertas de setores não financeiros, ao passo que estudos de eficiência bancária privilegiam a distância de cada instituição à fronteira de melhores práticas, e não a elasticidade do custo em relação à produção. Uma consequência dessa lacuna merece registro. A proporção de custos fixos e a existência de economias de escala são grandezas distintas, frequentemente tratadas como equivalentes, pois a primeira descreve a composição do custo em determinado nível de atividade e a segunda descreve como o custo reage a variações desse nível. A confusão entre ambas é fonte recorrente de inferência equivocada, particularmente quando se invoca ganho de escala para justificar operações de fusão e aquisição.

Diante desse contexto, esta pesquisa investiga duas questões. Qual é a estrutura de custos fixos e variáveis dos bancos brasileiros e como ela evoluiu no período recente? Existem economias de escala na produção bancária brasileira e, em caso afirmativo, elas beneficiam as instituições de maior porte?

O artigo pretende contribuir em quatro aspectos. Primeiro, ao estimar formalmente a elasticidade de escala da função custo bancária brasileira e ao decompô-la por segmento prudencial, o que permite identificar em que patamar de porte os ganhos se esgotam. Segundo, ao documentar a trajetória da rigidez da estrutura de custos ao longo de uma década que abrange recessão, pandemia e transformação digital. Terceiro, ao demonstrar que a medida de nível de atividade altera de forma substantiva as conclusões, uma vez que a receita bancária incorpora o ciclo de juros e produz viés de atenuação. Quarto, ao aplicar ao setor bancário o modelo de custos assimétricos, cuja evidência nacional concentra-se em empresas não financeiras.

O trabalho está organizado em cinco seções, além desta introdução. A segunda apresenta o referencial teórico sobre comportamento de custos, produção bancária e economias de escala. A terceira detalha os procedimentos metodológicos, a amostra e os dois modelos estimados. A quarta analisa os resultados. A quinta apresenta as conclusões, as limitações e as sugestões de pesquisa futura.

2 REFERENCIAL TEÓRICO

2.1 Custos e despesas: conceito, classificação e comportamento

Os custos de uma organização correspondem aos gastos incorridos para manter as atividades de produção de bens ou de prestação de serviços, enquanto as despesas se referem aos gastos necessários à realização dos processos administrativos. O somatório desses gastos constitui o total de custos e despesas da

entidade, cujo comportamento não é uniforme (Martins & Rocha, 2010; Richartz & Ensslin, 2014).

A classificação tradicional distingue custos fixos e variáveis conforme sua reação ao volume de atividade. Os custos variáveis oscilam com o volume produzido ou vendido, ao passo que os custos fixos permanecem estáveis dentro de determinado intervalo relevante de produção (Martins, 2010; Martins & Rocha, 2010). A proporção entre esses dois componentes define a estrutura de custos da entidade e determina seu grau de alavancagem operacional, uma vez que estruturas com maior participação de custos fixos ampliam a sensibilidade do resultado a variações na receita, elevando simultaneamente o potencial de ganho e o risco operacional (Horngren, Sundem & Stratton, 2004; Guerra, Rocha & Corrar, 2007).

Essa classificação dicotômica pressupõe simetria, ao admitir que os custos se comportem da mesma forma quando o volume de atividade cresce e quando ele diminui. Anderson, Banker e Janakiraman (2003) demonstraram que tal pressuposto não se sustenta empiricamente. Ao examinar despesas de vendas, gerais e administrativas de empresas norte-americanas, os autores constataram que essas despesas aumentam mais intensamente diante de elevações nas vendas do que se reduzem diante de quedas de mesma magnitude, comportamento que denominaram custos assimétricos.

A explicação proposta é econômica e não contábil. Quando a demanda recua, o gestor decide entre manter recursos ociosos ou dispensá-los, e a remoção de recursos comprometidos implica custos de ajustamento, tais como rescisões, perda de capital humano específico e recontratação caso a demanda retorne. Se a queda é avaliada como transitória, manter a capacidade instalada é a decisão racional, e a assimetria observada reflete essa escolha (Anderson et al., 2003; Banker, Byzalov & Chen, 2013).

A literatura subsequente refinou o argumento. Banker, Byzalov e Plehn-Dujowich (2014) demonstraram que a intensidade da assimetria depende das expectativas do gestor quanto à demanda futura, pois quando a queda corrente é precedida de outra queda, o gestor revisa suas expectativas e passa a cortar recursos com mais intensidade, atenuando ou revertendo a assimetria. Banker e Byzalov (2014) sistematizaram as condições sob as quais o comportamento assimétrico deve ser esperado, destacando o papel da intensidade de ativos e da intensidade de mão de obra como determinantes da magnitude dos custos de ajustamento.

No contexto brasileiro, Medeiros, Costa e Silva (2005) encontraram aplicação apenas parcial do modelo de assimetria, resultado posteriormente corroborado por Richartz e Borgert (2014). Costa, Marques, Santos e Lima (2013) estenderam a análise à América Latina, e Richartz, Borgert e Lunkes (2014) investigaram o papel dos gastos com mão de obra e da proporção de custos fixos na assimetria de empresas brasileiras de capital aberto, concluindo que companhias com maior participação de custos fixos tendem a apresentar maior assimetria. Richartz e Borgert (2021) aprofundaram a identificação dos fatores explicativos do fenômeno entre as companhias listadas na B3.

Duas contribuições recentes são relevantes para o desenho deste estudo. Ibrahim, Ali e Aboelkheir (2022), em revisão sistemática de vinte e sete anos de pesquisa, apontam a escassez de trabalhos em setores regulados como lacuna persistente. Bubeck e Silva (2025), examinando 184 companhias abertas brasileiras entre 2010 e 2023, demonstram que o comportamento assimétrico varia conforme a fase do ciclo econômico, achado pertinente ao período aqui examinado, que abrange a recessão de 2015 e 2016, o choque da pandemia e a recuperação subsequente.

2.2 A produção bancária e a mensuração de custos no setor

A aplicação dos conceitos de custo a instituições financeiras exige uma definição prévia do que constitui o produto bancário, questão sobre a qual a literatura oferece duas abordagens concorrentes.

Na abordagem da produção, o banco é tratado como prestador de serviços que processa contas e transações, os produtos são medidos pelo número de contas ou de operações, e os custos considerados são apenas os operacionais, excluindo as despesas de captação. Na abordagem da intermediação, proposta por Sealey e Lindley (1977) e predominante na literatura contemporânea, o banco é tratado como intermediário que capta recursos de poupadores e os transforma em ativos rentáveis. Nessa concepção, os depósitos e demais captações são insumos, não produtos, e o custo total inclui as despesas de intermediação financeira, ao lado das despesas de pessoal e administrativas.

A escolha entre as abordagens não é indiferente par os resultados. Berger e Humphrey (1997), em revisão de 130 estudos de eficiência em instituições bancárias, observam que a abordagem da intermediação é mais adequada quando o objeto de análise é a instituição como um todo, enquanto a abordagem da produção se ajusta melhor ao exame de unidades específicas, como agências. Como o interesse deste trabalho recai sobre a instituição, foi adotada a abordagem da intermediação.

A forma funcional predominante para representar a tecnologia bancária é a função custo translog, introduzida por Christensen, Jorgenson e Lau (1973). Trata-se de uma aproximação de segunda ordem em logaritmos que, por não impor restrições prévias sobre elasticidades de substituição ou sobre o formato da curva de custo médio, permite que a existência e o sentido das economias de escala sejam determinados pelos dados. Essa flexibilidade é a principal razão de sua adoção generalizada na literatura bancária (Berger & Mester, 1997).

A adequação dessa forma funcional ao sistema financeiro brasileiro é atestada pelo próprio regulador. O Banco Central do Brasil (2023), ao avaliar a eficiência do Sistema Financeiro Nacional, emprega função translog com três insumos, captação, administração e capital, e cinco produtos, crédito, títulos e valores mobiliários, outros ativos usuais, receita de serviços e depósitos, estimada por fronteira estocástica sobre 232 instituições entre 2004 e 2023. A especificação adotada aqui dialoga com essa referência, o que permite comparabilidade com o parâmetro oficial, ainda que o objetivo seja distinto, pois o regulador mede a distância de cada instituição à fronteira eficiente e o presente estudo estima a elasticidade de escala.

2.3 Economias de escala em bancos

Economias de escala existem quando o custo total cresce menos que proporcionalmente ao volume de produção, de modo que o custo médio decresce com o tamanho. Formalmente, a medida é dada pela soma das elasticidades do custo em relação a cada produto, sendo que valores inferiores à unidade indicam economias de escala, iguais à unidade indicam retornos constantes e superiores indicam deseconomias.

Cabe destacar uma distinção frequentemente negligenciada. A proporção de custos fixos na estrutura de custos não constitui medida de economia de escala. Duas instituições podem apresentar participações idênticas de custos fixos e elasticidades de escala inteiramente distintas, uma vez que a primeira grandeza descreve a composição do custo em um dado nível de atividade e a segunda descreve como o custo reage a variações desse nível. A confusão entre ambas é fonte recorrente de

inferência equivocada, e sua separação constitui uma das preocupações metodológicas deste trabalho.

A evidência internacional evoluiu ao longo do tempo. Estudos das décadas de 1980 e 1990 identificavam economias de escala apenas em instituições de pequeno porte, com exaustão dos ganhos a partir de patamares modestos de ativo (Berger & Humphrey, 1997). Wheelock e Wilson (2012), empregando métodos não paramétricos, encontraram economias de escala mesmo entre os maiores bancos norte-americanos, e Hughes e Mester (2013) chegaram em conclusão semelhante ao considerar a atitude dos gestores diante do risco. Cerasi e Daltung (2000) oferecem contraponto teórico ao mostrar que os custos de monitoramento crescem com a diversificação e com o tamanho, o que estabelece limite econômico à expansão.

No caso brasileiro, a discussão possui relevância particular. O sistema apresenta elevado grau de concentração, e o argumento de ganhos de escala é recorrentemente invocado para justificar operações de fusão e aquisição. A verificação empírica dessa premissa permanece escassa, pois as análises disponíveis concentram-se em medidas de eficiência técnica, e não em elasticidades de escala, de modo que a pergunta sobre ganhos associados ao tamanho permanece em aberto.

2.4 Transformação digital e estrutura de custos bancários

O período examinado neste estudo abrange transformações relevantes na tecnologia de produção bancária no Brasil. A entrada de instituições de modelo digital, sem rede física de atendimento, e a implantação do Pix em novembro de 2020 alteraram simultaneamente a estrutura de receitas e a configuração de custos do setor.

Freitas e Kirch (2019) investigaram gastos e investimentos em tecnologia da informação em bancos brasileiros e identificaram relação positiva com indicadores de rentabilidade e negativa com despesas administrativas, sugerindo que tais dispêndios podem representar simultaneamente redução de custos e ampliação de resultado. A substituição de rede física por infraestrutura digital, contudo, não elimina custos fixos, ela os transfere de agências e pessoal para plataformas tecnológicas e capital humano especializado, com implicações ainda pouco documentadas sobre a rigidez da estrutura de custos e sobre a escala mínima eficiente.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Amostra e fonte dos dados

Os dados econômico-financeiros foram extraídos da base IF.data do Banco Central do Brasil, por meio da interface OData do serviço Olinda, abrangendo o período de dezembro de 2014 a março de 2026. Foram coletados os relatórios de Resumo, Ativo, Passivo, Demonstração de Resultado, Informações de Capital e Segmentação, nos níveis de conglomerado prudencial e de instituição individual, resultando em mais de doze milhões de registros contábeis.

A análise foi conduzida no nível de conglomerado prudencial, que evita a dupla contagem entre entidades pertencentes a um mesmo grupo econômico. A identificação das instituições utiliza o código atribuído pelo Banco Central e o CNPJ da instituição líder, e não a denominação social, uma vez que esta se altera ao longo do período analisado.

A base foi consolidada em frequência anual. Testes preliminares mostraram que a elasticidade do custo em relação à atividade é praticamente idêntica em frequência anual e trimestral, com diferença na terceira casa decimal, ao passo que o

poder explicativo intragrupo é sistematicamente maior no dado anual. A variação intra-ano é dominada por sazonalidade contábil, que constitui ruído para o objeto deste estudo.

O IF.data publica as contas de resultado acumuladas dentro de cada semestre, de modo que o valor do segundo trimestre corresponde ao acumulado de janeiro a junho e o do quarto trimestre ao acumulado de julho a dezembro. O fluxo anual é obtido, portanto, pela soma dos valores do segundo e do quarto trimestres. As contas patrimoniais são tomadas na posição de dezembro.

O campo de Tipo de Consolidado Bancário do cadastro do Banco Central foi utilizado para delimitar o universo de instituições bancárias. Permanecem na amostra os consolidados B1, que reúne bancos comerciais e bancos múltiplos com carteira comercial, B2, que reúne bancos múltiplos sem carteira comercial, bancos de investimento e de câmbio, e B4, que reúne bancos de desenvolvimento. Foram excluídos os consolidados N1, N2, N3 e N4, que compreendem sociedades de crédito, financiamento e investimento, arrendamento mercantil, corretoras e distribuidoras de títulos, administradoras de consórcio e instituições de pagamento, por não constituírem bancos.

As cooperativas de crédito dos consolidados B3S e B3C integram a amostra apenas quando o ativo mediano da instituição no período alcança um bilhão de reais, patamar equivalente ao primeiro quartil do ativo dos bancos. O critério preserva as cooperativas de porte comparável ao dos bancos e afasta as instituições de base estritamente local, cuja tecnologia de produção é distinta. Exigiu-se ainda um mínimo de três anos de observação por instituição. A amostra final reúne 245 instituições e 2.341 observações anuais entre 2015 e 2024, distribuídas em 103 bancos do consolidado B1, 38 do B2, 4 do B4, 77 cooperativas singulares e 23 cooperativas centrais.

A partir da data-base de março de 2025 entrou em vigor novo Padrão Contábil das Instituições Reguladas pelo Banco Central, instituído pela Resolução CMN nº 4.966/2021 e pela Resolução BCB nº 352/2023, com alteração dos códigos e da composição de diversas contas. A extensão da amostra ao exercício de 2025 foi testada e descartada por incompatibilidade contábil entre os dois regimes, de modo que a amostra se encerra em 2024, preservando a homogeneidade da série.

As variáveis contínuas foram winsorizadas nos percentis de 1% e 99%, procedimento que preserva a observação e limita apenas o valor extremo. Os resultados sem winsorização também são reportados. Todos os valores monetários foram deflacionados pelo Índice Nacional de Preços ao Consumidor Amplo, a preços de 2025.

3.2 Definição das variáveis

Adotamos a metodologia da intermediação (Sealey & Lindley, 1977). O custo total compreende as despesas de pessoal, as outras despesas administrativas, as outras despesas operacionais e as despesas de captação. Para a decomposição entre custos fixos e variáveis emprega-se o custo operacional, que exclui as despesas de captação, por serem estas, função direta do volume captado e da taxa de juros, e não da estrutura operacional da instituição.

O nível de atividade é medido pelo volume de operação, definido como a soma da carteira de crédito, dos depósitos e dos títulos e valores mobiliários. A escolha decorre de teste específico. A receita bancária é o produto de volume por taxa de juros, e a taxa Selic oscilou entre 14,25% e 2,00% no período, de modo que a variação anual da receita agregada do sistema apresenta desvio padrão de 24,2 pontos

percentuais, contra 6,5 pontos do crédito e 3,9 pontos do próprio custo operacional. Utilizar a receita como medida de atividade equivale a estimar a resposta do custo ao ciclo de juros, e não à variação do volume operado. A receita é mantida como especificação alternativa. Registre-se que o IF.data não divulga o número de funcionários por instituição, o que impede a definição exata do preço do trabalho, e a sensibilidade dos resultados a essa escolha é reportada na Seção 4. O Quadro 1 detalha a construção de cada variável.

Variável	Construção
Custo operacional (CO)	Despesas de pessoal + despesas administrativas + outras despesas operacionais
Custo total (CT)	CO + despesas de captação
Volume de operação (ATV)	Carteira de crédito + depósitos totais + títulos e valores mobiliários
Receita total (REC)	Receitas de intermediação financeira + rendas de serviços + rendas de tarifas
Produtos (y_1, y_2, y_3)	Carteira de crédito; títulos e valores mobiliários; rendas de serviços e tarifas
Preço do trabalho (w_1)	Despesas de pessoal ÷ ativo total
Preço do capital físico (w_2)	Despesas administrativas ÷ ativo permanente
Preço do funding (w_3)	Despesas de captação ÷ depósitos totais
Controles	Patrimônio líquido e tendência temporal, com termo quadrático
Indicadora de queda (D)	Assume 1 quando a variação do volume de operação é negativa
Intensidade de ativo e de mão de obra	Logaritmo do ativo permanente e das despesas de pessoal, ambos divididos pelo volume de operação

Quadro 1. Definição operacional das variáveis

Fonte: Elaborado pelos autores com base no IF.data do Banco Central do Brasil.

3.3 Modelo A: função custo translog

Estima-se a função custo translog (Christensen, Jorgenson & Lau, 1973) com três produtos e três preços de insumos, na forma apresentada na equação 1.

$$\ln(CT/w_3) = \alpha_0 + \Sigma_i \alpha_i \ln y_i + \tfrac{1}{2} \Sigma_i\Sigma_j \alpha_{ij} \ln y_i \ln y_j + \Sigma_k \beta_k \ln(w_k/w_3) + \tfrac{1}{2} \Sigma_k\Sigma_l \beta_{kl} \ln(w_k/w_3) \ln(w_l/w_3) + \Sigma_i\Sigma_k \gamma_{ik} \ln y_i \ln(w_k/w_3) + \delta_1 \ln PL + \delta_2 t + \delta_3 t^2 + \mu_i + \varepsilon_{it} \quad (1)$$

Nessa especificação, i indexa a instituição, t o exercício e μ_i o efeito individual não observado. A homogeneidade linear nos preços é imposta pela normalização do custo total e dos demais preços pelo preço do funding. A simetria é imposta por α_{ij} igual a α_{ji} e β_{kl} igual a β_{lk}. As variáveis de tendência capturam mudança tecnológica e o patrimônio líquido controla diferenças de estrutura de capital (Hughes & Mester, 2013).

Todas as variáveis foram normalizadas pela respectiva média geométrica, de modo que os coeficientes lineares correspondem às elasticidades avaliadas na média amostral. A medida de economia de escala é a soma das elasticidades do custo em

relação aos produtos. Valores inferiores à unidade indicam economias de escala, iguais à unidade indicam retornos constantes e superiores indicam deseconomias. A hipótese nula de retornos constantes é testada por Wald, com erro padrão obtido pelo método delta. A medida é calculada na média amostral e, separadamente, por quartil de ativo, por segmento prudencial e por natureza do controle acionário.

3.4 Modelo B: assimetria de custos

Aplica-se a especificação de Anderson, Banker e Janakiraman (2003), adaptada ao setor bancário, conforme a equação 2.

$$\Delta \ln CT_{it} = \beta_0 + \beta_1 \Delta \ln ATV_{it} + \beta_2 (D_{it} \times \Delta \ln ATV_{it}) + \mu_i + \varepsilon_{it} \quad (2)$$

ATV é o nível de atividade, Δln denota a variação logarítmica anual e D_{it} é variável binária que assume valor um quando a atividade se reduz no exercício. O coeficiente β_1 mede a resposta do custo a aumentos de atividade e a soma de β_1 com β_2 mede a resposta a reduções. Valor de β_2 negativo e significante indica custos assimétricos.

A especificação é estendida em quatro direções. A primeira incorpora as expectativas do gestor, por meio de interação com variável binária de queda também no exercício anterior (Banker, Byzalov & Plehn-Dujowich, 2014). A segunda testa os determinantes da assimetria, interagindo o termo assimétrico com a intensidade de ativo fixo e com a intensidade de mão de obra. A terceira examina a heterogeneidade por tipo de instituição e por natureza do controle. A quarta compara os exercícios anteriores e posteriores à implantação do Pix, em novembro de 2020.

Todas as estimações empregam efeitos fixos de instituição, com erros padrão robustos agrupados por instituição, o que acomoda heterocedasticidade e autocorrelação serial dentro de cada entidade.

4 ANÁLISE DOS RESULTADOS

4.1 Caracterização da amostra

A Tabela 1 apresenta as estatísticas descritivas das principais variáveis, em milhões de reais de 2025. A distribuição é fortemente assimétrica, característica esperada do setor. O ativo mediano é de R$ 3,5 bilhões, enquanto a média alcança R$ 55,4 bilhões, o que reflete a presença de instituições de porte muito superior ao restante da amostra.

Tabela 1

Estatísticas descritivas, em R$ milhões de 2025

Variável	Média	Desvio padrão	1º quartil	Mediana	3º quartil
Custo operacional	2.384	11.168	69	153	486
Volume de operação	52.193	252.051	1.490	3.830	10.798
Receita total	6.515	30.267	207	459	1.382
Ativo total	55.394	262.578	1.467	3.544	11.104

Fonte: Elaborado pelos autores.

A composição do custo operacional é equilibrada, com as despesas de pessoal respondendo por 39,1% do total na mediana, as administrativas por 36,3% e as outras despesas operacionais por 22,2%. O índice de eficiência, medido pela razão entre custo operacional e receita total, apresenta mediana de 0,371 no período, com deterioração até 2020, quando alcança 0,471, e recuperação consistente a partir de então, chegando a 0,299 em 2024. O movimento acompanha o ciclo de juros, o que reforça a advertência quanto ao uso da receita como medida de atividade.

4.2 Estrutura de custos fixos e variáveis

A Tabela 2 apresenta a elasticidade do custo operacional em relação ao volume de operação, obtida por efeitos fixos. Em um modelo linear de custo, a elasticidade do custo em relação à quantidade corresponde à parcela variável, de modo que seu complemento estima a parcela fixa.

Tabela 2

Decomposição da estrutura de custos

Especificação	Elasticidade	Parcela fixa	Instituições	Observações
Bancos, sem cooperativas	0,417	58,3%	145	1.344
Bancos e cooperativas de grande porte	0,448	55,2%	245	2.341
Sem winsorização	0,400	60,0%	245	2.341
Mínimo de cinco anos por instituição	0,453	54,7%	239	2.321

Fonte: Elaborado pelos autores.

A parcela fixa situa-se entre 55% e 60% conforme a escolha metodológica, com estabilidade confortável entre as especificações. O resultado é substancialmente inferior às estimativas que circulam na literatura nacional para o setor, e a diferença tem origem identificável. Estimativas baseadas na receita como medida de atividade incorporam o ciclo de juros no regressor, o que configura erro de medida clássico e produz viés de atenuação, empurrando a elasticidade na direção de zero e, por consequência, elevando artificialmente a parcela fixa estimada.

4.3 A rigidez crescente da estrutura de custos

A Tabela 3 apresenta a elasticidade estimada em subperíodos, e o resultado é o achado mais expressivo deste trabalho.

Tabela 3

Evolução da estrutura de custos ao longo do período

Período	Elasticidade	Estatística t	Parcela fixa
2015 a 2017	0,320	5,0	68,0%
2018 a 2020	0,250	6,5	75,0%
2021 a 2022	0,294	3,3	70,6%
2023 a 2024	0,123	1,2	87,7%

Fonte: Elaborado pelos autores.

A parcela fixa dos custos bancários eleva-se de 68,0% no triênio inicial para 87,7% no biênio final. No último período a elasticidade perde significância estatística, o que significa que o custo operacional dos bancos brasileiros deixou de responder de forma detectável a variações no volume operado.

A interpretação econômica é consistente com a transformação tecnológica do setor, pois a substituição de rede física e de pessoal por plataformas digitais não elimina custos, ela os converte de variáveis em fixos, uma vez que a infraestrutura tecnológica precisa ser mantida independentemente do volume processado. O achado dialoga com Freitas e Kirch (2019), que identificaram relação negativa entre investimento em tecnologia da informação e despesas administrativas, e acrescenta que a redução de despesas veio acompanhada de aumento da rigidez. A implicação gerencial é relevante, pois estruturas mais rígidas ampliam a alavancagem operacional e, portanto, a exposição do resultado a variações na receita (Horngren, Sundem & Stratton, 2004).

4.4 Economias de escala

A Tabela 4 apresenta as elasticidades do custo total em relação a cada produto, estimadas pela função translog. O ajuste é elevado, com coeficiente de determinação intragrupo de 0,962.

Tabela 4

Função custo translog, elasticidades na média amostral

Produto	Elasticidade	Valor p
Carteira de crédito	0,451	0,000
Títulos e valores mobiliários	0,198	0,000
Receitas de serviços	0,101	0,004
Medida de escala, soma	**0,749**	

Fonte: Elaborado pelos autores. Erro padrão da soma igual a 0,057.

A soma das elasticidades é 0,749, com erro padrão de 0,057. A hipótese de retornos constantes é rejeitada com estatística t de menos 4,43 e valor p inferior a 0,0001. Existem, portanto, economias de escala na produção bancária brasileira, e um aumento de 10% na produção eleva o custo total em 7,5%.

A conclusão é robusta à definição do preço do trabalho, que constitui a principal limitação de dados deste estudo. Adotando o ativo total como denominador, a medida de escala é 0,749. Adotando as despesas administrativas, é 0,586. Adotando o ativo permanente, é 0,423. Em todas as especificações a hipótese de retornos constantes é rejeitada com valor p inferior a 0,0001.

A Tabela 5 apresenta a medida de escala calculada por segmento prudencial, classificação estabelecida pela Resolução CMN nº 4.553/2017 e aplicada pelo próprio regulador. A Figura 1 apresenta os mesmos resultados em forma gráfica.

Tabela 5

Economias de escala por segmento prudencial

Segmento	Ativo mediano	Medida de escala	Interpretação	Observações
S1	R$ 1.285 bilhões	1,248	Deseconomias	60
S2	R$ 119 bilhões	1,074	Deseconomias	60

S3	R$ 19 bilhões	0,861	Economias moderadas	418
S4	R$ 2,3 bilhões	0,664	Economias fortes	1.066
S5	R$ 2,0 bilhões	0,719	Economias fortes	146

Fonte: Elaborado pelos autores. Os segmentos S1 e S2 reúnem número reduzido de observações, o que recomenda cautela na leitura das estimativas desses grupos.

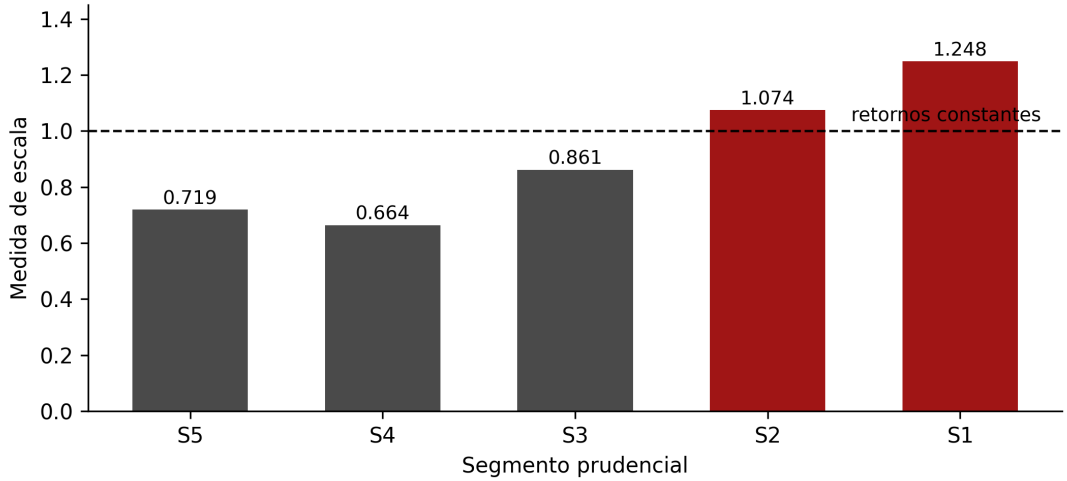


Figura 1. Medida de escala por segmento prudencial
Fonte: Elaborado pelos autores.

O padrão é evidente e responde diretamente à questão que motivou esta pesquisa. As economias de escala são fortes entre as instituições de menor porte, dissipam-se com o aumento do tamanho e convertem-se em deseconomias nos dois segmentos superiores. Nos bancos do segmento S1, que reúne as instituições sistemicamente importantes, um aumento de 10% na produção eleva o custo total em 12,5%. O resultado por quartil de ativo confirma a trajetória, com medida de escala de 0,556 no quartil inferior, 0,696 no segundo, 0,787 no terceiro e 0,959 no superior. Por natureza do controle, as instituições públicas apresentam medida de 0,988, próxima de retornos constantes, contra 0,712 nas privadas nacionais e 0,772 nas estrangeiras, diferença compatível com o porte médio superior das públicas.

Estes achados convergem com a literatura consolidada por Berger e Humphrey (1997), que identificava exaustão dos ganhos de escala a partir de patamares modestos de ativo, e divergem de Wheelock e Wilson (2012), que encontraram economias persistentes mesmo entre os maiores bancos norte-americanos. A discordância admite três explicações não excludentes, a maior concentração do mercado brasileiro, que reduz a pressão competitiva sobre os maiores, a diferença metodológica, uma vez que aqueles autores empregam métodos não paramétricos e incluem o risco na especificação, e o custo de complexidade organizacional previsto por Cerasi e Daltung (2000), segundo os quais os custos de monitoramento crescem com o tamanho e com a diversificação.

4.5 Assimetria de custos

A Tabela 6 apresenta a estimação do modelo de custos assimétricos com as duas medidas de atividade.

Tabela 6

Comportamento assimétrico dos custos

Medida de atividade	β_1, aumento	β_2, assimetria	$\beta_1 + \beta_2$, queda	Quedas na amostra
Volume de operação	0,227 (p = 0,000)	−0,062 (p = 0,406)	0,165	20,8%
Receita total	0,122 (p = 0,004)	−0,078 (p = 0,051)	0,043	32,0%

Fonte: Elaborado pelos autores.

O resultado depende da medida de atividade adotada, e essa dependência é ela própria informativa. Com o volume de operação, o termo assimétrico apresenta o sinal previsto pela teoria, porém não alcança significância estatística. Com a receita total, que é a medida usual na literatura, o termo assimétrico é significante ao nível de 10%, indicando que os custos caem menos diante de reduções de receita do que sobem diante de elevações equivalentes.

A divergência tem explicação direta. A receita registra reduções em 32,0% das observações, contra 20,8% do volume de operação, precisamente porque incorpora o ciclo de juros. O modelo identifica o parâmetro de interesse a partir dos episódios de queda, e a medida que produz mais quedas oferece mais poder estatístico. Resta em aberto, contudo, se as quedas adicionais capturadas pela receita correspondem a decisões gerenciais de ajuste de capacidade, que é o fenômeno de interesse, ou apenas à variação de taxa de juros, que não demanda ajuste de estrutura.

Entre as extensões, apenas a que incorpora as expectativas do gestor produz resultado digno de nota. Ao controlar por queda também no exercício anterior, o termo assimétrico passa a menos 0,124, com valor p de 0,077, e o termo de queda sucessiva torna-se positivo, em mais 0,158. O padrão é o previsto por Banker, Byzalov e Plehn-Dujowich (2014), segundo os quais reduções repetidas levam o gestor a revisar expectativas e a cortar recursos com maior intensidade, o que atenua a assimetria.

As demais extensões não produzem resultados significantes. Nem a intensidade de ativo fixo nem a de mão de obra explicam a magnitude da assimetria, a heterogeneidade por tipo de instituição e por natureza do controle não é estatisticamente detectável, e a comparação entre os exercícios anteriores e posteriores à implantação do Pix não revela alteração.

A ausência de assimetria estatisticamente robusta merece interpretação, e não deve ser tratada como falha. Duas razões concorrem para o resultado. O volume de operação de um banco raramente se reduz, o que limita os episódios que identificam o parâmetro. E, de forma mais substantiva, uma estrutura em que a parcela fixa alcança 88% deixa margem estreita de custo ajustável, de modo que custos que não variam com a atividade tampouco podem variar assimetricamente com ela.

5 CONCLUSÃO

Esta pesquisa examinou a estrutura de custos do setor bancário brasileiro, com dois objetivos. O primeiro foi estimar a decomposição entre custos fixos e variáveis e acompanhar sua evolução. O segundo foi verificar a existência de economias de escala na produção bancária e identificar a que instituições elas beneficiam.

Foram utilizados dados anuais do IF.data do Banco Central do Brasil, no nível de conglomerado prudencial, entre 2015 e 2024, com 245 instituições e 2.341 observações, compreendendo bancos comerciais, múltiplos, de investimento, de câmbio e de desenvolvimento, além de cooperativas de crédito de porte comparável ao dos bancos. Estimou-se função custo translog na abordagem de intermediação e

o modelo de custos assimétricos de Anderson, Banker e Janakiraman (2003), ambos por efeitos fixos.

Três achados se destacam. O primeiro é que a parcela fixa dos custos bancários situa-se entre 55% e 60% no conjunto do período, resultado estável entre as especificações testadas. O segundo, e mais expressivo, é que essa parcela se eleva de 68,0% no triênio de 2015 a 2017 para 87,7% no biênio de 2023 e 2024, com perda de significância estatística da elasticidade no período final. A estrutura de custos do setor tornou-se substancialmente mais rígida ao longo da década, movimento compatível com a substituição de rede física e pessoal por infraestrutura tecnológica, que converte custos variáveis em fixos.

O terceiro achado responde à questão que motivou a pesquisa. Existem economias de escala na produção bancária brasileira, com medida de escala de 0,749 e rejeição da hipótese de retornos constantes ao nível de significância de 1%. Elas se concentram, contudo, nas instituições de menor porte. A medida de escala é de 0,664 no segmento prudencial S4 e de 0,861 no S3, e converte-se em deseconomias nos dois segmentos superiores, alcançando 1,074 no S2 e 1,248 no S1. Grandes bancos não possuem menores custos por unidade produzida. O argumento de ganho de escala, recorrentemente invocado para justificar operações de fusão e aquisição entre instituições de grande porte, não encontra sustentação empírica nos dados examinados.

Duas questões permaneceram sem resposta satisfatória. A primeira é o comportamento assimétrico dos custos, cujo termo apresenta o sinal previsto pela teoria em todas as especificações, mas alcança significância apenas com a receita como medida de atividade, e ao nível de 10%. Não é possível determinar se essa assimetria reflete decisões gerenciais de ajuste de capacidade ou apenas a variação da taxa de juros incorporada àquela medida. A segunda é a causa da rigidez crescente, pois a trajetória documentada é compatível com a transformação digital do setor, mas o desenho adotado não permite atribuir causalidade, e o teste em torno da implantação do Pix não produziu resultado significante.

O trabalho apresenta limitações que devem ser consideradas. A base do Banco Central não divulga o número de funcionários por instituição, o que impediu a definição exata do preço do trabalho e exigiu o uso do ativo total como denominador. A conclusão sobre economias de escala mostrou-se robusta a três definições alternativas, mas a magnitude da medida varia entre elas. Os segmentos prudenciais S1 e S2 reúnem número reduzido de observações, o que recomenda cautela na leitura das estimativas desses grupos. A amostra foi encerrada em 2024 por incompatibilidade contábil decorrente da entrada em vigor de novo padrão em 2025.

Para pesquisas futuras, sugere-se quatro caminhos. O primeiro é a incorporação de dados de rede de atendimento por instituição, disponíveis na Estatística Bancária Mensal por município a partir de 2019, que permitiriam separar o efeito da capilaridade física do efeito da escala financeira. O segundo é a estimação de fronteira estocástica, que complementaria a medida de escala com escores de eficiência e permitiria examinar se as instituições que operam sob deseconomias de escala compensam essa condição com maior eficiência técnica. O terceiro é a investigação da rigidez crescente por meio de desenho quase experimental que explore a entrada de bancos digitais como choque exógeno. O quarto é a extensão da série sob o novo padrão contábil, quando houver observações suficientes, com reconciliação explícita entre os dois regimes.

REFERÊNCIAS

Anderson, M. C., Banker, R. D., & Janakiraman, S. N. (2003). Are selling, general, and administrative costs “sticky”? *Journal of Accounting Research, 41*(1), 47-63.

Banco Central do Brasil. (2017, 2021, 2023). *Resolução CMN nº 4.553/2017; Resolução CMN nº 4.966/2021; Resolução BCB nº 352/2023*. Brasília: BCB.

Banco Central do Brasil. (2023). Evolução da eficiência do Sistema Financeiro Nacional sob o ponto de vista da otimização de recursos. *Relatório de Economia Bancária 2023*, Boxe 4, 66-71.

Banker, R. D., & Byzalov, D. (2014). Asymmetric cost behavior. *Journal of Management Accounting Research, 26*(2), 43-79.

Banker, R. D., Byzalov, D., & Chen, L. T. (2013). Employment protection legislation, adjustment costs and cross-country differences in cost behavior. *Journal of Accounting and Economics, 55*(1), 111-127.

Banker, R. D., Byzalov, D., & Plehn-Dujowich, J. M. (2014). Demand uncertainty and cost behavior. *The Accounting Review, 89*(3), 839-865.

Berger, A. N., & Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research, 98*(2), 175-212.

Berger, A. N., & Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking & Finance, 21*(7), 895-947.

Bubeck, S. K., & Silva, A. da. (2025). Sticky costs e anti-sticky costs: uma análise pela perspectiva macro accounting de ciclos econômicos. *Revista Catarinense da Ciência Contábil, 24*, e3607.

Cerasi, V., & Daltung, S. (2000). The optimal size of a bank: Costs and benefits of diversification. *European Economic Review, 44*(9), 1701-1726.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1973). Transcendental logarithmic production frontiers. *The Review of Economics and Statistics, 55*(1), 28-45.

Costa, P. S., Marques, A. V. C., Santos, C. K. S., & Lima, F. D. C. (2013). Análise do comportamento assimétrico dos custos nas companhias abertas dos países da América Latina. *Anais do Congresso Brasileiro de Custos*, 20.

Freitas, O. D., & Kirch, G. (2019). Performance dos bancos brasileiros no contexto de digitalização. *Revista Brasileira de Finanças, 17*(2), 38-55.

Guerra, A. R., Rocha, W., & Corrar, L. J. (2007). Análise do impacto das variações de receitas nos lucros das empresas com diferentes estruturas de custos. *Revista de Administração da USP, 42*(2), 227-238.

Horngren, C. T., Sundem, G. L., & Stratton, W. O. (2004). *Contabilidade gerencial* (12a ed.). São Paulo: Prentice Hall.

Hughes, J. P., & Mester, L. J. (2013). Who said large banks don’t experience scale economies? Evidence from a risk-return-driven cost function. *Journal of Financial Intermediation, 22*(4), 559-585.

Ibrahim, A. E. A., Ali, H. M. H. O., & Aboelkheir, H. N. E. R. (2022). Cost stickiness: A systematic literature review of 27 years of research and a future research agenda. *Journal of International Accounting, Auditing and Taxation, 46*, 100439.

Martins, E. (2010). *Contabilidade de custos* (10a ed.). São Paulo: Atlas.

Martins, E., & Rocha, W. (2010). *Métodos de custeio comparados*. São Paulo: Atlas.

Medeiros, O. R., Costa, P. S., & Silva, C. A. T. (2005). Testes empíricos sobre o comportamento assimétrico dos custos nas empresas brasileiras. *Revista Contabilidade & Finanças, 16*(38), 47-56.

Richartz, F., & Borgert, A. (2014). O comportamento dos custos das empresas brasileiras listadas na BM&FBovespa entre 1994 e 2011. *Contaduría y Administración, 59*(4), 39-70.

Richartz, F., & Borgert, A. (2021). Fatores explicativos para o comportamento assimétrico dos custos das empresas listadas na B3. *Revista Universo Contábil, 16*(3).

Richartz, F., Borgert, A., & Lunkes, R. J. (2014). Comportamento assimétrico dos custos nas empresas brasileiras listadas na BM&FBovespa. *Advances in Scientific and Applied Accounting, 7*(3), 339-361.

Richartz, F., & Ensslin, S. R. (2014). Comportamento dos custos: análise da produção científica. *Revista Contemporânea de Contabilidade, 11*(22), 3-24.

Sealey, C. W., & Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance, 32*(4), 1251-1266.

Wheelock, D. C., & Wilson, P. W. (2012). Do large banks have lower costs? New estimates of returns to scale for U.S. banks. *Journal of Money, Credit and Banking, 44*(1), 171-199.