



UnB

UNIVERSIDADE DE BRASÍLIA

**FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E GESTÃO
DE POLÍTICAS PÚBLICAS**

DEPARTAMENTO DE ECONOMIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

**EFICIÊNCIA, PRIVATIZAÇÃO E TARIFAS:
ANÁLISE SOBRE O SETOR DE SANEAMENTO BRASILEIRO**

PEDRO RICELLY GAMA DE OLIVEIRA

Brasília
2026

Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas
Departamento de Economia

**Eficiência, privatização e tarifas:
análise sobre o setor de saneamento brasileiro**

Pedro Ricelly Gama de Oliveira

Tese de Doutorado submetida ao Programa de
Pós-Graduação em Economia da Universidade
de Brasília como parte dos requisitos necessários
para obtenção do grau de Doutor.

Orientador: Prof. Dr. Marcelo de Oliveira Torres

Brasília
2026

FICHA CATALOGRÁFICA

Oliveira, Pedro Ricelly Gama de.

 Eficiência, privatização e tarifas: análise sobre o setor de saneamento brasileiro / Pedro Ricelly Gama de Oliveira; orientador Marcelo de Oliveira Torres.

-- Brasília, 2026.

 263 p.

 Tese de Doutorado (Programa de Pós-Graduação em Economia) --
Universidade de Brasília, 2026.

 1. Saneamento básico. 2. Privatização. 3. Eficiência. 4. Regulação tarifária. 5. Diferenças em diferenças. I. Torres, Marcelo de Oliveira, orient. II. Título.

Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas
Departamento de Economia

Eficiência, privatização e tarifas: análise sobre o setor de saneamento brasileiro

Pedro Ricelly Gama de Oliveira

Tese de Doutorado submetida ao Programa de Pós-Graduação em Economia da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Doutor.

Trabalho aprovado. Brasília, 27 de maio de 2026:

Prof. Dr. Marcelo de Oliveira Torres,
UnB/FACE
Orientador

Profa. Dra. Ana Carolina Pereira Zoghbi,
UnB/FACE
Examinadora interna

Prof. Dr. Lucas Vitor de Carvalho Sousa,
UnB/FACE
Examinador interno

Profa. Dra. Adriana Pereira de Sousa,
Universidade Estadual de Goiás
Examinador externo

Prof. Dr. Carlos Cesar Santejo Saiani,
Universidade Federal de Uberlândia
Examinador externo

*Este trabalho é dedicado aos meus antepassados que,
trazidos a este país em navios como cargas,
só puderam sonhar com liberdade e bem-viver.
Hoje eu vivo o que vocês sonharam.*

Agradecimentos

Agradeço ao meu orientador por me permitir trabalhar com autonomia, mas sobretudo por me fazer boas perguntas. O que é a ciência senão uma reunião de boas perguntas?

Agradeço à minha família por serem meu porto seguro: minha mãe Vanuza, meu irmão Fabio e meu pai Edilson. Agradeço também ao meu cachorro Bento, que me fez companhia e me levou para caminhar durante todos os anos da pós-graduação, lembrando que havia vida além da tela do computador.

Agradeço de forma especial ao meu marido Ricardo, que acredita no meu potencial até quando eu mesmo duvido, e cujo amor tornou essa jornada não apenas possível, mas feliz.

Agradeço à Universidade de Brasília pela oportunidade de cursar o doutorado e à CAPES pelo apoio financeiro que o tornou viável.



Resumo

Esta tese investiga os efeitos da participação privada no saneamento básico brasileiro a partir de quatro perspectivas complementares: conceitual, bibliométrica e duas empíricas. A hipótese central é a de que a propriedade, por si só, não determina a eficiência dos prestadores nem garante a modicidade tarifária, o que prevalece é o ambiente regulatório e o desenho dos contratos de concessão. O primeiro artigo caracteriza o mercado de água e saneamento como monopólio natural e revisa experiências internacionais de privatização, evidenciando que os resultados dependem criticamente da qualidade da regulação. O segundo artigo realiza uma revisão bibliométrica de 392 documentos publicados entre 1972 e 2025 nas bases Scopus e Web of Science, mapeando a agenda de pesquisa para o Brasil e identificando sete eixos prioritários de investigação. O terceiro artigo estima uma função de custo variável translogarítmica pelo método de Análise de Fronteira Estocástica (SFA) para 254 prestadores brasileiros no período 2015–2022, encontrando eficiência média de 75,4% e ausência de efeito estatisticamente significativo da propriedade privada sobre os custos. O quarto artigo avalia os efeitos causais das concessões plenas de abastecimento de água celebradas a partir do Novo Marco Legal do Saneamento Básico (NMLSB) sobre tarifas, cobertura, eficiência operacional, qualidade do serviço e morbidade hospitalar, utilizando o estimador de diferenças em diferenças escalonado combinado com pareamento por escore de propensão em um painel de municípios para o período 2017–2024. Os resultados indicam que as concessões produziram ganhos operacionais mensuráveis no curto prazo, como redução de despesas com pessoal e expansão da macromedicação, mas não se traduziram em melhora dos indicadores de qualidade do serviço nem em efeitos detectáveis sobre saúde pública. Os efeitos tarifários são heterogêneos entre coortes e dependem do modelo de leilão adotado: o critério de maior outorga está associado a aumentos tarifários, enquanto o de menor tarifa está associado a reduções. Em conjunto, os achados sugerem que o sucesso das concessões depende do desenho do processo licitatório, da completude e monitorabilidade dos contratos e da capacidade técnica das entidades reguladoras.

Palavras-chave: saneamento básico; privatização; eficiência; regulação tarifária; diferenças em diferenças.

Abstract

This dissertation investigates the effects of private sector participation in Brazilian water and sanitation services from four complementary perspectives: conceptual, bibliometric, and two empirical. The central hypothesis is that ownership, by itself, does not determine provider efficiency or guarantee tariff affordability, what prevails is the regulatory environment and the design of concession contracts. The first essay characterizes the water and sanitation market as a natural monopoly and reviews international privatization experiences, showing that outcomes depend critically on regulatory quality. The second essay conducts a bibliometric review of 392 documents published between 1972 and 2025 in the Scopus and Web of Science databases, mapping the research agenda for Brazil and identifying seven priority axes of investigation. The third essay estimates a translogarithmic variable cost function using Stochastic Frontier Analysis (SFA) for 254 Brazilian providers over the period 2015–2022, finding an average efficiency score of 75.4% and no statistically significant effect of private ownership on costs. The fourth essay evaluates the causal effects of full water supply concessions signed under Brazil's New Sanitation Legal Framework on tariffs, coverage, operational efficiency, service quality, and hospital morbidity, using the staggered difference-in-differences combined with propensity score matching in a panel of municipalities over the period 2017–2024. The results indicate that concessions produced measurable short-term operational gains, such as reductions in personnel expenditures and expansion of bulk metering, but did not translate into improvements in service quality indicators or detectable effects on public health. Tariff effects are heterogeneous across cohorts and depend on the auction design: the highest-bonus criterion is associated with tariff increases, while the lowest-tariff criterion is associated with reductions. Taken together, the findings suggest that the success of concessions depends on the design of the bidding process, the completeness and enforceability of contracts, and the technical capacity of regulatory agencies.

Keywords: water and sanitation; privatization; efficiency; tariff regulation; difference-in-differences.

Lista de figuras

Figura 1.1	Perda de Peso Morto no Monopólio (Adaptado de Tirole (1988)).	26
Figura 1.2	Distribuição municipal das concessões privadas de saneamento básico no Brasil, segundo o período de concessão (anterior ou posterior ao NMLSB). Fonte: elaboração própria.	42
Figura 2.1	Fluxograma do processo de revisão bibliométrica utilizando as bases de dados Scopus e Web of Science. Fonte: elaboração própria.	51
Figura 2.2	Evolução temporal da produção científica sobre participação privada no saneamento, 1972–2025. A linha tracejada vermelha indica a tendência de crescimento. Fonte: elaboração própria.	52
Figura 2.3	Top 10 autores mais produtivos na literatura sobre participação privada no saneamento. Fonte: elaboração própria.	53
Figura 2.4	Top 10 artigos mais citados na literatura sobre participação privada no saneamento. Fonte: elaboração própria.	54
Figura 2.5	Top 10 referências mais citadas pelos artigos da amostra. Fonte: elaboração própria.	54
Figura 2.6	Top 10 países mais produtivos na literatura sobre participação privada no saneamento, por país do autor correspondente. Fonte: elaboração própria.	55
Figura 2.7	Número de citações e média de citações por artigo por país. Fonte: elaboração própria.	56
Figura 2.8	Top 10 afiliações mais produtivas na literatura sobre participação privada no saneamento. Fonte: elaboração própria.	56
Figura 2.9	Top 30 palavras-chave mais frequentes na literatura sobre participação privada no saneamento. Fonte: elaboração própria.	57
Figura 2.10	Rede de co-autoria entre os autores da literatura sobre participação privada no saneamento. O tamanho dos nós é proporcional ao número de publicações do autor. Fonte: elaboração própria.	58
Figura 2.11	Rede de co-citação das referências mais frequentes na literatura sobre participação privada no saneamento. O tamanho dos nós é proporcional ao número de co-citações. Fonte: elaboração própria.	59
Figura 2.12	Evolução temática dos resumos de 1972 a 2025, dividida em cinco períodos. A espessura dos fluxos é proporcional à intensidade da associação entre os temas. Fonte: elaboração própria.	60
Figura 2.13	Tópicos de tendência na literatura sobre participação privada no saneamento. O tamanho dos pontos é proporcional à frequência do termo. Fonte: elaboração própria.	62

Figura 2.14	Mapa temático do período 2023–2025 da literatura sobre participação privada no saneamento. Os quadrantes indicam: temas motores (alta centralidade e alta densidade), temas de nicho (baixa centralidade, alta densidade), temas emergentes ou em declínio (baixa centralidade e densidade) e temas básicos ou transversais (alta centralidade, baixa densidade). Fonte: elaboração própria.	63
Figura 3.1	Distribuição geográfica da amostra: prestadores por região e população atendida (2015–2022).	89
Figura 3.2	Evolução da eficiência técnica média por tipo de gestão – Modelo Translog Tempo-Variante (2015-2022).	95
Figura 3.3	Evolução da eficiência técnica média por tipo de gestão — Amostra de robustez 2017–2022 (308 prestadores)	99
Figura 3.4	Evolução da eficiência técnica média por tipo de gestão — Amostra de robustez 2018–2022 (339 prestadores)	100
Figura 4.1	Número anual de concessões municipais de serviços de saneamento. Brasil, 1994–2024.	108
Figura 4.2	Tendência do índice IFA2008 (incidência da despesa de pessoal próprio nas despesas de exploração de água) para municípios tratados e controles. Brasil, 2017–2024.	108
Figura 4.3	Amostra pareada de municípios tratados e controles. Brasil, 2026. Fonte: Resultados da pesquisa.	120
Figura 4.4	Estudo de eventos da receita operacional direta média de usuários de água (IFA1001). Brasil, 2026. Fonte: Resultados da pesquisa.	123
Figura 4.5	Estudo de eventos do índice de atendimento da população total com rede de abastecimento de água (IAG0001). Brasil, 2026. Fonte: Resultados da pesquisa.	124
Figura 4.6	Estudo de eventos da incidência da despesa de pessoal próprio nas despesas de exploração de água (IFA2008). Brasil, 2026. Fonte: Resultados da pesquisa.	126

Lista de quadros

Quadro 1.1	Síntese de experiências internacionais de privatização do saneamento. . . .	37
Quadro 2.1	Artigos representativos do quinto período (2023–2025) identificados pelo mapa temático. Brasília, 2026.	65
Quadro 2.2	Agenda de pesquisa para o Brasil sobre participação privada no saneamento. Brasília, 2026.	69
Quadro 4.1	Variáveis utilizadas no PSM — Estratégia Flexível (2017). Brasil, 2026. . .	118

Lista de tabelas

Tabela 1.1	Informações globais do sistema de saneamento no Brasil em 2024 (%) . . .	43
Tabela 1.2	Prestadores de serviços segundo a abrangência.	44
Tabela 1.3	Natureza jurídica dos prestadores de serviços.	44
Tabela 1.4	Índices de atendimento de água e esgoto no Brasil.	45
Tabela 1.5	Perdas de água em sistemas de distribuição (%).	46
Tabela 1.6	Receitas e despesas médias praticadas (R\$/m ³).	46
Tabela 3.1	Síntese dos principais estudos internacionais sobre eficiência no setor de saneamento.	78
Tabela 3.2	Síntese dos principais estudos sobre eficiência no setor de saneamento brasileiro.	81
Tabela 3.3	Descrição das variáveis e correspondência com o SNIS.	87
Tabela 3.4	Estatísticas Descritivas das Variáveis por Tipo de Prestador (Valores Originais)	89
Tabela 3.5	Estatísticas Descritivas das Variáveis por Tipo de Prestador (Transformação Logarítmica)	90
Tabela 3.6	Distribuição das Variáveis Categóricas	91
Tabela 3.7	Resultados da Análise de Fronteira Estocástica Modelo Translog Tempo-Variante.	92
Tabela 3.8	Estatísticas de Eficiência Técnica por Modelo.	94
Tabela 3.9	Eficiência Técnica por Região — Modelo Translog Tempo-Variante.	94
Tabela 3.10	Evolução da Eficiência Técnica ao Longo do Tempo — Modelo Translog Tempo-Variante.	95
Tabela 3.11	Os cinco prestadores com maior eficiência técnica média (2015–2022).	96
Tabela 3.12	Os cinco prestadores com menor eficiência técnica média (2015–2022).	96
Tabela 3.13	Testes de Robustez: Modelo Translog Tempo-Variante com Diferentes Amostras	98
Tabela 4.1	Comparação das 3 estratégias de seleção de covariadas para PSM.	117
Tabela 4.2	Diferença de médias entre grupos tratamento e controle para amostras sem pareamento e com pareamento (2017). Brasil, 2026.	119
Tabela 4.3	Estatísticas descritivas dos indicadores da amostra 2017-2024. Brasil, 2026.	121
Tabela 4.4	Efeito médio e por coortes da concessão plena sobre receita operacional direta média de usuários de água (IFA1001). Brasil, 2026.	122
Tabela 4.5	Efeito médio e por coortes da concessão plena sobre o índice de atendimento da população total com rede de abastecimento de água (IAG0001). Brasil, 2026.	124
Tabela 4.6	Efeito médio e por coortes da concessão plena sobre a incidência da despesa de pessoal próprio nas despesas de exploração de água (IFA2008). Brasil, 2026.	125
Tabela 4.7	ATT Agregado — Grupo IAG. Estimador Callaway & Sant’Anna (2021). Brasil, 2017–2024.	127

Tabela 4.8	ATT Agregado — Grupo IFA. Estimador Callaway & Sant’Anna (2021). Brasil, 2017–2024.	128
Tabela 4.9	Efeito da concessão plena sobre morbidade hospitalar por doenças relacionadas ao saneamento. Estimador Callaway & Sant’Anna (2021). Brasil, 2017–2024.	131
Tabela 4.10	Análise de robustez — ATT médio por tamanho de amostra PSM. Estimador Callaway & Sant’Anna (2021). Amostra PSM-FLEXÍVEL com proporções de pareamento 1:1, 1:2 e 1:5. Brasil, 2017–2024.	133
Tabela 4.11	Análise de robustez — Efeito sobre morbidade hospitalar por tamanho de amostra PSM. Callaway & Sant’Anna (2021). Brasil, 2017–2024.	134
Tabela A.1	Harmonização dos Indicadores de Saneamento: Equivalências entre SINISA (2023-2024) e SNIS (2017-2022). Brasil, 2025.	170
Tabela A.2	Validação da Metodologia: Correlação e RMSE entre Indicadores Calculados e SNIS Original (municípios com prestador único). Brasil, 2025.	172

Lista de abreviaturas e siglas

ABCON SINDCON	Associação e Sindicato Nacional das Concessionárias Privadas de Serviços Públicos de Água e Esgoto
ANA	Agência Nacional de Águas e Saneamento Básico
ATT	Average Treatment Effect on the Treated
CADE	Conselho Administrativo de Defesa Econômica
DEA	Data Envelopment Analysis
DiD	Difference-in-Differences
DR	Doubly Robust
EUA	Estados Unidos da América
IBGE	Instituto Brasileiro de Geografia e Estatística
IPCA	Índice Nacional de Preços ao Consumidor Amplo
NMLSB	Novo Marco Legal do Saneamento Básico
ODS	Objetivos de Desenvolvimento Sustentável
OFWAT	Office of Water Services
OLS	Ordinary Least Squares
ONU	Organização das Nações Unidas
PIB	Produto Interno Bruto
PPP	Parcerias Público-Privadas
SFA	Stochastic Frontier Analysis
SICONFI	Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro
SINISA	Sistema Nacional de Informações em Saneamento Básico
SMD	Standardized Mean Difference
SNIS	Sistema Nacional de Informações sobre Saneamento

Sumário

Introdução	18
1 Economia do Saneamento: teoria, experiências internacionais e panorama brasileiro	22
1.1 Introdução	22
1.2 Monopólio natural	23
1.2.1 Comportamento de precificação do monopolista	25
1.3 Tarifas	27
1.4 Presença do Estado	28
1.5 Privatização	29
1.5.1 Reino Unido	29
1.5.2 Chile	31
1.5.3 França	32
1.5.4 Alemanha	33
1.5.5 Argentina	34
1.5.6 Bolívia	35
1.5.7 Estados Unidos	36
1.5.8 Síntese de experiências internacionais	37
1.6 Panorama do saneamento básico no Brasil	38
1.6.1 O Novo Marco Legal do Saneamento Básico	39
1.6.2 Características nacionais do saneamento	42
1.6.3 Características regionais do saneamento	44
1.7 Conclusões	46
2 Participação de empresas privadas no saneamento: revisão bibliométrica e agenda de pesquisa para o Brasil	48
2.1 Introdução	48
2.2 Materiais e Métodos	49
2.2.1 Fontes bibliográficas e estratégia de busca	49
2.2.2 Limpeza dos dados	50
2.2.3 Análise bibliométrica e visualização	51
2.3 Resultados	52
2.3.1 Análise descritiva	52
2.3.2 Redes de Colaboração e Co-citação	57
2.3.3 Evolução Temática	59
2.3.4 Agenda de pesquisa para o Brasil	65
2.4 Discussão	69

2.5	Conclusões	71
3	Eficiência dos prestadores de serviço de saneamento no Brasil: uma análise de fronteira estocástica entre 2015 e 2022	73
3.1	Introdução	73
3.2	Revisão de Literatura	74
3.2.1	A hipótese da propriedade e seus resultados inconclusivos	75
3.2.2	O papel mediador do ambiente regulatório	77
3.2.3	Evidências para o Brasil	79
3.2.4	O Novo Marco Legal e a condição necessária para ganhos de eficiência	80
3.3	Metodologia	81
3.3.1	Base de dados	81
3.3.2	Fundamentos teóricos: função de custo variável	82
3.3.3	Especificação: função de custo translogarítmica	83
3.3.4	Modelo de ineficiência e estratégia de estimação	85
3.3.5	Variáveis	86
3.4	Resultados	88
3.4.1	Amostra e estatísticas descritivas	88
3.4.2	Estimativas da fronteira de custo	91
3.4.3	Eficiência técnica média e comparação de modelos	94
3.4.4	Testes de robustez	97
3.5	Discussão	100
3.5.1	Nível de eficiência: posicionamento na literatura brasileira	100
3.5.2	Propriedade e gestão: resultado nulo e suas interpretações	101
3.5.3	O papel do ambiente regulatório e as implicações do NMLSB	103
3.6	Conclusões	104
4	Participação privada no mercado de água reduz tarifas? Evidências de uma reforma legal brasileira	107
4.1	Introdução	107
4.2	Revisão da literatura	110
4.2.1	Métodos de avaliação de privatizações	112
4.3	Metodologia	113
4.3.1	Construção dos indicadores	113
4.3.2	Estratégia empírica	114
4.4	Resultados	118
4.4.1	Pareamento e seleção da amostra	118
4.4.2	Efeitos sobre tarifas	122
4.4.3	Efeitos sobre atendimento	123
4.4.4	Efeitos sobre eficiência	125

4.4.5	Demais efeitos	126
4.4.6	Efeitos sobre morbidade hospitalar	131
4.4.7	Teste de Robustez	132
4.5	Discussão	135
4.6	Conclusões	140
Considerações Finais		143
Referências		147
Apêndice A Tabelas longas		169
Apêndice B Códigos de programação		174
B.1	Bibliometria	174
B.2	Fronteira Estocástica	182
B.3	Diferença em diferenças escalonado	222

Introdução

A água é reconhecida como um direito humano fundamental. Em 2010, a Assembleia Geral da Organização das Nações Unidas (ONU) declarou que o acesso à água potável segura e limpa e ao saneamento é um direito essencial para a vida e a dignidade humana, implicando que os Estados e organismos internacionais devem aumentar os esforços para fornecê-los a preços acessíveis para todos (United Nations, 2010).

O atingimento da meta de universalização da cobertura de água limpa e saneamento até 2030, definida como um dos Objetivos de Desenvolvimento Sustentável (ODS) pela ONU, requer que o atual ritmo de crescimento dessa cobertura se acelere consideravelmente (United Nations, 2024). A esse respeito, a literatura científica tem identificado a meta de “água segura e acessível” (ODS 6.1) como uma das mais investigadas (Basu; Dasgupta, 2021). Esforços mundiais têm sido empregados para aumentar e melhorar o acesso à água tratada e esgoto nas últimas décadas (United Nations, 2024): a proporção da população mundial com acesso a coleta e tratamento de esgoto, por exemplo, subiu de 49% para 57% entre 2015 e 2022 (United Nations, 2024). Ainda assim, o progresso permanece aquém do necessário: alcançar a cobertura universal de água gerenciada com segurança até 2030 exigiria multiplicar por seis o ritmo atual de avanço (Ogata *et al.*, 2026).

No Brasil, 84,9% da população é atendida com rede de água, mas há estados como Acre (48%) e Amapá (46,9%) em que essa cobertura não chega à metade de seus residentes (Brasil, 2023). Em se tratando de rede de esgoto, o cenário é ainda mais grave: a população total atendida é de 56%, com estados cujos índices não superam os 10%, como novamente o Acre (10,5%) e o Amapá (5,4%), mas também Rondônia (9%) (Brasil, 2023). Essa desigualdade não é exclusiva do Brasil: evidências para 64 países mostram que o crescimento do Produto Interno Bruto (PIB) per capita está associado a melhorias gerais no acesso à água, mas produz ganhos distributivos limitados, mais especificamente, as disparidades persistem entre os quintis mais pobres e nas populações rurais, ao passo que qualidade da governança e eficiência hídrica emergem como preditores mais robustos de acesso equitativo do que a renda per se (García-López; López-Ruiz; Cuadrado-Quesada, 2026).

Além de ser uma questão social, a universalização do saneamento é fundamentalmente um desafio financeiro (Perard, 2018). O setor exige investimentos significativos, carece de viabilidade e sustentabilidade financeira, possui uma demanda com baixa disposição a pagar e enfrenta pressões políticas para manter as tarifas baixas, além de complicados fatores institucionais como duração dos contratos e direitos de propriedade nem sempre bem definidos (Perard, 2018; Distefano, 2020; Meran; Siehlow; Hirschhausen, 2021). A compatibilização entre oferta de água de qualidade adequada, no momento e espaço corretos, e a um custo que a população possa e esteja disposta a pagar é um problema nada trivial (Ogata *et al.*, 2026): estudos revelam que

a disposição a pagar por melhorias no serviço é maior exatamente onde a qualidade é mais precária, e que a satisfação do consumidor afeta diretamente a arrecadação e a sustentabilidade financeira das concessionárias. Os elevados custos fixos, a alta intensidade de capital, o potencial relativamente baixo de inovação, a possibilidade de custos afundados substanciais, o longo período de retorno do investimento e a necessidade de operação em escala fazem desse mercado um caso peculiar da teoria econômica (Fabre; Straub, 2023; Hans, 2023; Lima; Brochado; Marques, 2021).

A gestão eficiente dos recursos também envolve a dimensão energética. No setor de tratamento de água, os custos de energia podem representar até 40% das despesas operacionais das concessionárias (Maziotis; Molinos-Senante, 2026), e a avaliação de eficiência energética das estações de tratamento é um campo em rápida expansão no contexto da transição para um ciclo urbano da água energeticamente neutro. Evidências para o Chile demonstram que ignorar a heterogeneidade entre plantas (em termos de tecnologia, fonte hídrica e estrutura de propriedade) leva a estimativas de eficiência sistematicamente enviesadas, com implicações diretas para o desenho de benchmarking regulatório (Maziotis; Molinos-Senante, 2026). No caso brasileiro, as assimetrias institucionais vão além da dimensão energética: o saneamento básico, definido legalmente como conjunto de quatro serviços interdependentes, evoluiu de forma desigual, com ênfase desproporcional no abastecimento de água e esgotamento sanitário em detrimento da drenagem urbana e do manejo de águas pluviais (Mendes; Alves, 2026). Essa fragmentação institucional constitui um dos principais entraves à universalização integral do setor.

Diante da distância ainda considerável em relação à universalização e do escasso espaço fiscal do Estado brasileiro, o setor privado emerge como alternativa para mobilizar os vultosos recursos de que o setor necessita. Por meio das Parcerias Público-Privadas (PPP), empresas privadas podem projetar, financiar, construir e operar os serviços, enquanto o setor público monitora e regula esse monopólio natural (Lima; Brochado; Marques, 2021). Trata-se de um arranjo largamente estudado no Brasil (Heller, 2007; Saiani; Azevedo, 2018; Oliveira; Saiani, 2022; Dassan *et al.*, 2024; Rufino; Cavalcante, 2024). Evidências recentes para o período 2006–2022, obtidas a partir do método de diferenças-em-diferenças sintéticas, indicam que a privatização eleva os fluxos de investimento realizados e a produtividade dos prestadores no curto prazo, melhora de forma robusta o tratamento de água e contribui para reduções modestas nas tarifas, ainda que a participação privada alcance cerca de 4% dos prestadores (Cardoso; Zoghbi; Resende, 2026). Esses resultados são encorajadores, mas precisam ser lidos com cautela: a literatura internacional registra que a alternativa privada também apresenta desafios como aumento de tarifas em determinados contextos, má qualidade do serviço e falta de transparência corporativa (Hanna; McDonald, 2021), o que em alguns casos motivou a remunicipalização como forma de recuperar o controle democrático e a prestação transparente (Warner, 2021; Warner, 2024).

No Brasil, esse debate ganhou nova dimensão com a aprovação do Novo Marco Legal do

Saneamento Básico (NMLSB), Lei n.º 14.026, de 15 de julho de 2020, que atribuiu à Agência Nacional de Águas e Saneamento Básico (ANA) a competência regulatória do setor, estabeleceu metas de universalização (99% de cobertura de água e 90% de esgoto até 2033), determinou a regionalização dos serviços e criou condições institucionais que aceleraram fortemente a transição de municípios para o modelo de provisão privada via concessões plenas. Além da dimensão privatizante, o NMLSB também demandou atenção crescente do Conselho Administrativo de Defesa Econômica (CADE): dos 30 atos de concentração analisados pelo órgão entre 2012 e 2024, 12 ocorreram após 2020, sinalizando o impacto positivo do novo marco sobre a dinâmica concorrencial do setor (Oliveira *et al.*, 2025).

A hipótese geral desta tese é que a participação privada no saneamento brasileiro não é, por si só, suficiente para garantir ganhos de eficiência e modicidade tarifária; seus efeitos dependem criticamente do ambiente regulatório, do desenho dos contratos de concessão e das condições institucionais em que os prestadores operam. Em outras palavras, a propriedade importa menos do que o contexto no qual ela se insere, o que é uma conclusão convergente com a evidência internacional de que governança e capacidade técnica são determinantes mais robustos de acesso equitativo do que a expansão econômica per se (García-López; López-Ruiz; Cuadrado-Quesada, 2026; Maziotis; Molinos-Senante, 2026).

Para investigar essa hipótese, a tese está organizada em quatro artigos. O primeiro (Capítulo 1) apresenta os conceitos fundamentais da economia da água e do saneamento, como seus aspectos microeconômicos, tarifas, privatização e experiências internacionais, bem como descreve o panorama brasileiro, com ênfase no NMLSB. O segundo artigo (Capítulo 2) mapeia a agenda de pesquisa para o Brasil por meio de uma revisão bibliométrica de 392 documentos publicados entre 1972 e 2025 nas bases Scopus e Web of Science, identificando lacunas, tendências temáticas e os principais eixos de investigação ainda em aberto. O terceiro artigo (Capítulo 3) estima uma função de custo variável translogarítmica pelo método de Análise de Fronteira Estocástica ou *Stochastic Frontier Analysis* (SFA) para 254 prestadores brasileiros no período 2015–2022, estabelecendo uma linha de base pré-reforma da eficiência técnica do setor e testando se a forma de propriedade determina os custos. O quarto artigo (Capítulo 4) avalia os efeitos causais das concessões plenas celebradas a partir do NMLSB sobre tarifas, cobertura, eficiência operacional, qualidade do serviço e morbidade hospitalar, utilizando o estimador de diferenças-em-diferenças escalonado de Callaway e Sant’Anna (2021) combinado com pareamento por escore de propensão em um painel de municípios brasileiros para o período 2017–2024.

Os principais resultados são: (i) a eficiência média dos prestadores no período pré-NMLSB foi de 75,4%, com margem de ineficiência de aproximadamente 25%, mas a propriedade privada não exerce efeito estatisticamente significativo sobre os custos, o que mais importa é o ambiente institucional; (ii) as concessões plenas produziram ganhos operacionais mensuráveis no curto prazo como a redução de despesas com pessoal, a expansão da macromedição e o aumento reportado de cobertura, mas não se traduziram em melhora dos indicadores de qualidade nem

em efeitos detectáveis sobre saúde pública; e *(iii)* os efeitos tarifários são heterogêneos entre coortes e dependem do modelo de leilão: o critério de maior outorga está associado a aumentos de tarifa, enquanto o de menor tarifa está associado a reduções. Esses achados dialogam com a literatura internacional recente (Cardoso; Zoghbi; Resende, 2026; García-López; López-Ruiz; Cuadrado-Quesada, 2026; Ogata *et al.*, 2026) e contribuem para o debate sobre as condições sob as quais a participação privada pode avançar a universalização do saneamento sem comprometer a equidade tarifária.

Por fim, cabe registrar uma consideração ética sobre o uso de ferramentas de inteligência artificial no desenvolvimento desta tese. Os códigos de análise de dados e de estimação econométrica foram desenvolvidos com o auxílio do modelo Claude (Anthropic) e estão todos disponíveis no Apêndice B; a organização e leitura dos artigos científicos contou com o suporte do NotebookLM (Google). Essas ferramentas foram utilizadas como recursos de produtividade, sem substituir o julgamento analítico e a interpretação substantiva do pesquisador, que permanecem de inteira responsabilidade do autor.

1 Economia do Saneamento: teoria, experiências internacionais e panorama brasileiro

1.1 Introdução

A água, embora essencial à vida, possui características naturais e sociais que a tornam um recurso economicamente escasso: sua distribuição é desigual no espaço e no tempo, sua disponibilidade é limitada e seu uso está sujeito à poluição e à sobreexploração. Diante disso, o acesso à água doce limpa e ao saneamento se torna um problema econômico central, exigindo decisões sobre produção, tratamento e alocação eficientes para atender múltiplos usos e garantir sustentabilidade no longo prazo (Meran; Siehlow; Hirschhausen, 2021).

O termo economia da água representa um “arranjo hidro-socioeconômico que consiste em (i) fontes de água, (ii) usuários de água, (iii) a infraestrutura física que conecta fontes e usuários, e (iv) as instituições que governam a alocação de água” (Dinar; Tsur, 2021, p. 45). As fontes de água podem ser naturais, como os aquíferos, riachos e lagos; e produzidas, como as usinas de reciclagem, que coletam e tratam o esgoto residencial e industrial, e as de dessalinização, que transformam água do mar em potável. Já os usuários são os grupos doméstico, que inclui os domicílios, escolas e hospitais; indústria; agricultura, que engloba a irrigação, aquicultura e pecuária; e meio ambiente, com a restauração de rios e suporte ecológico (Dinar; Tsur, 2021).

A infraestrutura física inclui bombas, dutos e filtros, além de capital específico para diferentes fontes (natural, reciclada, dessalinizada) e para a distribuição aos diferentes setores. As instituições governamentais, por sua vez, são aquelas que estabelecem direitos de propriedade e regras para a alocação da água (Dinar; Tsur, 2021).

Há que se considerar as características físicas da água: a mobilidade, variabilidade e o peso da água tornam seu armazenamento e transporte intensivos em capital, de modo que a infraestrutura necessária para fornecer água tratada e canalizada é de longa duração e com pouca modularidade, exigindo um sistema completo antes que se possa obter qualquer receita (Garrick; Hanemann; Hepburn, 2020).

Além de sua dimensão econômica, a água possui uma dimensão ambiental intrínseca que não pode ser dissociada de sua gestão. Em muitos contextos, como no acesso a aquíferos e no uso de corpos hídricos como receptores de efluentes, a água se comporta como um recurso de uso comum (*common pool resource*): rival no consumo, mas de difícil exclusão (Meran; Siehlow; Hirschhausen, 2021). Essa combinação de propriedades cria incentivos para a sobreexploração e para a tragédia dos comuns, em que a ausência de regras claras de acesso leva ao uso predatório

do recurso (Ostrom, 1990).

Um exemplo concreto é o esgoto doméstico e industrial: quando não há incentivos para o tratamento adequado, os custos recaem sobre terceiros na forma de externalidades negativas como poluição de rios, contaminação de aquíferos e degradação de ecossistemas aquáticos. Segundo dados da ONU-Água citados por Dinar e Tsur (2021), cerca de 80% do esgoto residencial em países em desenvolvimento é lançado sem tratamento em corpos d'água naturais. No Brasil, esse problema é amplificado pelas desigualdades regionais no acesso ao saneamento, tornando a gestão da água uma questão simultaneamente econômica, ambiental e de política pública.

A economia da água tem como escopo entender as relações entre oferta e demanda, a infraestrutura necessária, as políticas de alocação e precificação, e o impacto da gestão hídrica no bem-estar social e no desenvolvimento econômico (Dinar; Tsur, 2021). Este capítulo, portanto, objetiva esclarecer características do mercado de água e saneamento, quais sejam: o aspecto de monopólio natural, a provisão estatal e a privatização.

1.2 Monopólio natural

O termo mercado pode ser definido como o ambiente no qual compradores e vendedores interagem entre si para produzir e comercializar bens ou serviços. Ele será dito competitivo caso haja um grande número de vendedores sem qualquer influência sobre o preço praticado. Já, no extremo oposto, tem-se o monopólio em que há apenas um vendedor e, caso não seja regulado, está livre para determinar preço e quantidades ofertadas de modo a maximizar seu lucro (Sharkey, 1982).

Seja $C(q)$ o custo total da firma na produção de q , pode-se denotar $C(q)$ como o custo mínimo de uma combinação de insumos que permite a produção de q unidades de produto. Supõe-se que a função custo seja duas vezes diferenciável, exceto no caso da produção ser nula:

$$C(q) = \begin{cases} F + \int_0^q C'(x) dx & \text{for } q > 0 \\ 0 & \text{caso contrário,} \end{cases}$$

onde $F \geq 0$ denota o custo fixo da produção, aquele em que a firma deve incorrer em razão de produzir mas que não depende do nível de produção (Tirole, 1988).

Tomando o conceito de uma função de custo de produção para uma empresa monopolista, seja $q = (q_1, \dots, q_k)$ um vetor de produtos em um mercado específico, então $C(q)$ representará o custo dos insumos necessários para a empresa única produzir q . Em comparação a um mercado com múltiplas empresas, pode-se analisar o custo relativo da produção de q , considerando que todas as empresas possuem a mesma função custo C , então uma única empresa será mais eficiente do que um mercado competitivo se

$$C(q) < C(q^1) + \dots + C(q^k), \quad (1.1)$$

ou seja, será sempre mais barato ter apenas uma empresa nesse mercado (Baumol, 1977). Se a desigualdade (1.1) valer para qualquer desagregação possível do vetor de produção q , então C é dita ser subaditiva em q e o mercado em questão é considerado um monopólio natural em q (Sharkey, 1982).

A subaditividade é a condição necessária e suficiente para a caracterização do monopólio natural (Baumol, 1977), no entanto cabe definir outras questões sobre custos: as economias de escala e de escopo e os custos médios decrescentes. Existem economias de escala estritas na produção de N se para cada vetor inicial de insumos-produtos $(x_1, \dots, x_r, y_1, \dots, y_n)$ e para qualquer $w > 1$, seja factível um vetor $(wx_1, \dots, wx_r, v_1y_1, \dots, v_ny_n)$, onde todo $v_i \geq w + \delta, \delta > 0$ (Baumol, 1977).

Essa definição diz que haverá economias de escala quando, ao aumentar a quantidade de insumos w vezes, a produção aumente w mais um fator positivo, indicando um aumento mais que proporcional ao incremento de insumos (Baumol, 1977). Já as economias de escopo surgem quando, ao considerar q_1 e q_2 como duas quantidades de produtos diferentes com base em uma função de custo subaditiva:

$$C(q_1, 0) + C(0, q_2) > C(q_1, q_2),$$

neste caso, uma empresa que produz os dois bens será tecnologicamente mais eficiente do que duas firmas especializadas em uma das duas produções (Tirole, 1988).

Das economias de escala decorrem a definição restante, a dos custos médios decrescentes, que diz que os custos médios serão estritamente decrescentes se, para q_a e q_b tais que $0 < q_a < q_b$, temos:

$$\frac{C(q_b)}{q_b} < \frac{C(q_a)}{q_a}.$$

O que significa ser vantajoso para a firma sempre aumentar sua produção na medida que seu custo por unidade produzida se reduz (Tirole, 1988).

O mercado de água é frequentemente caracterizado como um monopólio natural devido a uma série de fatores inerentes à sua infraestrutura e à natureza do serviço (Dassan *et al.*, 2024). Os custos fixos são altos e as economias de escala, significativas, na medida em que os sistemas de água exigem grandes investimentos iniciais em infraestrutura como redes de tubulações, estações de tratamento e sistemas de armazenamento (Galiani; Gertler; Schargrodsky, 2005). Além disso, as empresas de água operam como monopólios verticalmente integrados, desde a captação da água bruta até a torneira do consumidor, o que reduz a possibilidade de competição em diferentes etapas do processo (Allen; Pryke, 2013), tornando a concorrência economicamente inviável e o consumidor não poderá escolher entre diferentes ofertantes (Oliveira; Saiani, 2022).

A regulação tarifária é particularmente crítica nesse contexto. Como aponta Lazzarini (2023, p. 280), uma regulamentação defeituosa pode fazer com que os preços estabelecidos sejam muito altos, o que por sua vez pode provocar divergências públicas e deslegitimar o modelo de provisão privada. Por outro lado, quando o governo tenta manter os preços artificialmente

baixos, o investimento privado pode ser insuficiente, comprometendo a qualidade e a expansão dos serviços. Esse dilema regulatório é central para o sucesso ou fracasso das privatizações no setor de saneamento.

1.2.1 Comportamento de precificação do monopolista

A implicação mais relevante, do ponto de vista do bem-estar social, decorrente do monopólio é o seu comportamento de precificação, cujo resultado é uma perda de peso morto (Tirole, 1988).

Sejam $q = D(p)$ a demanda pelo bem produzido pelo monopólio, com função de demanda inversa $p = P(q)$ e $C(q)$ o custo de produzir q unidades desse bem. Considerando que a demanda é diferenciável e decrescente com o preço (ou seja, $D'(p) < 0$), e que o custo é diferenciável e crescente com a produção, um monopolista maximizador de lucros estabelece um preço de monopólio p^m tal que:

$$\max_p [pD(p) - C(D(p))].$$

Cuja condição de primeira ordem é:

$$p^m - C'(D(p^m)) = -\frac{D(p^m)}{D'(p^m)},$$

ou:

$$\frac{p^m - C'}{p^m} = \frac{1}{\varepsilon}, \quad (1.2)$$

onde $\varepsilon = -D'p^m/D$ denota a elasticidade da demanda no preço de monopólio p^m . Definindo $q^m \equiv D(p^m)$ como a produção de monopólio, pode-se reescrever a condição de primeira ordem como a igualdade entre receita marginal e custo marginal:

$$RM(q^m) \equiv P(q^m) + P'(q^m)q^m = C'(q^m).$$

A equação 1.2 indica que o *markup* — a razão entre a margem de lucro (preço menos custo marginal) e o preço, também chamada de índice de Lerner — é inversamente proporcional à elasticidade da demanda (Tirole, 1988). Tal distorção de preço pode ser verificada em termos de excedente do consumidor por meio da Figura 1.1, na qual se observa o preço praticado no monopólio (p^m) que é superior ao de equilíbrio competitivo (p^c).

O excedente total é igual à soma dos excedentes do consumidor e do produtor, representado pela área ADG na Figura 1.1 para o caso do preço competitivo ($p^c = C'$), já no caso de um monopólio esse excedente passa a ser a área ADEF. Comparativamente, o excedente do consumidor (área acima do preço competitivo e abaixo da curva de demanda) é em parte apropriado pelo monopolista (área verde) e parte do excedente tanto do consumidor como do

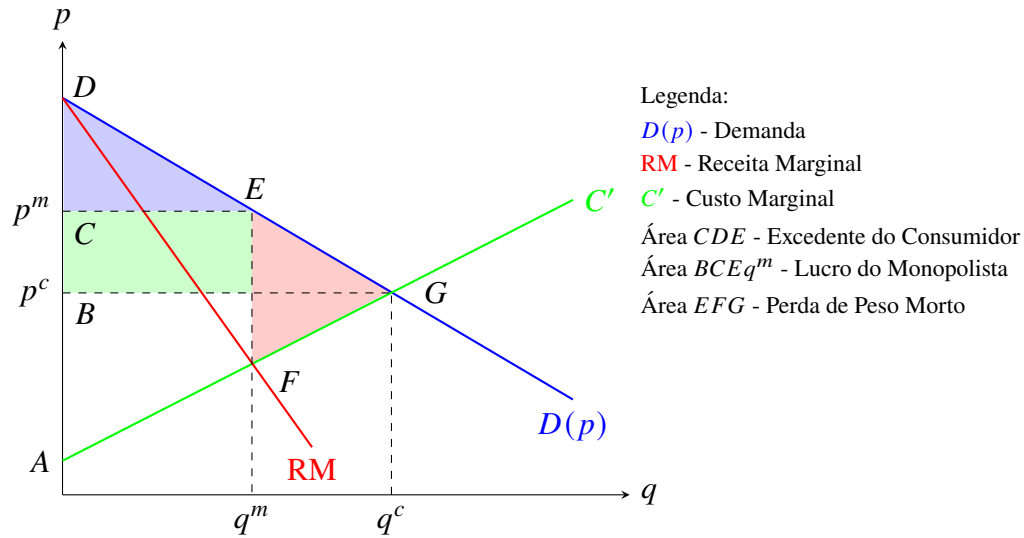


Figura 1.1 – Perda de Peso Morto no Monopólio (Adaptado de Tirole (1988)).

produtor (área vermelha) é perdida, o que leva o nome de perda de peso morto, um excedente que não é apropriado neste mercado, causando perda de bem-estar social (Tirole, 1988).

Outro aspecto a se destacar acerca da precificação do monopolista é o fato de que este sempre opera na região elástica da curva de demanda (Equação 1.2), ou seja, onde a elasticidade-preço da demanda ε satisfaz $|\varepsilon| > 1$ (Tirole, 1988). Isso decorre do fato de que, para $|\varepsilon| < 1$, um aumento no preço eleva a receita total, implicando que a receita marginal se torna negativa. Como a condição de maximização de lucro requer que a receita marginal iguale o custo marginal ($RM = C'$), a maximização do lucro nunca ocorre na região inelástica da demanda (Tirole, 1988).

Uma propriedade geral da precificação monopolista é que o preço de monopólio é uma função *não decrescente* do custo marginal: considere duas funções de custo distintas e diferenciáveis, $C_1(\cdot)$ e $C_2(\cdot)$, e que $C'_2(q) > C'_1(q) \quad \forall q > 0$. Sejam (p_1^m, q_1^m) o par preço-quantidade de monopólio associado a $C_1(q)$, e (p_2^m, q_2^m) o correspondente a $C_2(q)$ (Tirole, 1988). Como p_1^m maximiza o lucro sob $C_1(q)$, tem-se:

$$p_1^m q_1^m - C_1(q_1^m) \geq p_2^m q_2^m - C_1(q_2^m).$$

De forma análoga, dado que p_2^m é ótimo sob $C_2(q)$, obtém-se:

$$p_2^m q_2^m - C_2(q_2^m) \geq p_1^m q_1^m - C_2(q_1^m).$$

Somando as inequações acima:

$$(p_1^m q_1^m - C_1(q_1^m)) + (p_2^m q_2^m - C_2(q_2^m)) \geq (p_2^m q_2^m - C_1(q_2^m)) + (p_1^m q_1^m - C_2(q_1^m)),$$

o que implica:

$$[C_2(q_1^m) - C_2(q_2^m)] - [C_1(q_1^m) - C_1(q_2^m)] \geq 0.$$

Aplicando o Teorema do Valor Médio para funções diferenciáveis:

$$\int_{q_2^m}^{q_1^m} [C_2'(x) - C_1'(x)] dx \geq 0. \quad (1.3)$$

Como $C_2'(x) > C_1'(x)$ para todo $x > 0$, segue que a Equação 1.3 implica que $q_1^m \geq q_2^m$, ou seja, quanto maior o custo marginal, menor a quantidade ofertada pelo monopolista. Logo, o preço de monopólio é uma função *não decrescente* do custo marginal (Tirole, 1988).

Ocorre que a elasticidade-preço da demanda por água é pequena para cada uso, seja doméstico, industrial ou agrícola (Ciriacy-Wantrup, 1961). Apesar da heterogeneidade observada, os estudos empíricos mais recentes mostram que a demanda por água é inelástica aos preços observados (Bruno; Jessoe, 2021). Desse modo, o comportamento típico de precificação do monopolista, demonstrado anteriormente, não se aplica.

1.3 Tarifas

Idealmente, as tarifas de água e esgoto apresentam algumas características desejáveis como sustentabilidade financeira e ambiental, eficiência na alocação de recursos e equidade social (Garrick; Hanemann; Hepburn, 2020). A suficiência de receitas relaciona-se com a capacidade das tarifas de cobrir todos os custos na prestação do serviço (Meran; Siehlow; Hirschhausen, 2021). Caso os custos não sejam totalmente cobertos, não será possível operar e gerir o sistema de abastecimento de água de forma eficaz e eficiente, com impactos sobre os investimentos que, por consequência, resultará em um serviço pior (Meran; Siehlow; Hirschhausen, 2021). A piora do serviço gera insatisfação e acarreta uma diminuição na disposição a pagar dos consumidores.

Em termos ambientais, como a conservação dos recursos hídricos desempenha um papel crucial para alcançar a sustentabilidade, as tarifas devem ser capazes de incentivar a proteção da água na natureza (Meran; Siehlow; Hirschhausen, 2021). Para tanto, devem levar em conta o princípio da recuperação dos custos dos serviços de água, incluindo os custos ambientais e de recurso, o que significa incorporar externalidades ambientais e o valor *in situ* (ou preço sombra) da água natural nos preços de usuário (Meran; Siehlow; Hirschhausen, 2021).

A eficiência econômica ou alocativa é o atributo segundo o qual a política de precificação deve alocar a água para os usuários que mais se beneficiam de seu recebimento (Meran; Siehlow; Hirschhausen, 2021). Se a água é alocada para usuários com baixos benefícios marginais, enquanto outros com benefícios marginais mais altos não são atendidos, o princípio da eficiência econômica é violado, uma vez que não maximiza a renda econômica agregada de todos os consumidores de água (Meran; Siehlow; Hirschhausen, 2021).

Outro aspecto de eficiência carregado pelo preço é o desincentivo ao uso ineficiente de recursos, o preço da água tem um papel econômico importante em sinalizar o valor de escassez e o custo de oportunidade da água (Meran; Siehlow; Hirschhausen, 2021). Teoricamente, para se garantir o uso eficiente de uma *commodity*, ela deve ser precificada ao seu custo marginal de longo prazo. Entretanto, conforme apontado na seção anterior, em setores com custos médios decrescentes (monopólios naturais), a precificação estritamente ao custo marginal pode resultar em déficits de receita (Meran; Siehlow; Hirschhausen, 2021).

Já para países em desenvolvimento, cidadãos e localidades de baixa renda, o aspecto mais relevante das tarifas é a equidade social, que inclui considerações sobre acessibilidade (*affordability*) e justiça (*fairness*) (Meran; Siehlow; Hirschhausen, 2021). A acessibilidade pode ser avaliada pela relação entre as despesas com o serviço de água e a renda total da família, com limiares sugeridos de 3 a 5% da renda familiar (Meran; Siehlow; Hirschhausen, 2021).

A estrutura tarifária mais comum em países de baixa e média renda é a tarifa em blocos crescentes (*IBT - Increasing Block Tariff*) (Andres *et al.*, 2019). Nessa estrutura, os níveis de consumo são divididos em faixas, com um preço unitário diferente aplicado dentro de cada faixa. O preço unitário cobrado nas faixas de consumo mais baixas é menor (e pode até ser gratuito) do que os preços unitários cobrados nas faixas de consumo mais altas (Andres *et al.*, 2019).

Em termos práticos, independentemente do orçamento, todos deveriam ter acesso ao nível de subsistência, ou seja, 20 litros per capita por dia de água potável necessária para atingir os níveis mínimos essenciais para saúde e higiene (World Health Organization WEDC, 2013). Outra medida frequentemente proposta para o aspecto social é fornecer um certo limite de consumo necessário a um preço baixo ou zero, este limite é frequentemente quantificado em 50 litros/dia por pessoa (Massarutto, 2020).

O desconto de preço para este limite essencial poderia ser financiado por tributação geral ou por subsídios cruzados provenientes de blocos de consumo ou renda mais elevados, conferindo um aspecto de justiça à tarifa (Massarutto, 2020; Dassan *et al.*, 2024). Em teoria, a estrutura IBT permite que as concessionárias atinjam os objetivos de equidade (através das faixas mais baixas) e sustentabilidade ambiental (através das faixas mais altas) (Andres *et al.*, 2019).

1.4 Presença do Estado

O saneamento possui consideráveis externalidades sociais, especialmente relacionadas à saúde e ao meio ambiente, além de ser um bem essencial e, em alguns contextos, considerado um direito humano (Kresch, 2020; Oliveira; Saiani, 2022; Dassan *et al.*, 2024). O setor requer altos investimentos iniciais com retornos de longo prazo (Oliveira; Saiani, 2022), e o governo muitas vezes desempenha papel fundamental no financiamento ou na garantia desses investimentos (Dassan *et al.*, 2024). Dadas as características de monopólio natural com demanda de baixa elasticidade, o setor enfrenta o desafio de não cobrir os custos de operação e manutenção, nem de

recuperar o capital investido com a receita tarifária (Perard, 2018; Andres *et al.*, 2019; Garrick; Hanemann; Hepburn, 2020). Essa insuficiência decorre, em parte, da baixa disposição a pagar dos consumidores, que pressionam politicamente para manter as tarifas baixas (Perard, 2018). No caso do esgoto, o cenário é ainda mais desafiador: a baixa disposição a pagar faz com que os governos não priorizem investimentos nessa área em comparação ao abastecimento de água tratada (Perard, 2018), fato corroborado pela discrepância entre os percentuais de cobertura de água e de coleta de esgoto nos domicílios brasileiros (Brasil, 2023).

Adicionalmente, o setor possui dimensões de qualidade não perfeitamente contratáveis: é difícil antecipar todas as eventualidades e monitorar adequadamente o desempenho, especialmente em relação a aspectos da qualidade não diretamente observáveis pelos usuários, como impacto na saúde e no ambiente (Saiani; Azevedo, 2018; Oliveira; Saiani, 2022). A ausência de mecanismos de arbitragem de mercado, como a escolha do consumidor, torna a regulação ou provisão pública necessária não apenas para conter preços, mas também para garantir os níveis desejados de qualidade (Oliveira; Saiani, 2022). Por fim, a provisão estatal visa frequentemente garantir o acesso universal, uma vez que levar o serviço a áreas mais pobres ou menos desenvolvidas pode não ser atrativo para provedores privados focados no lucro (Kresch, 2020).

Teoricamente, um trabalho bastante influente nessa discussão é o de Hart, Shleifer e Vishny (1997), no qual os autores analisaram as condições que determinam se é mais eficiente para o governo prover um serviço público diretamente ou delegá-lo à iniciativa privada. Seus achados teóricos são, entre outros, de que quando há a possibilidade de que reduções de custos não contratuais possuam grandes efeitos deletérios sobre a qualidade e quando as inovações em qualidade são irrelevantes, a provisão pública é a mais adequada (Hart; Shleifer; Vishny, 1997).

1.5 Privatização

Há um certo consenso de que entidades públicas são geralmente menos eficientes ou rentáveis que firmas privadas, isso porque prestadores de serviço privados teriam incentivos mais fortes à eficiência produtiva via redução de custos e elevação de produtividade e rentabilidade (Galiani; Gertler; Schargrodsky, 2005; Oliveira; Saiani, 2022; Dassan *et al.*, 2024).

Tais argumentos subsidiam a ideia de que a privatização dos serviços de água e esgoto possa ser um caminho para a melhoria de acesso, qualidade e preço. No entanto, preocupações emergem acerca da possibilidade de comprometimento do acesso à água como um direito humano. Nesta seção são apresentadas algumas experiências de privatização do saneamento ao redor do mundo.

1.5.1 Reino Unido

A privatização do setor de água na Inglaterra e no País de Gales ocorreu em 1989 e consistiu na venda de dez autoridades regionais de água para o setor privado, formando dez companhias

privadas monopolistas regionais (Bakker, 2001). Essas empresas, também responsáveis pelos serviços de esgoto, foram listadas na bolsa de valores de Londres (Bakker, 2001).

Esse é um caso emblemático porque privatizou completamente os sistemas de abastecimento de água e esgoto através da venda de ativos, diferentemente dos modelos de concessão usados em outros países (Bakker, 2001).

Anteriormente, os critérios de desempenho não priorizavam a eficiência técnica ou econômica, mas sim a segurança dos sistemas de abastecimento de água e a saúde pública de modo que as tarifas não eram determinadas pela necessidade de recuperação de custos, mas pelos objetivos de provisão universal e acesso equitativo (Bakker, 2001). Nesse sentido, a propriedade pública era considerada necessária devido a falhas de mercado como monopólio e externalidades, para garantir a provisão universal e o acesso equitativo (Bakker, 2001).

O processo de venda das empresas ocorreu em um contexto de necessidade de atender às exigências de qualidade da água da União Europeia, em que se julgou que o setor privado teria mais eficiência e maior capacidade de financiar os investimentos requeridos (Dore; Kushner; Zumer, 2004). Cada empresa recebeu um monopólio regional por 25 anos e três agências reguladoras foram estabelecidas para supervisionar o setor: o Office of Water Services (OFWAT), regulador econômico responsável pelos interesses do consumidor, padrões de serviço e preços; a Agência de Meio Ambiente e a Agência de Proteção à Saúde (Dore; Kushner; Zumer, 2004).

Cabe notar que, para tornar as empresas atraentes aos compradores, houve uma política de equalização de custos no período anterior à privatização, na qual o governo cancelou dívidas das autoridades públicas, ofereceu subsídios e regulações favoráveis (Bakker, 2001). Isso penalizou as autoridades de água regionais que controlavam seus custos e beneficiou aquelas com custos mais altos, removendo incentivos para conter custos e levando a preços globais mais altos do que o necessário (Bakker, 2001).

A eficiência se tornou um critério de desempenho primordial, de modo que as empresas têm um dever estatutário de conservar água e a Agência de Meio Ambiente examina essas medidas de eficiência técnica antes de conceder novas licenças de captação (Bakker, 2001). A regulação evoluiu ao longo do tempo, com a OFWAT se tornando um órgão regulador cada vez mais intrusivo, ao controlar resultados detalhados e impor mais obrigações sociais (Cashman; Lewis, 2007).

Em termos de eficiência e produtividade, houve melhorias modestas na eficiência técnica, porém atribuídas mais à regulação e à pressão por desempenho do que à privatização em si (Saal; Parker, 2001; Saal; Parker; Weyman-Jones, 2007). Ao se comparar a eficiência entre empresas públicas e privadas, não foi encontrada uma clara vantagem para o setor privado (Walter *et al.*, 2009).

De fato, após a privatização, as empresas investiram na modernização da infraestrutura e no cumprimento de normas ambientais e sanitárias mais rígidas (Dore; Kushner; Zumer, 2004). A maior parte dos gastos foi direcionada para melhorar a qualidade da água potável e minimizar

os impactos ambientais (Bakker, 2001).

Para financiar a melhoria dos serviços, houve um rápido aumento nas tarifas de água e esgoto desde 1989 (Bakker, 2001; Byatt, 2013). Os preços da água no Reino Unido cresceram mais rapidamente do que a inflação, especialmente na década de 1990, o que gerou críticas de consumidores e da mídia (Dore; Kushner; Zumer, 2004).

Os altos preços cobrados por tarifas fixas (que independem do consumo), e a regulação que limitou o "diferencial", preço cobrado quando os consumidores ultrapassam a cobrança mínima, foram responsáveis pela mudança de grande parte dos consumidores para tarifas baseadas no consumo, mesmo tendo que pagar pela instalação do medidor (Byatt, 2013).

A aceitabilidade política foi uma questão desde o início do processo, sendo a privatização da água bastante impopular, e o incidente da seca em Yorkshire entre 1995 e 1996, combinada com os altos níveis de vazamento não controlados pela empresa responsável, só agravou a situação (Haughton, 1998; Bakker, 2000; Byatt, 2013).

Os reguladores também foram criticados por não terem previsto a crise e por não terem garantido investimentos adequados pela empresa antes do evento (Haughton, 1998; Bakker, 2000). Como resposta, destaca-se a imposição de um imposto sobre os lucros excessivos (*windfall tax*) sobre as companhias privatizadas (Bakker, 2000). A oposição à privatização na Inglaterra e no País de Gales contribuiu para que a Escócia mantivesse seu setor de água em propriedade pública (Gibbs, 2024).

1.5.2 Chile

O Chile privatizou a água por meio do Código de Águas de 1981, separando os direitos de uso da água da propriedade da terra e os transformou em direitos privados e comercializáveis, de modo que a atuação estatal foi bastante reduzida (Bauer, 1997; Correa-Parra; Vergara-Perucich; Aguirre-Núñez, 2020).

Entre 1998 e 2000, houve a venda das companhias públicas regionais para entidades privadas; já entre 2001 a 2004 houve outra forma de privatização, a concessão de direito de operação por períodos longos, como 30 anos; finalmente em 2011 o governo vendeu parte das ações remanescentes, ficando minoritário nas companhias (Molinos-Senante; Sala-Garrido, 2015).

Quanto à primeira fase, esperava-se criar um mercado autorregulado, eficiente e livre, no entanto isso gerou forte concentração de propriedade e problemas de especulação (Bauer, 1997; Correa-Parra; Vergara-Perucich; Aguirre-Núñez, 2020). A atividade de mercado foi limitada e desigual, com efeitos positivos localizados, mas com muitos problemas jurídicos, sociais e administrativos (Bauer, 1997). Em particular, os incentivos à eficiência foram enfraquecidos por ausência de taxas sobre direitos não utilizados (Bauer, 1997).

Em comparação das duas formas de privatização adotadas pelo Chile, Molinos-Senante e

Sala-Garrido (2015) usaram uma abordagem quantitativa de produtividade, com base indicador de Luenberger para 18 empresas de saneamento entre 1997 e 2013. Seus resultados dão conta do aumento na produtividade, via melhorias técnicas, das companhias totalmente privatizadas; já o oposto ocorreu com as que receberam concessões operacionais, sendo que o modelo de regulação adotado, benchmarking (empresa hipotética eficiente), funcionou apesar das assimetrias de informação (Molinos-Senante; Sala-Garrido, 2015).

Correa-Parra, Vergara-Perucich e Aguirre-Núñez (2020) calcularam o índice de Gini para os direitos de uso de água no Chile com base em dados de 2020, encontrando níveis elevados de desigualdade e concentração em todas as regiões do país, com Gini nacional de 0,9537 para direitos de consumo superficial. Os autores não realizam comparação com o período anterior à privatização, mas argumentam que o próprio desenho institucional do Código das Águas de 1981 é estruturalmente gerador dessa concentração.

1.5.3 França

Na França, a responsabilidade legal pela provisão de serviços de água e saneamento pertence às autoridades públicas locais, que decidem ofertar diretamente via empresa pública ou delegar a um operador privado (Barraqué, 2012). Geralmente, a infraestrutura permanece de propriedade do município e a operação é delegada por meio de contratos (Barraqué, 2012).

Os contratos assumem diferentes formas: concessão, em que o concessionário recebe o direito de operar o serviço por um longo período (30 anos ou mais), arca com os custos operacionais e de capital e é remunerado pela tarifa; arrendamento, similar à concessão com a diferença de que financia apenas os custos operacionais e de manutenção pela tarifa (a infraestrutura continua sob financiamento da autoridade local), sendo o tipo de contrato mais comum; gestão com bônus, em que o operador recebe as contas dos clientes; e gestão simples (contrato de serviço), em que não há contato com os clientes, apenas com a autoridade pública (Barraqué, 2012).

Historicamente, houve grande crescimento da participação privada no saneamento francês, passando de 30% em meados da década de 1950 para 75% na década de 1990 (Barraqué, 2012). No entanto, a gestão pública ainda representava cerca de 35% dos prestadores de serviços de água em 2010 (Barraqué, 2012).

As comparações de desempenho entre servidores públicos e privados franceses normalmente utilizam variáveis como preços, qualidade, eficiência e níveis de dívida (Chong; Saussier; Silverman, 2015; Porcher, 2017). Com dados em painel de quase 500 autoridades de água entre 1998 e 2008, Chong, Saussier e Silverman (2015) estimaram um preço esperado com base nas características observáveis dos municípios para identificar possíveis sobrepreços usando regressões com efeitos fixos.

A gestão privada, em média, tende a cobrar tarifas mais altas que a gestão pública direta

(Chong; Saussier; Silverman, 2015; Porcher, 2017). No entanto, ao considerar as dívidas acumuladas pelos serviços públicos (custos ocultos), essa diferença desaparece (Porcher, 2017).

A gestão pública direta tende a operar com preços cobrados mais baixos no curto prazo, mas acumula níveis de dívida mais altos no orçamento do serviço de água (Porcher, 2017). A gestão privada, com preços cobrados mais altos, tende a ter níveis de dívida mais baixos, o que pode indicar uma maior adesão ao princípio de recuperação de custos totais por meio das tarifas (Porcher, 2017).

Outro aspecto analisado faz referência ao tamanho dos municípios, no qual o prêmio de preço associado à gestão privada é concentrado em municípios menores, com menor poder de barganha, enquanto que nos municípios maiores a diferença de preço é insignificante (Chong; Saussier; Silverman, 2015).

Em termos de qualidade e eficiência, a literatura aponta para uma melhora da qualidade da água ao longo do tempo (Chong; Saussier; Silverman, 2015), mas não encontra evidências de que o setor privado tenha uma vantagem de eficiência absoluta (Dore; Kushner; Zumer, 2004).

A água fornecida por operadores privados, em alguns estudos, mostrou melhor qualidade em termos de bactérias, produtos químicos e vazamentos em comparação com a gestão pública, no entanto cabe notar que esses resultados são devidos a uma maior intensidade de tratamento que, por sua vez, estão associadas a preços mais elevados (Chong; Saussier; Silverman, 2015).

O modelo francês, por vezes, careceu de uma regulamentação econômica e monitoramento adequados dos contratos, a ausência de informações para o controle social foi marcante (Lobina; Hall, 2007). Escândalos de corrupção foram centrais para a criação da Lei Sapin de 1993, que buscou aumentar a transparência e instituiu a necessidade de licitação para a renovação de contratos, caso contrário a gestão deveria retornar ao setor público (Lobina; Hall, 2007; Barraqué, 2012).

Em particular, duas cidades merecem atenção devido ao processo de reestatização dos serviços: Paris e Grenoble (Lobina; Hall, 2007; Barraqué, 2012). Grenoble privatizou os serviços em 1989 e voltou a ser público em 2001, já em Paris, o serviço foi reestatizado em 2010, após décadas de gestão privada. Nas duas localidades, a volta ao setor público trouxe melhora na transparência, investimento e controle social, sem prejuízo à qualidade do serviço (Lobina; Hall, 2007; Barraqué, 2012).

1.5.4 Alemanha

Diferentemente de países como Chile e Reino Unido, que privatizam de forma ampla o setor de água, a Alemanha deu autonomia aos municípios para adotarem formas variadas de parcerias público-privadas e privatizações (Ruester; Zschille, 2010; Beveridge; Hüesker; Naumann, 2014). Como resultado, apenas 6% das companhias foram totalmente privadas, enquanto a maioria opera sob modelos cooperativos ou com participação majoritária pública.

Destaca-se o caso de Berlim, que vendeu 49,9% das ações da companhia municipal a entidades privadas em 1999 sob um contexto de grande endividamento público e tentativa de atrair investimentos privados (Beveridge; Hüesker; Naumann, 2014). A privatização gerou insatisfação pública devido à falta de transparência, tarifas altas e perda de controle público, o que resultou na remunicipalização entre 2012 e 2013 (Beveridge; Hüesker; Naumann, 2014).

Análise econométrica com base em dados de 765 empresas alemãs mostrou que a participação do setor privado está associada a preços mais altos ao consumidor, mesmo após controle por variáveis técnicas e estruturais como densidade populacional, tipo de água (subterrânea ou superficial) e eficiência de rede (Ruester; Zschille, 2010).

Politicamente, a insatisfação da sociedade civil pressionou os partidos políticos a incluírem a oposição à privatização da água em seus manifestos eleitorais (Tosun; Triebkorn, 2020). O movimento *Right2Water*, iniciativa europeia contra a liberalização dos serviços de água, teve grande apoio na Alemanha, especialmente entre sindicatos e partidos de esquerda (Tosun; Triebkorn, 2020).

1.5.5 Argentina

A privatização dos serviços de água na Argentina iniciou nos anos 1990, como parte de um amplo programa de reformas estruturais para reduzir déficits fiscais e atrair investimento privado. Foram adotadas concessões de longo prazo (30 anos), nas quais operadores privados assumiram a operação e financiamento dos serviços e cujos contratos envolviam metas de expansão de cobertura e criação de agências reguladoras (Estache; Trujillo, 2003; Casarin; Delfino; Delfino, 2007).

A cidade de Buenos Aires foi o caso mais relevante, com a concessão da empresa Aguas Argentinas em 1993 (Estache; Trujillo, 2003; Casarin; Delfino; Delfino, 2007). Em termos de saúde, o resultado foi positivo: houve redução de 8% na mortalidade infantil em áreas que privatizaram, sendo que o efeito foi ainda mais forte (26%) nas áreas mais pobres, onde a expansão da rede foi maior (Galiani; Gertler; Schargrotsky, 2005).

Essa investigação de Galiani, Gertler e Schargrotsky (2005) possui relevante impacto científico por inovar metodologicamente na avaliação de impacto: os autores empregaram o método de diferenças em diferenças para medir o efeito da privatização sobre a mortalidade infantil.

Uma análise do desempenho da concessão de Buenos Aires em relação a metas de cobertura, tarifas e padrões de serviço, no entanto, encontrou descumprimento de metas, tarifas elevadas e qualidade de serviço aquém do esperado, apesar da expansão da cobertura (Casarin; Delfino; Delfino, 2007). A explicação, nesse caso, foi a de que a ausência de força e experiência do regular tenha favorecido o comportamento oportunista do concessionário e a crise macroeconômica de 2001 comprometeu o equilíbrio financeiro da concessão, levando a sucessivas tentativas frustradas

de renegociação e, finalmente, ao cancelamento do contrato pelo governo e à reestatização do serviço em 2006 (Casarin; Delfino; Delfino, 2007).

Outras cidades também foram estudadas pela literatura como Tucumán, Santa Fé, Córdoba, Mendoza, Corrientes e a província de Salta (Clarke; Kosec; Wallsten, 2009; Hailu; Osorio; Tsukada, 2012). Em geral, a parcela de domicílios conectados à água encanada e esgoto e a qualidade física, química e bacteriana melhorou, após a introdução da participação do setor privado no país (Clarke; Kosec; Wallsten, 2009).

Em Tucumán, as tarifas de água após o controle privado dobraram, o que contribuiu para a resistência dos consumidores, razão fundamental para a rescisão da concessão (Israel, 2007). A concessão em Santa Fé (iniciada em 1998) também foi renacionalizada em 2006 (Hailu; Osorio; Tsukada, 2012).

Em termos macroeconômicos, um modelo de equilíbrio geral computável (CGE) foi utilizado para estimar os impactos distributivos da privatização e da regulação (Chisari; Estache; Romero, 1999). Os benefícios macroeconômicos encontrados foram significativos, mas os ganhos da privatização favoreceram mais as classes altas em razão da regulação ter sido ineficiente (Chisari; Estache; Romero, 1999).

Já ao se tratar de produtividade, houve ganhos de eficiência via produtividade total dos fatores, com base em dados de quatro operadoras que se mostraram bastante heterogêneas (Estache; Trujillo, 2003). A falta de dados e novamente a regulação fraca são apontadas como razões que comprometeram avaliações mais precisas (Estache; Rossi, 2002).

Politicamente, mesmo pessoas que ganharam acesso à água em virtude da privatização tinham opiniões moderadamente negativas sobre esse processo (Di Tella; Galiani; Schargrodsky, 2012). No caso concentro da Argentina, a exposição à propaganda política do governo contra a empresa privada moldou crenças sobre a privatização, o que reduziu significativamente o apoio popular (Di Tella; Galiani; Schargrodsky, 2012).

1.5.6 Bolívia

Na Bolívia, as cidades de La Paz, Santa Cruz de La Sierra, El Alto e Cochabamba tiveram o saneamento privatizado nos anos 1990 via concessão com o objetivo de melhorar os serviços de água e ampliar a cobertura (Israel, 2007; Clarke; Kosec; Wallsten, 2009; Hailu; Osorio; Tsukada, 2012). As taxas de conexão para domicílios no quintil de renda inferior em cidades bolivianas mostram que em 1994 apenas 13.45% dessas residências tinham água encanada (Israel, 2007).

Em média, o resultado foi positivo na medida em que cobertura de água encanada e esgoto é significativamente maior em cidades com participação do setor privado ao controlar por população, renda e uma tendência temporal (Clarke; Kosec; Wallsten, 2009).

Não há evidências de que a privatização tenha prejudicado os pobres, pois quando se analisa seu impacto na cobertura para domicílios de baixa renda, no quintil de renda inferior, os

resultados sugerem que tanto a cobertura de água encanada quanto de esgoto sanitário interno é maior nas cidades com participação do setor privado (Clarke; Kosec; Wallsten, 2009).

Já em estudo de caso, percebe-se que o resultado foi positivo para La Paz e El Alto, neutro para Santa Cruz de La Sierra e negativo para Cochabamba, quando se considera o quintil de renda mais baixo dessas localidades entre os anos de 1996 e 2005 (Hailu; Osorio; Tsukada, 2012). O motivo do sucesso das duas primeiras relaciona-se com o contrato de concessão, cuja licitação foi conduzida com base em metas de cobertura, em vez de reduções de tarifa: a concessionária se comprometeu a aumentar a cobertura de esgoto para 90% em El Alto e 95% em La Paz (Clarke; Kosec; Wallsten, 2009).

O caso negativo de Cochabamba também se relaciona com o contrato de concessão, que incluía uma garantia de lucratividade para a operadora (Nickson; Vargas, 2002; Hailu; Osorio; Tsukada, 2012). A população se revoltou contra os aumentos nas tarifas de água, e o contrato que foi celebrado em 1999 foi cancelado em 2000 devido à desordem civil generalizada (Nickson; Vargas, 2002; Clarke; Kosec; Wallsten, 2009; Hailu; Osorio; Tsukada, 2012).

1.5.7 Estados Unidos

Nos Estados Unidos da América (EUA), a privatização dos serviços de água ocorreu predominantemente por meio de contratos de operação e manutenção, concessões e, em alguns casos, venda de ativos para empresas privadas (Warner; Bel, 2008). As razões que motivaram esse processo incluem pressões fiscais, necessidade de investimentos em infraestrutura, busca por eficiência, modernização dos sistemas de distribuição em razão de crises ou falhas no serviço público (Wallsten; Kosec, 2008b; Wait; Petrie, 2017; Hanna; McDonald, 2021).

A privatização levou a certa redução de custos operacionais, porém frequentemente resultou em aumento de tarifas (Wait; Petrie, 2017; Onda; Tewari, 2021; Hanna; McDonald, 2021; Warner, 2021). Em média notou-se que clientes de sistemas privados pagam cerca de US\$ 20 a mais por mês (Onda; Tewari, 2021).

Ao comparar municípios que privatizaram com outros semelhantes que não o fizeram, antes e depois da privatização, foi identificado que o aumento nas tarifas de água após a privatização não veio acompanhado de ganhos em eficiência ou qualidade de serviço (Hanna; McDonald, 2021), e que os benefícios da redução de custo são limitados e frequentemente não repassados aos consumidores (Warner, 2021).

Nos casos em que não houve licitação ou concorrência robusta, as privatizações se deram por meio de negociações diretas, o que levanta preocupações acerca dos incentivos econômicos e transparência (Warner; Bel, 2008). O controle público e a transparência foram bastante prejudicados, o que fez com que as populações vulneráveis, em particular, enfrentassem riscos de exclusão e dificuldade de participação (Jepson; Lee Brown, 2014).

A oposição pública decorrente de experiências negativas com empresas privadas, como

aumento de tarifas, problemas de transparência, ou quebra de confiança pública levou a algumas cidades a retomarem o controle público do saneamento (Warner, 2024).

1.5.8 Síntese de experiências internacionais

O quadro 1.1 resume os principais resultados dos casos de privatização do saneamento nos países supracitados. Nele, pode-se notar que há modelos diversos, no entanto os resultados e tendências caminham em direções semelhantes.

Quadro 1.1 – Síntese de experiências internacionais de privatização do saneamento.

País	Modelo	Principais resultados	Tendência atual
Reino Unido	Venda total de ativos e monopólios privados	Melhora modesta na eficiência; tarifas altas; regulação intrusiva	Regulação crescente; críticas persistem
Chile	Venda e concessões operacionais	Aumento de produtividade (privadas); desigualdade no acesso	Modelo mantido, mas com críticas sociais
França	Concessões e arrendamentos locais	Tarifas mais altas com privados; qualidade variável	Reestatizações pontuais (Paris, Grenoble)
Alemanha	Modelos locais com PPPs e controle público	Tarifas mais altas em privadas; insatisfação pública	Retorno à gestão pública em alguns casos
Argentina	Concessões com metas e venda parcial	Cobertura e saúde melhoraram; metas descumpridas	Várias reestatizações após crise
Bolívia	Concessões com metas de cobertura	Sucesso em algumas cidades; protestos em outras	Cancelamentos e reestatizações (ex: Cochabamba)
EUA	Contratos operacionais e venda de ativos	Tarifas mais altas; pouca melhoria na eficiência	Cresce a remunicipalização

Como afirma Lazzarini (2023, p. 51), em vez de escolher a dedo um argumento teórico e defender de forma genérica a atuação de empresas públicas ou privadas, o exercício adequado é analisar a fundo os prós e contras de cada alternativa e propor maneiras de lidar com seus dilemas inerentes.

Nesse sentido, as experiências internacionais de privatização, como as do Reino Unido, Chile e Argentina, mostram que a participação privada pode trazer melhorias em eficiência e cobertura, mas também destacam que, sem transparência e regulação robusta, podem surgir riscos sociais significativos que levaram inclusive à remunicipalização em diversos casos. A regulação eficaz é essencial não apenas para evitar aumentos abusivos de tarifas, mas também para garantir a qualidade dos serviços e a inclusão de populações vulneráveis.

1.6 Panorama do saneamento básico no Brasil

Historicamente, o setor de água e saneamento no Brasil teve o governo como principal fonte de financiamento e prestador de serviços (Dassan *et al.*, 2024). O Plano Nacional de Saneamento (PLANASA), dos anos 1970, criou uma companhia de saneamento para cada estado da federação, as Companhias Estaduais de Saneamento Básico (CESBs), com base nas ideias de economia de escala e eficiência, segundo as quais partes mais ricas dos estados subsidiariam o saneamento das mais pobres, um subsídio cruzado (Kresch, 2020). No entanto, o PLANASA agravou as assimetrias de acesso a água e saneamento devido à sua abordagem orientada a mercado (Heller, 2007): a decisão das cidades mais ricas de criarem ou manterem a oferta do serviço a cargo de empresas municipais prejudicou o objetivo de subsídios cruzados entre cidades dentro do estado (Kresch, 2020). Verificou-se ainda que o Plano preteriu o esgotamento sanitário em favor da água e não aumentou a cobertura nos municípios mais ricos nem nos menores, por razões de viabilidade econômico-financeira (Heller, 2007).

Na década de 1990, a crise fiscal do Brasil e a aprovação da Lei de Concessões nº 8.987/1995, que trata da concessão e permissão da prestação de serviços públicos pela iniciativa privada, permitiu o investimento privado nos setores de infraestrutura e seguiu-se uma série de privatizações de serviços de água e esgoto (Dassan *et al.*, 2024; Saiani; Azevedo, 2018).

No final dos anos 1990, o estado do Paraná vendeu parte da sua CESB para a iniciativa privada e os resultados em termos de acesso da população foram inócuos (Clarke; Kosec; Wallsten, 2009). O acesso melhorou tanto das cidades que privatizaram como nas que se mantiveram públicas, mas o aumento foi maior no segundo grupo (Clarke; Kosec; Wallsten, 2009).

Já em análise nacional das privatizações desse mesmo período, notou-se diferenças entre os tipos de empresas resultantes, sendo que as do tipo privada local apresentaram melhora de qualidade do serviço (medida em morbidade e mortalidade decorrente de doenças veiculadas por água) enquanto que as híbridas, sociedades de economia mista, não apresentaram melhora nem piora (Saiani; Azevedo, 2018).

Com dados entre 2001 e 2003, observou-se aumento robusto no acesso ao saneamento decorrente da provisão privada, com destaque para os municípios mais pobres (Oliveira, 2008). As evidências com relação às tarifas, nesse caso, não são tão significativas mas sugerem que os serviços em municípios que privatizaram são relativamente mais baratos (Oliveira, 2008).

Em 2007, a Lei 11.445 que ficou conhecida como o Marco Legal do Saneamento Básico, permitiu contratos de concessão no setor de água e saneamento com os processos de licitação realizados pelas agências reguladoras de nível regional ou local (Dassan *et al.*, 2024). Além disso, dirimiu as controvérsias sobre titularidade herdadas do PLANASA, na medida em que estabeleceu claramente as competências dos municípios de serem a autoridade concedente, podendo eles cederem aos estados (Kresch, 2020).

As privatizações ocorridas no Brasil em decorrência da reforma legal de 2007 foram

analisadas por Dassan *et al.* (2024), que encontraram queda absoluta nas tarifas dos municípios com serviço privado em comparação aos que permaneceram públicos, embora os reajustes tarifários tenham sido mais elevados nos municípios privatizados ao longo dos primeiros quatro anos. Além disso, verificou-se aumento do acesso em cobertura de água, coleta e tratamento de esgoto.

Outro aspecto notável do saneamento brasileiro é a significativa heterogeneidade na eficiência dos serviços prestados entre os municípios, com média nacional abaixo de 50% (Tourinho *et al.*, 2022). Municípios com prestadores locais públicos tendem a apresentar maior eficiência do que aqueles atendidos por companhias estaduais ou privadas regionais (Tourinho *et al.*, 2022).

Em relação à estrutura tarifária, o modelo predominante é a combinação entre tarifas fixas mínimas e blocos crescentes, o que garante estabilidade financeira às empresas, mas pode implicar em cobranças desproporcionais às famílias de baixa renda (Silva; Ledo, 2023b). Em muitas localidades, a prática de subsidiar o serviço de esgoto com as receitas da água agrava as desigualdades, beneficiando principalmente as populações já atendidas por ambos os serviços (Silva; Ledo, 2023b).

Por fim, a questão regulatória carece de maior uniformidade e capacidade técnica. Segundo o Instituto Trata Brasil (2021), há grande disparidade entre as agências reguladoras locais, e muitas não realizam revisões tarifárias adequadas nem acompanham indicadores de universalização. Embora o NMLSB tenha atribuído à ANA a responsabilidade de editar normas de referência, a efetividade dessas diretrizes depende da adesão voluntária das agências estaduais ou municipais e de sua capacidade institucional (Instituto Trata Brasil, 2021).

1.6.1 O Novo Marco Legal do Saneamento Básico

Com o objetivo de atingir uma cobertura de 99% para o abastecimento de água potável e 90% para coleta e tratamento de esgoto até 31 de dezembro de 2033, o NMLSB, Lei 14.026, foi sancionado em julho de 2020 (BRASIL, 2020). A lei atualizou o arcabouço jurídico do setor com incentivos ainda maiores para participação do setor privado e regionalização, para que se torne mais atrativo financeiramente (Narzetti; Marques, 2021).

Além das metas de cobertura, o NMLSB estabelece metas progressivas de redução de perdas de água, regulamentadas pela Portaria MCID nº 788, de 1º de agosto de 2024 (BRASIL. Ministério das Cidades, 2024a). Para comprovação do cumprimento do índice de perda de água na distribuição, os valores dos indicadores devem ser menores ou iguais a: (i) 35% e 303,0 litros/ligação/dia, até 2025; (ii) 30% e 263,0 litros/ligação/dia para os anos de 2026 a 2032; e (iii) 25% e 216,0 litros/ligação/dia a partir do ano de 2033 (BRASIL. Ministério das Cidades, 2024a).

Essas metas de eficiência operacional são fundamentais para a sustentabilidade dos sistemas, considerando que, conforme dados do SINISA de 2024, as perdas totais de água na distribuição

alcançam 39,5% em nível nacional, valor significativamente acima das metas estabelecidas.

Com base na lei de 2007, os municípios delegavam os serviços de saneamento ao Estado-membro via convênio de cooperação, que em seguida utilizava-se de uma empresa pública ou sociedade de economia mista estadual, por meio de um contrato de programa com o município (Leite; Moita Neto; Bezerra, 2022).

O Artigo 10 do NMLSB, em particular, inova ao vedar a prestação do serviço de saneamento básico sob as formas de “contrato de programa, convênio, termo de parceria ou outros instrumentos de natureza precária” (BRASIL, 2020). Como muitos municípios não possuem contrato assinado e formalizado, a nova lei reforça a contratualização nos contratos de concessão e na regulamentação (Narzetti; Marques, 2022).

Isso significa dizer que a legislação eliminou a possibilidade de celebração de contratos de programa sem licitação. A prestação de serviços de saneamento básico por entidade que não integre a administração do titular passou a depender obrigatoriamente de concessão mediante prévia licitação, nos termos da Lei nº 8.987/1995 (BRASIL, 2020). Não é mais possível realizar novas concessões aos Estados-membros nem celebrar contrato de programa com os entes estaduais ou empresas estatais de outra unidade federativa sem processo competitivo (Leite; Moita Neto; Bezerra, 2022), o que reforça a concorrência e a transparência nesse mercado.

Os processos licitatórios para concessão dos serviços têm adotado diferentes critérios de seleção, sendo os mais comuns: (i) maior valor de outorga a ser pago ao poder concedente; e (ii) menor tarifa a ser cobrada dos usuários. A escolha do critério tem implicações importantes sobre a acessibilidade dos serviços. Nos leilões de Alagoas e Rio de Janeiro, por exemplo, o critério de maior outorga resultou em contratos que estabeleceram limites arbitrários ao número de residências com acesso à tarifa social, contribuindo para a elevação das tarifas e impactando negativamente o orçamento das famílias de baixa renda (Fagundes; Cunha Marques; Malheiros, 2024). Esse modelo de seleção, embora maximize a receita imediata do poder concedente, pode criar incentivos contrários à universalização se não houver salvaguardas adequadas de acessibilidade e inclusão social nos editais (Lazzarini, 2023).

A busca por ganhos de escala levou à atualização do conceito de prestação regionalizada dos serviços: “modalidade de prestação integrada de um ou mais componentes dos serviços públicos de saneamento básico em determinada região cujo território abranja mais de um Município” (BRASIL, 2020), que pode ser estruturada em região metropolitana, unidade regional de saneamento básico ou bloco de referência. Em muitos casos não é o município que decide a concessão, mas sim a companhia estadual, o modelo de regionalização transfere essa decisão para o estado.

Outra mudança significativa se deu sobre a regulação do setor. A situação prévia era a de que cada titular do serviço, no caso os municípios, poderia instituir órgão regulador próprio, sem observância de normas gerais (Leite; Moita Neto; Bezerra, 2022). O artigo 3º da Lei 14.026 estabelece a competência da ANA de “instituir normas de referência para a regulação dos serviços

públicos de saneamento básico” (BRASIL, 2020), entre as quais estão a regulação tarifária e governança das entidades reguladoras.

Considerando que tal mudança legislativa era bastante esperada, o número de concessões aumentou de forma significativa após a sanção do NMLSB. Antes de 2020, 284 municípios (5.10% do total) concederam algum ou todos os serviços de saneamento à iniciativa privada. Os que fizeram o mesmo partir de 2020 foram 563 municípios (10.11%). No entanto, a maior parte das cidades brasileiras 4723 (84.79%) segue sem qualquer tipo de concessão (ABCON SINDCON, 2024).

A situação em 2024 dos municípios por tipo de prestador pode ser visualizada na Figura 1.2. Os municípios em azul realizaram concessões anteriores ao NMLSB, enquanto os em verde concederam a partir de 2020; os municípios em cinza permanecem sem concessão privada. A distribuição espacial revela dois padrões distintos. As concessões anteriores a 2020 (azul) estão dispersas por todo o território nacional, com presença relevante no Centro-Oeste e no interior do Nordeste. Já as concessões pós-NMLSB (verde) apresentam padrão mais concentrado geograficamente: destacam-se o Amapá, onde praticamente todo o estado foi concedido, o Rio Grande do Sul, com grande mancha contínua no sul do estado, Mato Grosso do Sul no Centro-Oeste, e municípios pontuais no litoral nordestino, correspondentes aos blocos de Alagoas, e do Rio de Janeiro. Esse padrão é coerente com a estratégia de regionalização induzida pelo NMLSB, que agrupa municípios em blocos para viabilizar economias de escala (ABCON SINDCON, 2024).

Destaca-se nesse cenário a regionalização das concessões, uma vez que todas as concessionárias que assumiram os municípios a partir de 2020 pertencem a holdings que contém operações em outras cidades (ABCON SINDCON, 2024).

O NMLSB representa um avanço ao incentivar a participação privada e a regionalização, mas também levanta preocupações quanto à acessibilidade e equidade, especialmente para populações de baixa renda. Lazzarini (2023, p. 24) observa que a privatização pode ser especialmente problemática quando operadores privados têm grandes lucros potenciais ao ignorar dimensões sociais ou públicas difíceis de especificar em contrato.

Há que se levar em consideração, portanto, que provedores privados podem se concentrar nas populações de alta renda que conseguem pagar não só tarifas variáveis pela água consumida, mas também tarifas fixas para cobrir investimentos iniciais em infraestrutura, deixando as áreas mais carentes sem acesso adequado ou forçando-as a buscar fontes alternativas menos seguras (Lazzarini, 2023, p. 58).

Em análise dos primeiros efeitos do NMLSB, foi encontrado que a concessão resultante de leilões com base em critério de maior preço oferecido pelas competidoras prejudicou os mais pobres (Fagundes; Cunha Marques; Malheiros, 2024). Os casos analisados foram de Alagoas e Rio de Janeiro, cujos contratos previram limite arbitrário de residências com acesso à tarifa social, o que contribuiu para a elevação das tarifas e consequente impacto negativo sobre o orçamento das

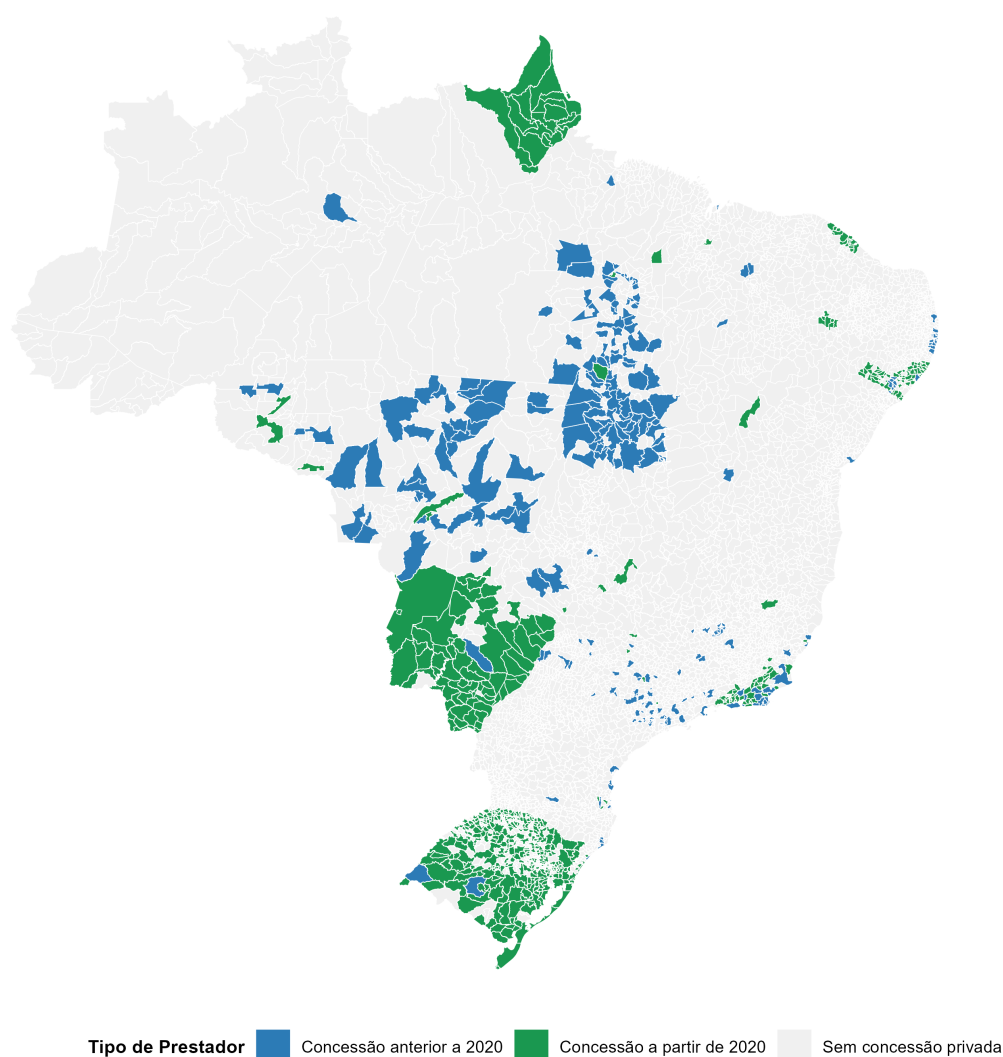


Figura 1.2 – Distribuição municipal das concessões privadas de saneamento básico no Brasil, segundo o período de concessão (anterior ou posterior ao NMLSB). Fonte: elaboração própria.

famílias (Fagundes; Cunha Marques; Malheiros, 2024). Fagundes, Cunha Marques e Malheiros (2024) defendem que análises obrigatórias de acessibilidade sejam incorporadas antes e depois dos leilões, evitando que o modelo de concessão se torne financeiramente insustentável em regiões de baixa renda.

1.6.2 Características nacionais do saneamento

O Sistema Nacional de Informações sobre Saneamento (SNIS), criado em 1995, foi a principal base de dados do setor no Brasil, coletando e divulgando informações sobre abastecimento de água, esgotamento sanitário, resíduos sólidos e drenagem urbana (BRASIL. Ministério das Cidades, 2024b). Ele serviu como ferramenta essencial para monitorar a cobertura e a qualidade dos serviços de saneamento, utilizando dados populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE) para calcular indicadores de atendimento (BRASIL. Ministério das Cidades, 2024b).

Em 2024, o SNIS foi substituído pelo Sistema Nacional de Informações em Saneamento Básico (SINISA), instituído pelo NMLSB. O SINISA moderniza a coleta de dados, separando informações de água e esgoto, incluindo detalhes sobre populações urbana e rural, e introduzindo novos módulos, como o de Gestão Municipal, que avalia a estrutura administrativa dos municípios (BRASIL. Ministério das Cidades, 2024b).

Além de desagregar informações por área urbana e rural, o sistema introduz mecanismos de auditoria e certificação de dados, como a metodologia ACERTAR, garantindo maior confiabilidade nas estatísticas (BRASIL. Ministério das Cidades, 2024b). Os dados a seguir são provenientes dos resultados do SINISA 2025, com informações coletadas sobre o ano de 2024 (Brasil. Ministério das Cidades, 2025a; Brasil. Ministério das Cidades, 2025b).

Tabela 1.1 – Informações globais do sistema de saneamento no Brasil em 2024 (%)

Informação	Indicador	Valor
Atendimento da população total com rede de abastecimento de água	IAG0001	84,10
Atendimento da população urbana com rede de abastecimento de água	IAG0002	92,30
Atendimento da população rural com rede de abastecimento de água	IAG0003	22,90
Atendimento dos domicílios totais com rede de abastecimento de água	IAG0004	80,70
Atendimento dos domicílios urbanos com rede de abastecimento de água	IAG0005	88,30
Atendimento dos domicílios rurais com rede de abastecimento de água	IAG0006	23,10
Perdas de faturamento de água	IAG2012	39,10
Perdas totais de água na distribuição	IAG2013	39,50
Atendimento da população total com rede coletora de esgoto	IES0001	62,30
Atendimento da população urbana com rede coletora de esgoto	IES0002	68,40
Atendimento da população rural com rede coletora de esgoto	IES0003	4,80
Atendimento dos domicílios totais com rede coletora de esgoto	IES0004	55,40
Atendimento dos domicílios urbanos com rede coletora de esgoto	IES0005	60,90
Atendimento dos domicílios rurais com rede coletora de esgoto	IES0006	3,50
Esgoto coletado referido à água consumida	IES2002	61,80
Esgoto tratado referido à água consumida	IES2003	51,80
Esgoto tratado referido ao esgoto coletado	IES2004	85,20

Fonte: SINISA 2025.

A Tabela 1.1 apresenta os principais indicadores nacionais de cobertura e eficiência dos serviços de saneamento no Brasil, expressos em porcentagem. Os dados revelam uma disparidade significativa entre áreas urbanas e rurais. Por exemplo, enquanto 92,3% da população urbana é atendida com rede de abastecimento de água, apenas 22,9% da população rural tem acesso a esse serviço.

A cobertura de esgoto é ainda mais desigual: apenas 4,8% da população rural é atendida, contra 68,4% da urbana. Além disso, os indicadores mostram níveis elevados de perdas de água: 39,5% de perdas totais na distribuição e 39,1% de perdas de faturamento. Quanto ao tratamento de esgoto, 61,8% do esgoto coletado referem-se à água consumida, sendo que 51,8% são efetivamente tratados, o que representa 85,2% do esgoto coletado.

1.6.3 Características regionais do saneamento

A Tabela 1.2 mostra a distribuição dos prestadores de serviços de saneamento conforme sua participação no sistema SINISA e sua abrangência territorial. O Brasil conta com centenas de prestadores locais e algumas dezenas de prestadores regionais. No serviço de água, há 1.249 prestadores locais participantes do SINISA, além de 54 regionais. Já no esgotamento sanitário, são 1.201 locais e 45 regionais. Há também prestadores que não participam do sistema, evidenciando a diversidade e fragmentação institucional do setor.

Tabela 1.2 – Prestadores de serviços segundo a abrangência.

Participantes do SINISA-Água	1.249 prestadores locais 54 prestadores regionais
Não participantes do SINISA-Água	172 prestadores locais 2 prestadores regionais
Participantes do SINISA-Esgoto	1.201 prestadores locais 45 prestadores regionais
Não participantes do SINISA-Esgoto	154 prestadores locais 1 prestador regional

Fonte: SINISA 2025.

Na Tabela 1.3, observa-se que a maioria dos prestadores de serviços são órgãos da administração direta ou autarquias. Para o serviço de água, predominam a administração direta (643) e as autarquias (458), com menor presença de empresas privadas (148), sociedades de economia mista (28) e empresas públicas (5). Para esgoto, o padrão se repete, com ainda maior concentração na administração direta (770) e autarquias (327).

Tabela 1.3 – Natureza jurídica dos prestadores de serviços.

Água	643 administração direta 458 autarquia 5 empresa pública 28 sociedade de economia mista 148 empresa privada 21 associação privada
Esgoto	770 administração direta 327 autarquia 6 empresa pública 25 sociedade de economia mista 116 empresa privada 2 associação privada

Fonte: SINISA 2025.

O cenário brasileiro do saneamento é, portanto, marcado pela coexistência de diferentes modelos de provisão. Segundo Lazzarini (2023, p. 116), na prática o que frequentemente se observa é a coexistência de formas públicas, privadas e híbridas de entrega de serviços. Essa diversidade de arranjos institucionais reflete não apenas diferenças regionais, capacidades fiscais e escolhas políticas locais, mas também o reconhecimento de que diferentes contextos podem demandar soluções distintas.

A fragmentação evidenciada pelos dados do SINISA, com mais de 1.200 prestadores locais e dezenas de prestadores regionais, ilustra essa heterogeneidade institucional. Nos casos em que os governos decidem manter as estatais, ter um concorrente do setor privado pode melhorar o desempenho dos gestores públicos, uma vez que a competição pode funcionar como um mecanismo indutor de eficiência e redução das falhas de governo (Lazzarini, 2023, p. 117).

1.6.3.1 Índices de atendimento

A Tabela 1.4 detalha os índices de atendimento por região. A região Sudeste possui os melhores indicadores tanto para água quanto para esgoto, enquanto a região Norte apresenta os piores índices, especialmente no atendimento total com rede de coleta de esgoto (24,7%) e no tratamento de esgoto como percentual da água consumida (22,5%). No geral, o tratamento do esgoto coletado é mais eficiente (acima de 81% em todas as regiões), mas a coleta ainda é limitada em várias partes do país.

Tabela 1.4 – Índices de atendimento de água e esgoto no Brasil.

Região	Atendimento com rede (%)				Tratamento dos esgotos (%)	
	Água		Coleta de esgoto		% da água consumida	% do esgoto coletado
	Total	Urbano	Total	Urbano		
	IAG0001	IAG0002	IES0001	IES0002	IES2003	IES2004
Norte	62,8	74,9	24,7	27,7	22,5	90,6
Nordeste	73,7	85,3	35,9	43,5	34,6	81,4
Centro-Oeste	89,7	97,0	68,3	73,3	54,7	95,6
Sudeste	92,1	96,2	81,9	85,8	62,3	82,8
Sul	88,3	97,5	57,7	63,3	46,6	94,6

Fonte: SINISA 2025.

1.6.3.2 Perdas de água na distribuição

Na Tabela 1.5, evidencia-se o alto índice de perdas de água em todas as regiões, especialmente no Norte (58,6% de perdas de faturamento e 49,4% de perdas totais) e Nordeste (45,5% e 46,7%, respectivamente). A menor perda de faturamento ocorre no Sul (34,5%), seguido do Sudeste (36,2%), refletindo maior eficiência no sistema. Contudo, mesmo nas regiões com melhor desempenho, as perdas continuam elevadas em padrões internacionais.

Tabela 1.5 – Perdas de água em sistemas de distribuição (%).

Região	Perdas de faturamento de água	Perdas totais de água na distribuição
	IAG2012	IAG2013
Norte	58,6	49,4
Nordeste	45,5	46,7
Centro-Oeste	35,0	32,4
Sudeste	36,2	37,6
Sul	34,5	36,1

Fonte: SINISA 2025.

1.6.3.3 Tarifas e despesas

A Tabela 1.6 compara receitas e despesas médias por metro cúbico de água e esgoto. O Sul apresenta as maiores tarifas de água (R\$7,12) seguido pelo Centro-Oeste (R\$6,62), enquanto o Nordeste tem as menores (R\$4,89). Para esgoto, o Centro-Oeste apresenta as maiores tarifas (R\$6,16), seguido pelo Sul (R\$5,90). A diferença entre receita e despesa varia regionalmente, refletindo desigualdades na sustentabilidade financeira dos prestadores. Em algumas regiões, como Norte e Centro-Oeste no serviço de água, as despesas superam as receitas, indicando desafios de sustentabilidade.

Tabela 1.6 – Receitas e despesas médias praticadas (R\$/m³).

Região	Água		Esgoto	
	Receita média	Despesa média	Receita média	Despesa média
	IFA1001	IFA2002	IFE1001	IFE2002
Norte	5,49	6,91	4,87	4,44
Nordeste	4,89	6,02	5,61	4,44
Centro-Oeste	6,62	6,70	6,16	5,77
Sudeste	5,54	6,19	5,36	4,01
Sul	7,12	6,26	5,90	4,98

Fonte: SINISA 2025.

1.7 Conclusões

O artigo apresentou uma análise da economia da água e do saneamento, destacando os desafios e as particularidades do setor no contexto brasileiro. A água, como recurso essencial e escasso, demanda uma gestão eficiente que equilibre aspectos econômicos, sociais e ambientais.

O estudo evidenciou que o mercado de água e saneamento é caracterizado como um monopólio natural, devido aos altos custos fixos, economias de escala e a necessidade de

infraestrutura extensiva e de longo prazo. Essas características justificam a intervenção estatal ou a regulação rigorosa para evitar distorções de preço e garantir o acesso universal.

No cenário brasileiro, o setor de saneamento enfrenta desafios significativos, como a disparidade na cobertura entre áreas urbanas e rurais, altas taxas de perda de água e a necessidade de investimentos em tratamento de esgoto. Além disso, a eficiência dos serviços varia bastante entre municípios, a regulação ainda é fraca e desigual entre agências, comprometendo a universalização e o controle de tarifas.

O Novo Marco Legal do Saneamento Básico (NMLSB) representa um avanço ao incentivar a participação privada e a regionalização, mas também levanta preocupações quanto à acessibilidade e equidade, especialmente para populações de baixa renda.

As experiências internacionais de privatização, como as do Reino Unido, Chile e Argentina, mostram que a participação privada pode trazer melhorias em eficiência e cobertura, mas também destacam que sem transparência pode haver riscos sociais, que levaram inclusive à remunicipalização. Nesse sentido, ressalta-se a importância de uma regulação robusta e transparente para evitar aumentos abusivos de tarifas e garantir a qualidade dos serviços.

No Brasil, os resultados das privatizações têm sido heterogêneos, com alguns casos de sucesso e outros que demandam ajustes para assegurar benefícios sociais. O NMLSB ampliou a participação privada e fortaleceu a regionalização, exigindo licitação e contratos formais. No entanto, efeitos iniciais mostram que os critérios de leilão (maior oferta financeira) podem penalizar os mais pobres se não houver salvaguardas de acessibilidade.

Em síntese, a universalização do saneamento básico no Brasil requer um modelo que combine eficiência econômica, sustentabilidade financeira e equidade social. A atuação conjunta do setor público e privado, aliada a uma regulação eficaz e à priorização de investimentos em regiões menos desenvolvidas, é essencial para superar os desafios históricos do setor. O NMLSB abre caminho para avanços, mas sua implementação deve ser acompanhada de mecanismos que garantam o acesso justo e a qualidade dos serviços.

2 Participação de empresas privadas no saneamento: revisão bibliométrica e agenda de pesquisa para o Brasil

2.1 Introdução

A participação de empresas privadas no saneamento é um tema de crescente relevância para a política pública brasileira. O NMLSB, aprovado em 2020, inaugurou um novo ciclo de concessões e parcerias público-privadas no setor, com meta de universalização do acesso à água e ao esgotamento sanitário até 2033 (BRASIL, 2020). Compreender o estado do conhecimento internacional sobre os efeitos dessa participação privada é, portanto, condição necessária tanto para orientar a agenda de pesquisa nacional quanto para informar as escolhas regulatórias que ainda estão em aberto.

A literatura científica sobre o tema é extensa, interdisciplinar e dispersa por diferentes tradições teóricas e empíricas. Trabalhos seminais como o de Galiani, Gertler e Schargrodsky (2005), que examinou a privatização da água na Argentina usando um desenho de diferenças-em-diferenças, e revisões abrangentes como a de Fabre e Straub (2023) sobre parcerias público-privadas em infraestrutura coexistem com uma vasta produção de estudos de eficiência, análises regulatórias e avaliações de impacto distribuídos por centenas de periódicos. Mapear esse campo de forma sistemática exige uma metodologia capaz de processar grandes volumes de dados bibliográficos sem perder a capacidade de identificar estruturas intelectuais e tendências emergentes.

A bibliometria oferece esse conjunto de instrumentos (Donthu *et al.*, 2021). Por meio dela é possível investigar colaborações entre autores, padrões de citação e estruturas conceituais em amostras de centenas ou milhares de documentos (Donthu *et al.*, 2021), o que é particularmente útil para a formulação de agendas de pesquisa em campos complexos e interdisciplinares (Durán-Sánchez *et al.*, 2020). Diferentes abordagens bibliométricas revelam dimensões complementares do campo: a análise de co-autoria mapeia as redes de colaboração; a análise de co-citação expõe a estrutura intelectual subjacente, mostrando quais obras fundamentais formam os alicerces teóricos da área; a análise de co-ocorrência de palavras-chave identifica temas centrais e emergentes; e a análise de evolução temática rastreia como os tópicos dominantes se transformam ao longo do tempo, evidenciando transições paradigmáticas e lacunas de pesquisa (Donthu *et al.*, 2021).

Assim, o presente estudo tem como objetivo realizar uma análise bibliométrica sobre a participação de empresas privadas no saneamento, com foco na identificação de tendências da área de pesquisa ao longo dos anos e nas oportunidades de pesquisa que possam ser desenvolvidas

para o Brasil. Os resultados apontam que, em termos teóricos, o campo está ancorado na teoria de incentivos e regulação e na teoria de contratos (Laffont; Tirole, 1993), enquanto em termos empíricos o foco recai sobre estudos econométricos de impacto. Para o Brasil especificamente, a agenda identificada aponta para sete eixos prioritários: estudos econométricos pós-NMLSB, análises de eficiência comparativa, investigação sobre escala e regionalização, regulação e captura regulatória, equidade tarifária, comportamento do consumidor, e determinantes e *design* da privatização.

2.2 Materiais e Métodos

São muitas as análises que um estudo bibliométrico pode realizar; neste trabalho, optou-se por uma análise descritiva seguida de uma análise de evolução temática e de redes de colaboração e citação. Donthu *et al.* (2021) indicam que a análise de “co-palavras”, como palavras notáveis mais frequentes nas implicações e agendas de pesquisa futuras dos textos completos, pode ser selecionada para caracterizar o futuro do campo em análise.

2.2.1 Fontes bibliográficas e estratégia de busca

Para capturar a literatura científica mais relevante sobre a participação privada no setor de água e saneamento, foram consultadas duas bases de dados: Scopus, mantida pela Elsevier, e Web of Science, pela Clarivate, ambas acessadas por meio do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A busca foi realizada em 23/11/2025.

A escolha dessas duas bases se justifica por serem as mais abrangentes e criteriosamente curadas no campo das ciências sociais aplicadas. A Scopus e a Web of Science apresentam cobertura complementar: enquanto a primeira tende a incluir maior volume de periódicos internacionais, a segunda tem reconhecida profundidade histórica e rigor na indexação de periódicos de alto impacto em economia (Aria; Cuccurullo, 2017). A utilização conjunta de ambas reduz o viés de cobertura que ocorreria com o uso de apenas uma fonte.

Procurou-se por artigos na área de economia, publicados em inglês ou português, sem restrição de intervalo de datas. Uma consulta foi elaborada para capturar uma ampla gama de trabalhos, pesquisando nos títulos, resumos e palavras-chave dos artigos. A estratégia de busca combina termos relativos ao setor privado (“private sector”, “privat*”) com termos referentes ao setor de saneamento (“water”, “wastewater”, “sewerage”, “sanitation”) e com termos que delimitam o escopo analítico econômico (regulação, impacto econômico, análise custo-benefício, saúde, impacto social e ambiental). Essa combinação visa equilibrar sensibilidade e especificidade da busca, priorizando artigos com abordagem econômica explícita.

Cabe ressaltar, contudo, que a estratégia de busca adotada introduz vieses inerentes à escolha das palavras-chave. A restrição ao campo da economia e aos idiomas inglês e português pode subrepresentar contribuições relevantes produzidas em outros idiomas ou publicadas em

periódicos interdisciplinares. Além disso, termos como “privat*” podem capturar estudos sobre privatização em outros setores além do saneamento, o que foi mitigado pelos termos de restrição setorial. Estudos futuros poderiam ampliar o escopo a outros campos do conhecimento e bases de dados adicionais, como o Google Scholar, para recuperar literatura cinzenta não indexada.

Na Scopus, a seguinte consulta resultou em 293 artigos:

(TITLE-ABS-KEY (“private sector” OR privat* AND (water OR wastewater OR sewerage OR sanitation) AND (regulat* OR “economic impact” OR “economic evaluation” OR “cost-benefit analysis” OR “economic policy” OR health OR “public health” OR “social impact” OR “environmental impact”))) AND (LIMIT-TO (SUBJAREA, “ECON”)) AND (LIMIT-TO (DOCTYPE, “ar”)) AND (LIMIT-TO (LANGUAGE, “English”) OR LIMIT-TO (LANGUAGE, “Portuguese”))

Por sua vez, a Web of Science forneceu 225 artigos com a seguinte consulta:

TS=((“private sector” OR privat*) AND (water OR wastewater OR sewerage OR sanitation) AND (regulat* OR “economic impact” OR “economic evaluation” OR “cost-benefit analysis” OR “economic policy” OR health OR “public health” OR “social impact” OR “environmental impact”)) AND WC=(Economics) AND DT=(Article) AND (LA=(English) OR LA=(Portuguese))

onde TS representa o tópico, ou seja, a busca pelos termos selecionados nos títulos, resumos e listas de palavras-chave.

2.2.2 Limpeza dos dados

Os metadados brutos foram baixados da Scopus (.csv) e da Web of Science (.txt). Esses conjuntos de dados contêm informações bibliográficas de cada artigo, como título, resumo, autores, DOI, palavras-chave, referências e citações. A união dos conjuntos resultou em 518 registros, dos quais 126 documentos duplicados foram eliminados, restando 392 artigos para análise bibliométrica.

Para a construção da análise de evolução temática, foi necessário realizar um pré-processamento dos resumos com a remoção de palavras funcionais sem conteúdo semântico relevante para o campo, procedimento conhecido na literatura de mineração de texto como filtragem por *stopwords* (Donthu *et al.*, 2021). Trata-se de termos de alta frequência que não contribuem para a identificação de temas substantivos, como artigos, preposições, conjunções e verbos auxiliares, cuja presença distorceria a análise de co-ocorrência de palavras ao inflar artificialmente a importância de termos genéricos em detrimento dos conceitos analiticamente relevantes.

A lista de *stopwords* utilizada foi definida a priori pelo pesquisador com base no vocabulário recorrente nos resumos da amostra e inclui os seguintes termos em língua inglesa: *the, and, of, to,*

in, a, is, for, that, this, with, as, are, on, by, from, at, an, be, or, has, have, was, were, been, their, which, can, these, not, but, more, than, such, study, results, using, also, between, through, both, may, used, paper, our, we, most, into, all, based, data, however, other, show, well, found, different, analysis. A inclusão de termos como “study”, “results”, “paper” e “analysis” justifica-se pelo caráter metodológico dessas palavras, que aparecem com alta frequência em resumos científicos independentemente do tema tratado e que, portanto, não discriminam entre os cortes temporais analisados.

2.2.3 Análise bibliométrica e visualização

Para manipulação, análise e visualização, utilizou-se o pacote *bibliometrix* (Aria; Cuccurullo, 2017) na linguagem *R* (R Core Team, 2024), uma poderosa ferramenta de código aberto para realizar análises bibliométricas, especialmente devido à sua interface web amigável, *biblioshiny*.

Esse pacote permite executar todas as etapas necessárias em pesquisas bibliométricas, incluindo importação e conversão de dados, fusão de diferentes bases, estatísticas descritivas e criação de redes para acoplamento bibliográfico, co-citação, colaboração e análise de coocorrência, além de visualizações como mapeamento de estrutura conceitual e de redes (Aria; Cuccurullo, 2017).

Os processos de importação, conversão e fusão foram realizados diretamente no software de código aberto *RStudio*, assim como as análises descritiva e gráfica, enquanto o conjunto de dados limpo foi importado para o *biblioshiny* para realizar a análise de evolução temática.

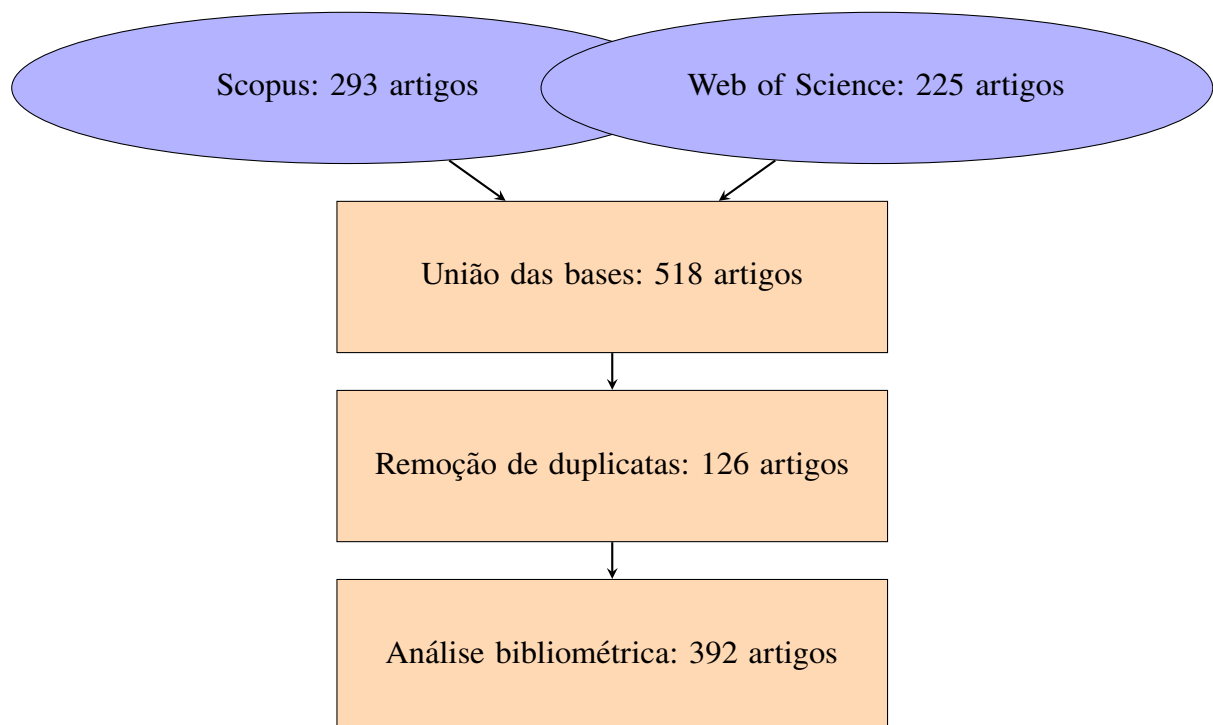


Figura 2.1 – Fluxograma do processo de revisão bibliométrica utilizando as bases de dados Scopus e Web of Science. Fonte: elaboração própria.

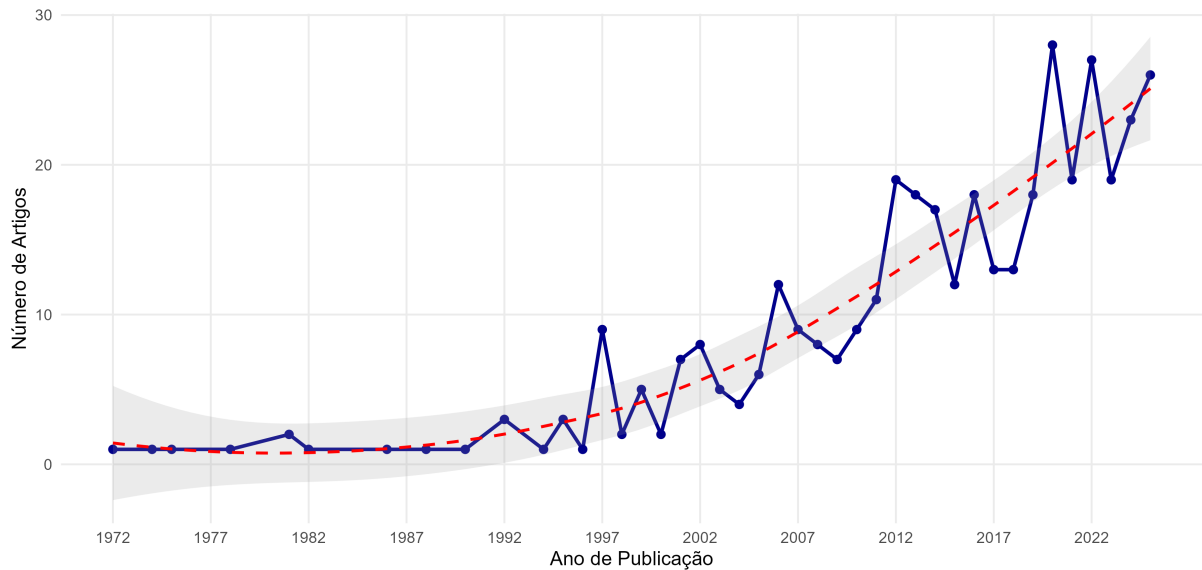


Figura 2.2 – Evolução temporal da produção científica sobre participação privada no saneamento, 1972–2025. A linha tracejada vermelha indica a tendência de crescimento. Fonte: elaboração própria.

A Figura 2.1 resume o fluxograma do processo de busca, limpeza e análise dos dados. O algoritmo que permite a reprodução desse estudo encontra-se na seção B.1 do Apêndice B.

A visualização da evolução temática ao longo dos cinco períodos foi realizada por meio de diagrama de Sankey, que representa fluxos direcionados entre nós, em que a espessura de cada fluxo é proporcional à intensidade da associação entre os temas em períodos consecutivos (Lim; Chou; Chen, 2025).

Conforme demonstrado por Lim, Chou e Chen (2025), os diagramas de Sankey superam as visualizações bibliométricas tradicionais em rede ao simplificar a leitura de conexões entre elementos, reduzindo a poluição visual (*visual clutter*) sem perda de informação substantiva sobre as relações dominantes entre clusters temáticos. No contexto da evolução temática, cada nó representa um tema-chave de um período e cada fluxo representa a continuidade ou transformação desse tema no período seguinte; temas que desaparecem ou surgem sem fluxo de continuidade indicam, respectivamente, declínio ou emergência temática no campo.

2.3 Resultados

2.3.1 Análise descritiva

A análise bibliométrica abrange o período de 1972 a 2025 e inclui um total de 392 documentos publicados em 215 periódicos distintos. A idade média dos documentos é de 11,1 anos, com uma média de 29,49 citações por documento e 2,66 citações por ano por documento, totalizando 12.534 referências no conjunto de dados. A taxa de crescimento anual da produção científica é de 6,34%, o que evidencia a expansão contínua do interesse acadêmico pelo tema.

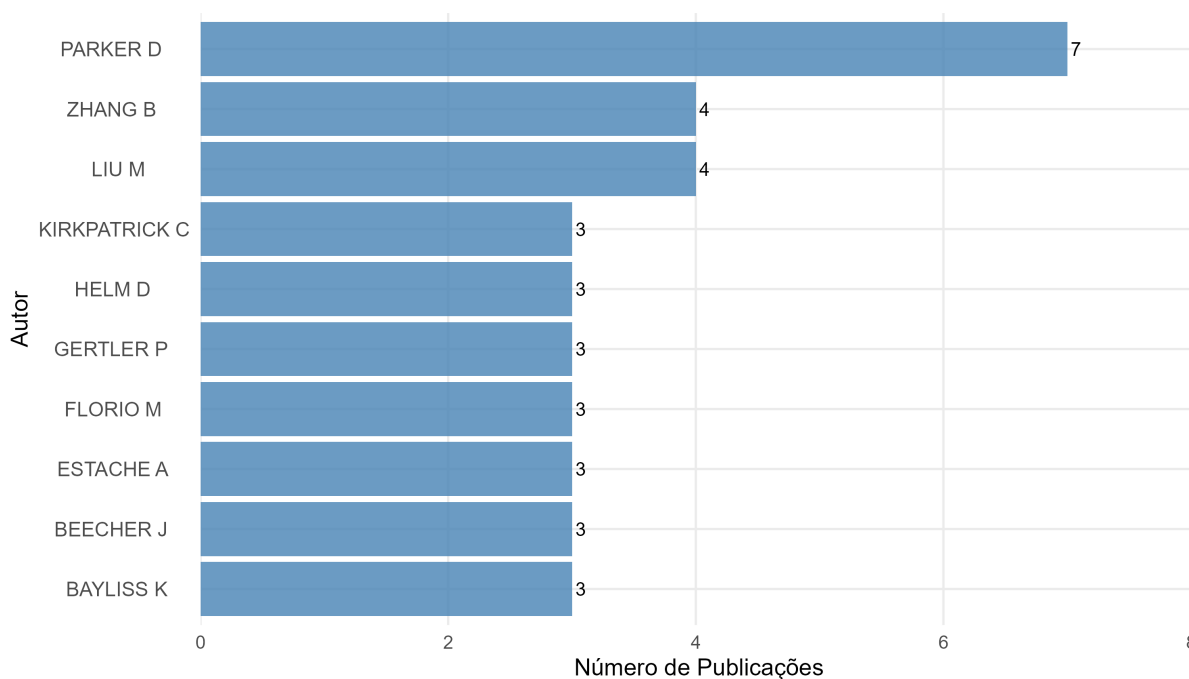


Figura 2.3 – Top 10 autores mais produtivos na literatura sobre participação privada no saneamento. Fonte: elaboração própria.

A Figura 2.2 ilustra a evolução temporal da produção científica. Nota-se produção esparsa até meados dos anos 1990, com aceleração expressiva a partir de 1997, coincidindo com as grandes ondas de privatização na América Latina, Europa e Ásia. A tendência de crescimento se intensifica após 2010 e atinge seus picos mais recentes entre 2020 e 2025, período que também coincide com as discussões e a aprovação do NMLSB no Brasil.

Um total de 944 autores contribuíram para as 392 publicações, fazendo 1.020 aparições de autores. Destes, 98 autores têm documentos de autoria única, e a média é de 2,6 co-autores por documento. Os 10 autores mais produtivos, ilustrados na Figura 2.3, foram: Parker D (n=7), Liu M (n=4), Zhang B (n=4), Bayliss K (n=3), Beecher J (n=3), Estache A (n=3), Florio M (n=3), Gertler P (n=3), Helm D (n=3) e Kirkpatrick C (n=3).

Os artigos mais citados na amostra foram: Polasky *et al.* (2011) (n=679), seguido por Galiani, Gertler e Schargrodsky (2005) (n=413), Cai, Chen e Gong (2016) (n=328), Domenech *et al.* (2019) (n=288), Liu, Shadbegian e Zhang (2017) (n=268), Havstad *et al.* (2007) (n=260), Sathirathai e Barbier (2001) (n=239), Hazilla e Kopp (1990) (n=215), Saal, Parker e Weyman-Jone (2007) (n=207) e Devoto *et al.* (2012) (n=201), conforme ilustrado na Figura 2.4.

Já as referências mais citadas pelos artigos da amostra, representadas na Figura 2.5, foram Galiani, Gertler e Schargrodsky (2005) (n=14), Megginson e Netter (2001) (n=12), McKenzie e Mookherjee (2003) (n=9), Estache e Rossi (2002) (n=9), além de referências relacionadas aos trabalhos de Vickers e Yarrow (1988), Teeple e Glyer (1987), Laffont e Tirole (2001), Feigenbaum e Teeple (1983a), Crain e Zardkoohi (1978a) e Byrnes, Grosskopf e Hayes (1986), todas com 8 citações cada. A recorrência de Galiani, Gertler e Schargrodsky (2005) tanto entre

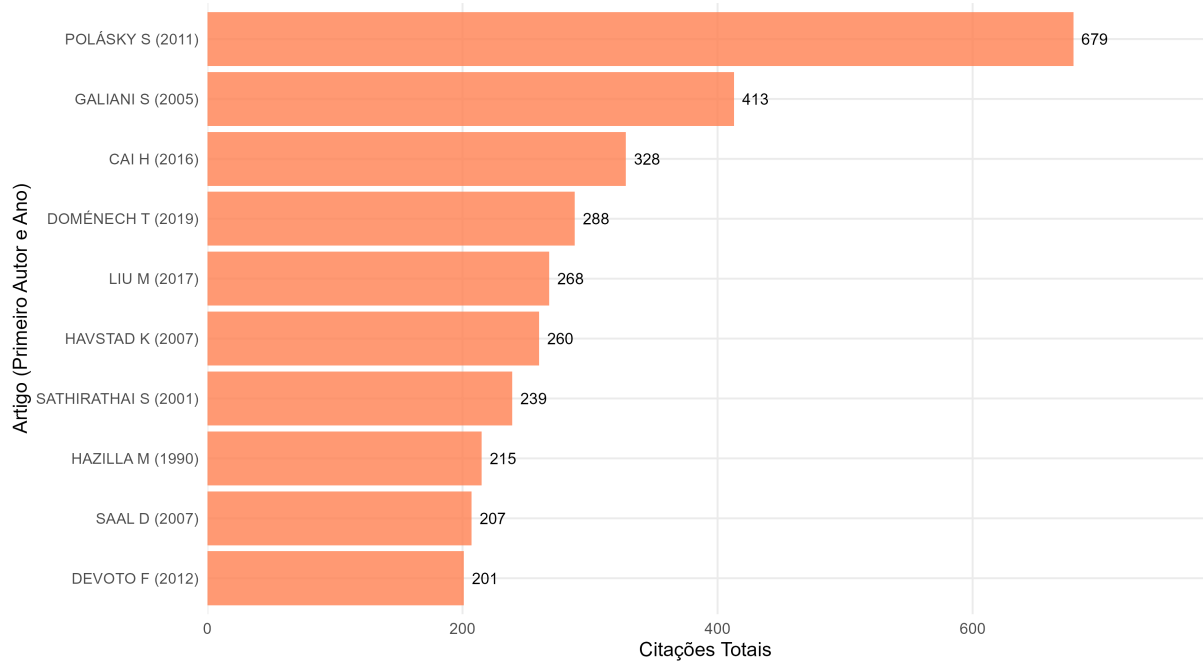


Figura 2.4 – Top 10 artigos mais citados na literatura sobre participação privada no saneamento. Fonte: elaboração própria.

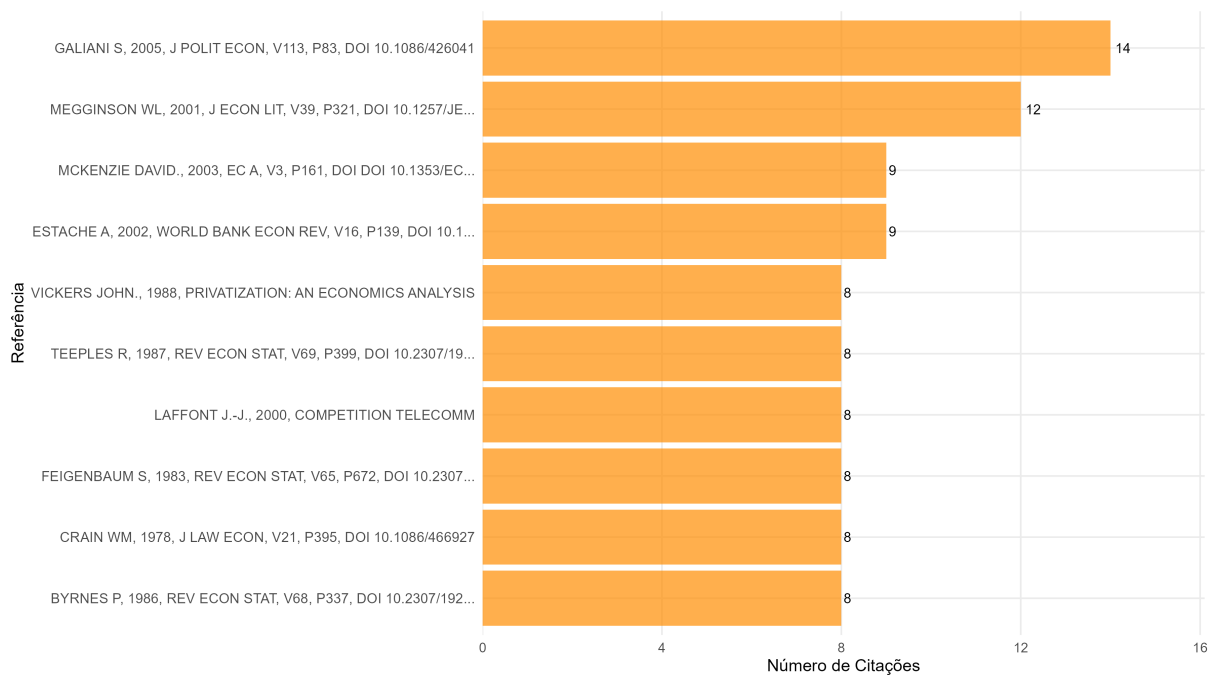


Figura 2.5 – Top 10 referências mais citadas pelos artigos da amostra. Fonte: elaboração própria.

os mais citados quanto entre as referências mais utilizadas reforça seu papel de estudo seminal no campo.

A distribuição dos artigos pelos países dos autores correspondentes, apresentada na Figura 2.6, é liderada pelos EUA (n=193), Reino Unido (n=60), França (n=44), China (n=31), Brasil (n=26) e Austrália (n=26), entre outros. O Brasil ocupa a quinta posição em produção, o

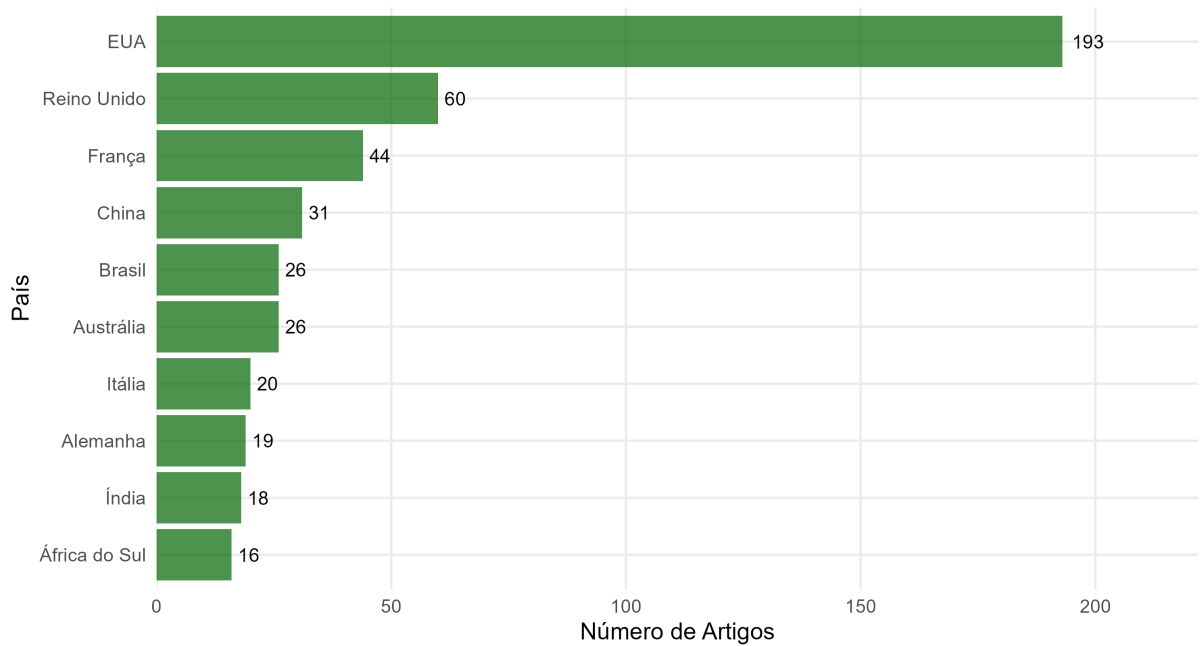


Figura 2.6 – Top 10 países mais produtivos na literatura sobre participação privada no saneamento, por país do autor correspondente. Fonte: elaboração própria.

que indica participação relevante no debate científico internacional, embora ainda distante dos primeiros colocados em termos absolutos.

A Figura 2.7 apresenta o número total de citações e a média de citações por artigo por país. Os EUA lideram em volume absoluto de citações, o que é reforçado pela presença de quatro instituições americanas entre as dez afiliações mais produtivas (Universidade da Califórnia, Banco Mundial, Universidade Duke e National Bureau of Economic Research), conforme ilustra a Figura 2.8.

Destaca-se, porém, que quando a métrica é a média de citações por artigo, países como Bélgica, Canadá, Bolívia, China, Itália e Alemanha se sobressaem, o que pode indicar a existência de artigos altamente influentes nesses contextos. O caso da Bolívia é ilustrativo: um único artigo amplamente citado sobre privatização da água em Cochabamba eleva exponencialmente a média do país.

A análise das palavras-chave mais frequentes, apresentada na Figura 2.9, revela que os termos mais recorrentes na amostra são “Water”, “Privatization”, “Sanitation”, “Regulation” e “Efficiency”, seguidos por “Infrastructure”, “Management”, “Health” e “Ownership”. A co-ocorrência frequente de “privatization” e “regulation” sugere que o campo trata o tema da propriedade não de forma isolada, mas sempre em relação ao marco regulatório que o governa. Adicionalmente, termos como “Sustainability”, “Climate Change” e “Groundwater” indicam que preocupações ambientais e de longo prazo integram cada vez mais as análises econômicas do setor.

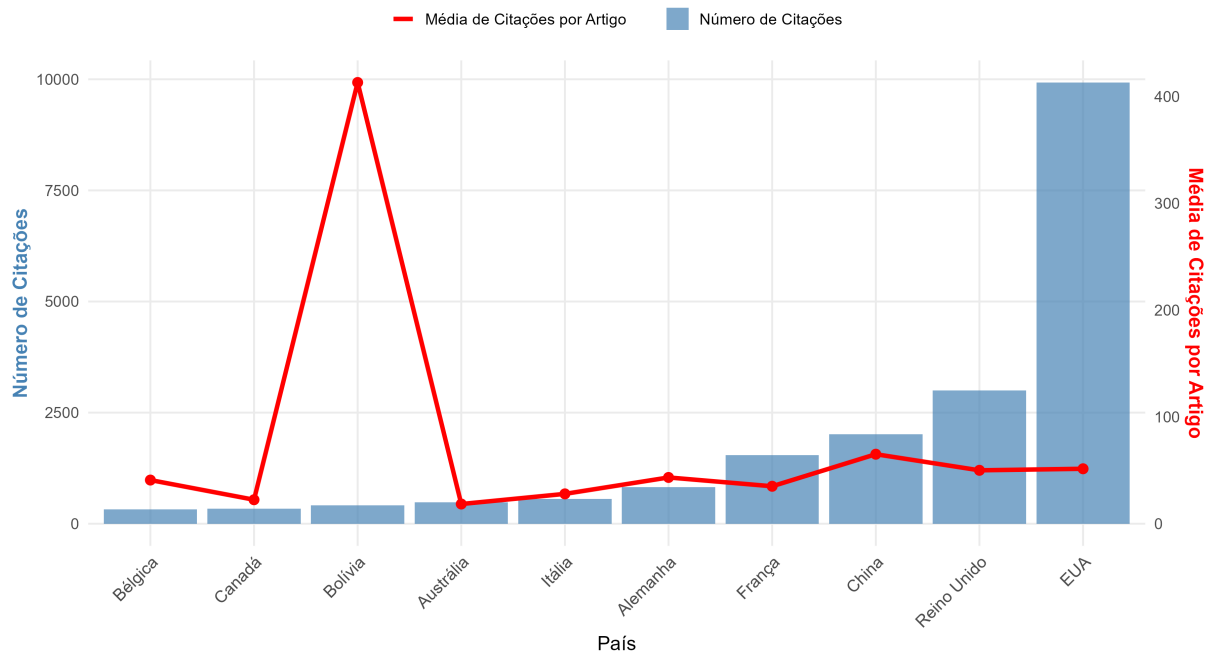


Figura 2.7 – Número de citações e média de citações por artigo por país. Fonte: elaboração própria.

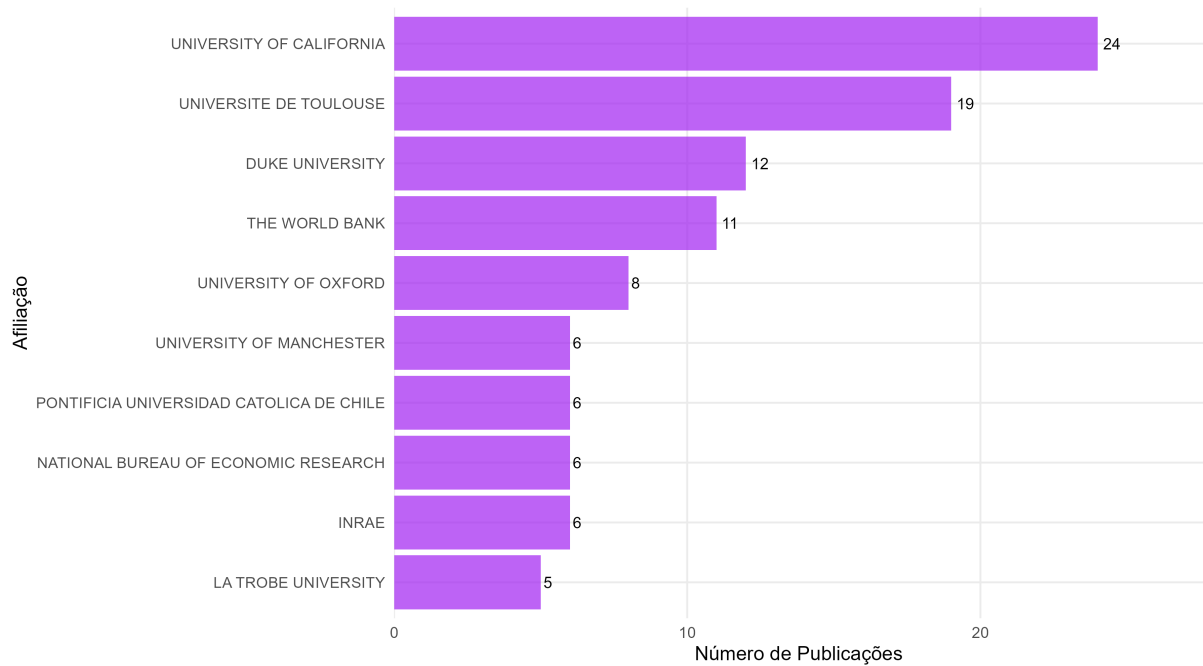


Figura 2.8 – Top 10 afiliações mais produtivas na literatura sobre participação privada no saneamento. Fonte: elaboração própria.

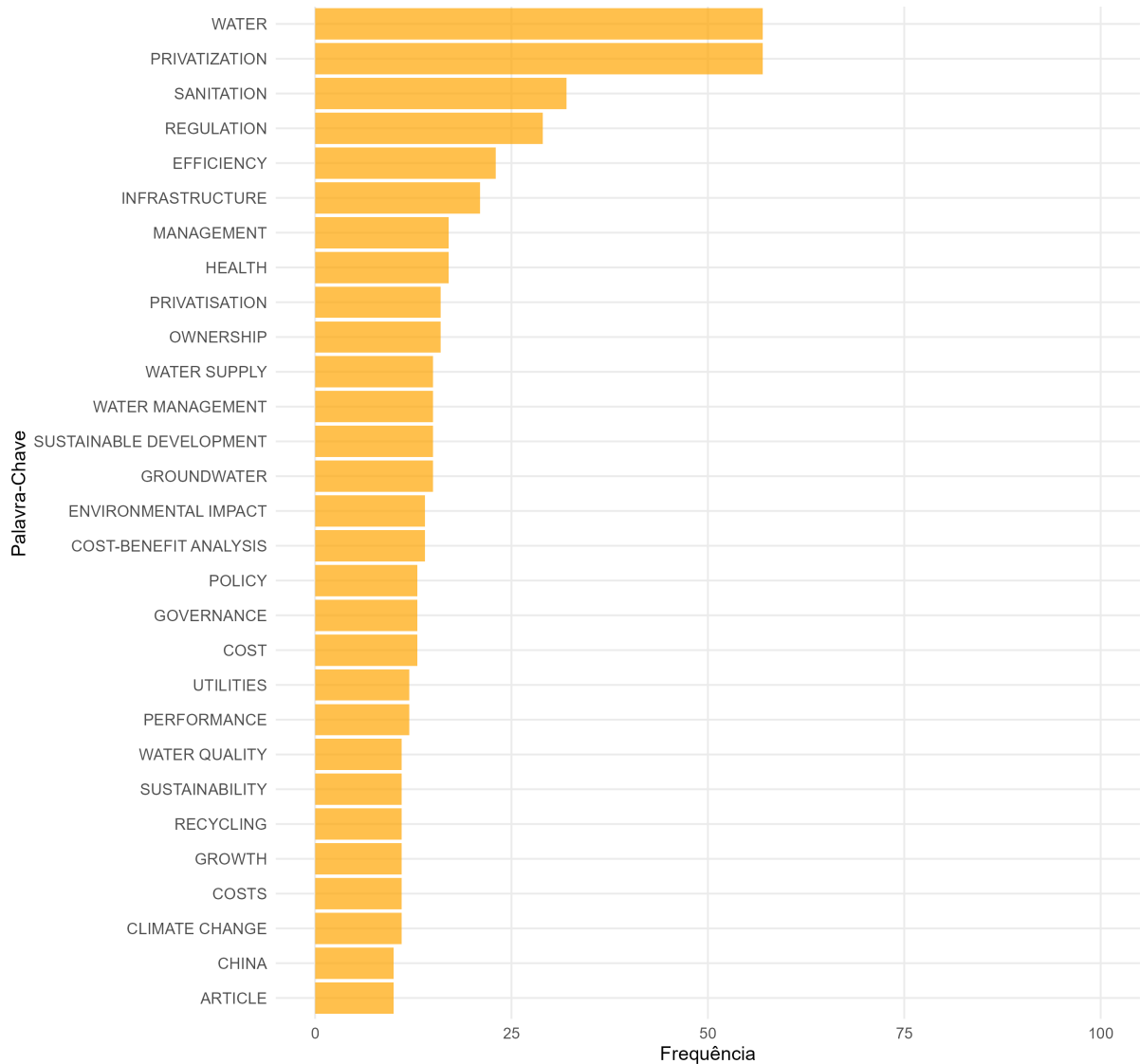


Figura 2.9 – Top 30 palavras-chave mais frequentes na literatura sobre participação privada no saneamento.
Fonte: elaboração própria.

2.3.2 Redes de Colaboração e Co-citação

A Figura 2.10 apresenta a rede de co-autoria entre os autores da amostra. A estrutura da rede é caracterizada por vários pequenos clusters isolados, sem um componente central que conecte os grupos entre si. Isso indica que, embora existam grupos coesos internamente, como o cluster formado por Parker D, Saal D e Kirkpatrick C, que são os autores com maior produção e figuram no maior nó da rede, a colaboração internacional entre grupos distintos ainda é limitada. O índice de co-autoria internacional é de 15,82%, o que, embora expressivo, sugere espaço para maior integração entre grupos de pesquisa de diferentes países.

O cluster mais proeminente (em vermelho) reúne Parker D, Kirkpatrick C e Saal D, pesquisadores britânicos que têm se dedicado ao estudo da eficiência e regulação de serviços de água e esgoto. Outro cluster notável (em azul escuro) reúne Liu M, Zhang B e Shadbeian

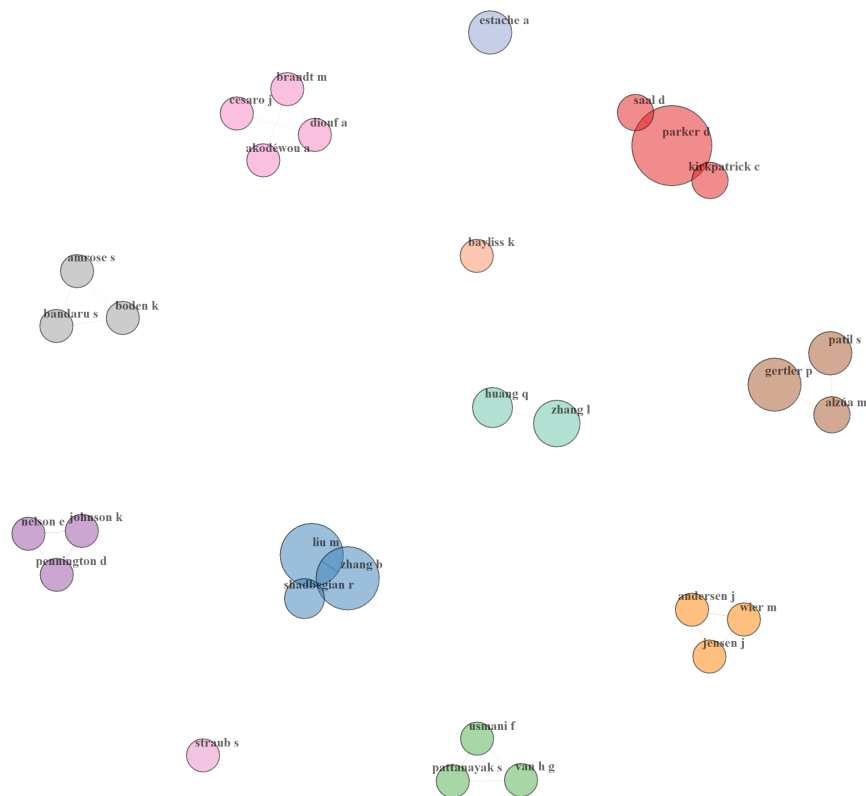


Figura 2.10 – Rede de co-autoria entre os autores da literatura sobre participação privada no saneamento. O tamanho dos nós é proporcional ao número de publicações do autor. Fonte: elaboração própria.

R, pesquisadores com foco em análise ambiental e de poluição na China. O isolamento relativo de Estache A, um dos autores mais citados no campo de infraestrutura e desenvolvimento, e de Bayliss K, pesquisadores com contribuições relevantes sobre privatização, evidencia que sua influência se dá mais por meio de seus textos do que por uma rede de co-autoria densa.

A rede de co-citação, apresentada na Figura 2.11, revela a estrutura intelectual do campo, ou seja, os pares de obras que tendem a ser citadas em conjunto e, portanto, formam núcleos teóricos e empíricos coesos. Identificam-se dois clusters principais. O primeiro e mais denso (em vermelho) agrupa obras seminais sobre privatização, regulação e eficiência em serviços públicos de água e saneamento: Galiani, Gertler e Schargrodsky (2005), Megginson e Netter (2001), Estache e Rossi (2002), McKenzie e Mookherjee (2003), Laffont e Tirole (2001), Guasch (2004), Wallsten e Kosec (2008a) e Ostrom (1990), entre outros. Esse cluster representa o núcleo central da literatura econômica sobre o tema e ancora a maior parte das discussões teóricas e empíricas da área.

O segundo cluster (em azul) é composto principalmente por obras de Sueyoshi e Goto (2012) e demais autores, voltadas para análise de eficiência por meio de Análise Envolvória de Dados ou *Data Envelopment Analysis* (DEA) em serviços de utilidade pública, particularmente

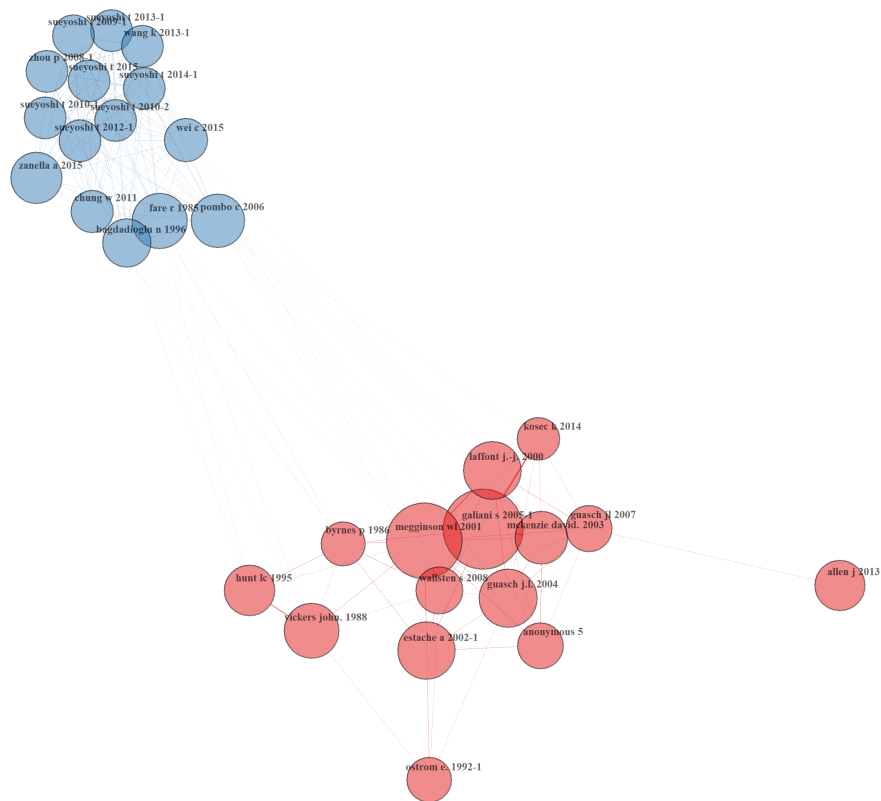


Figura 2.11 – Rede de co-citação das referências mais frequentes na literatura sobre participação privada no saneamento. O tamanho dos nós é proporcional ao número de co-citações. Fonte: elaboração própria.

no contexto japonês. Embora geograficamente e metodologicamente distinto do núcleo principal, esse cluster evidencia a relevância de análises de fronteira de eficiência no campo.

2.3.3 Evolução Temática

O banco de dados foi dividido em cinco períodos de tempo, e na Figura 2.12 estão apresentadas as palavras mais relevantes de cada período com base nos resumos dos artigos. O diagrama de Sankey, produzido via *biblioshiny*, permite visualizar não apenas quais temas dominam cada período, mas também como eles se comportam entre cortes temporais.

No primeiro período, de 1972 a 2007, os termos dominantes são “water”, “rights”, “management” e “regulation”, refletindo uma produção ainda incipiente, centrada nas questões jurídico-institucionais do acesso à água e nos primeiros debates sobre regulação econômica de serviços públicos.

O trabalho mais citado desse período foi Galiani, Gertler e Schargrotsky (2005), que explora a privatização de companhias de água na Argentina nos anos 1990 como um experimento natural para verificar seu efeito sobre a mortalidade infantil. Os autores obtêm evidências de que

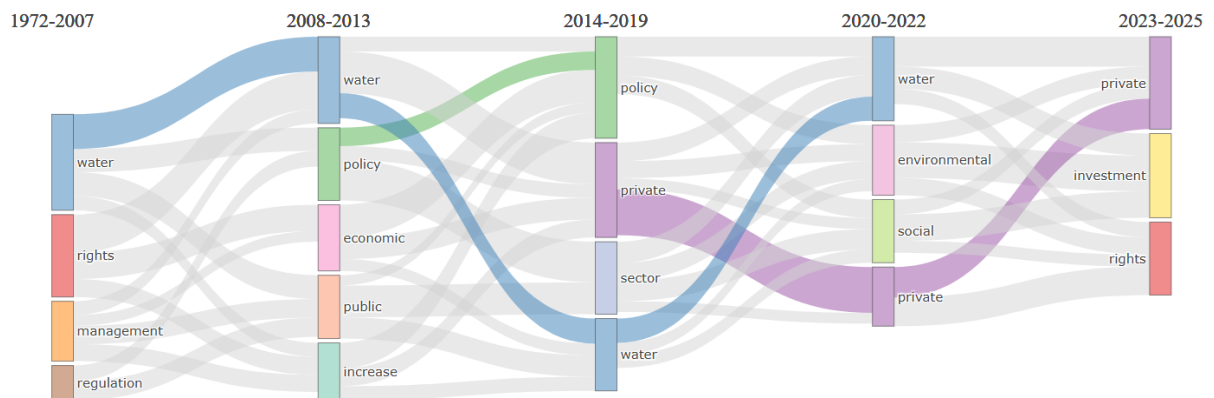


Figura 2.12 – Evolução temática dos resumos de 1972 a 2025, dividida em cinco períodos. A espessura dos fluxos é proporcional à intensidade da associação entre os temas. Fonte: elaboração própria.

a privatização reduziu a mortalidade infantil por doenças infecciosas e parasitárias em 8% na média, e que em áreas mais pobres essa redução chegou a 26% (Galiani; Gertler; Schargrotsky, 2005).

Além disso, houve aumento do investimento em infraestrutura física e da qualidade dos serviços em decorrência da melhora de eficiência das empresas, o que propiciou aumento significativo do acesso ao serviço de água, notadamente entre os mais pobres, fato que se traduz em redução da desigualdade em saúde (Galiani; Gertler; Schargrotsky, 2005).

No segundo grupo de artigos, de 2008 a 2013, emergem os termos “policy”, “economic” e “public”, enquanto “water” se mantém dominante. Esse deslocamento reflete a maturação do debate: saindo de questões de direitos e regulação formal, o campo passa a enfatizar a avaliação econômica das políticas de provisão de serviços públicos e, em particular, os desafios de desenho e implementação de contratos de concessão.

Trabalho destacado desse período é o estudo de Guasch, Laffont e Straub (2008) sobre renegociação de contratos de concessão na América Latina nos setores de transporte e água entre 1989 e 2000. Os autores analisam uma base de 307 concessões e documentam taxas alarmantes de renegociação: 53% no setor de transporte e 76% no de água, ocorridas em média apenas 3,1 e 1,6 anos após a assinatura do contrato, respectivamente.

Seus resultados empíricos demonstram que a probabilidade de renegociação é significativamente reduzida pela existência de um órgão regulador no momento da concessão, mas aumenta quando se adota regulação por price cap (mais frágil a choques), quando há exigências de investimento no contrato, em períodos pós-eleitorais, durante crises econômicas e, de modo geral, em ambientes institucionais caracterizados por maior corrupção e menor qualidade burocrática (Guasch; Laffont; Straub, 2008).

O terceiro corte temporal, de 2014 a 2019, é marcado pela ascensão conjunta de “policy”, “private” e “sector”, com o fluxo de “water” mantendo expressão. Essa combinação indica o

amadurecimento do debate sobre a participação privada em serviços essenciais, agora ancorado em evidências empíricas robustas sobre demanda e disposição a pagar por melhorias.

Trabalho representativo desse período é a meta-análise de Houtven *et al.* (2017), que sintetiza resultados de 60 estudos com 171 estimativas de disposição a pagar (DAP) para melhorias no acesso à água em países de baixa e média renda. Os autores demonstram que a DAP das famílias varia entre aproximadamente US\$3 e US\$30 por mês e é sensível ao escopo (magnitude da melhoria no serviço), à renda domiciliar e ao método de elicitação utilizado. Crucialmente, seus resultados confirmam que a DAP excede os custos de provisão em regiões com baixa cobertura, validando a viabilidade econômica de políticas de universalização quando há comprometimento político-financeiro sustentado.

No quarto período, de 2020 a 2022, observa-se uma diversificação temática expressiva: os termos “water” e “private” mantêm relevância, mas surgem com força “environmental” e “social”, sugerindo que a literatura passou a incorporar preocupações com sustentabilidade ambiental e equidade social na análise das formas de provisão de serviços de saneamento. Essa ampliação temática é consistente com a intensificação dos debates sobre os Objetivos de Desenvolvimento Sustentável (ODS) e com a aprovação do NMLSB no Brasil em 2020, mas reflete também uma reavaliação crítica das prescrições econômicas convencionais como privatização, precificação e reforma de direitos de propriedade.

Conforme documentado por Garrick, Hanemann e Hepburn (2020), essas receitas falharam em diversos contextos por não considerarem adequadamente a política da água e os conflitos distributivos associados. Os autores identificam cinco características físicas e sociais distintivas da água (mobilidade, variabilidade temporal e espacial, peso, multiplicidade de usos e não-rivalidade) que tornam a precificação via mercado e os direitos de propriedade instrumentos insuficientes quando aplicados isoladamente.

A crescente dependência de águas subterrâneas e de sistemas não conectados em rede (especialmente no Sul Global) exacerba esses desafios, demandando governança híbrida que combine papéis públicos e privados, sinais de quantidade além de preços, e estratégias políticas para enfrentar conflitos distributivos de frente (Garrick; Hanemann; Hepburn, 2020).

No quinto e mais recente período, de 2023 a 2025, os termos “private”, “investment” e “rights” emergem como os mais proeminentes, com o fluxo de “private” recebendo contribuições de múltiplos períodos anteriores, o que confirma a permanência desse tema como eixo central do campo. A ascensão de “investment” como termo de destaque é particularmente significativa e reflete o desafio global de mobilizar capital privado para atingir as metas de universalização previstas nos ODS.

Fabre e Straub (2023), em revisão abrangente publicada no *Journal of Economic Literature*, sintetizam as evidências sobre o impacto das parcerias público-privadas (PPPs) em infraestrutura, incluindo água e saneamento. Seus achados revelam que os efeitos das PPPs variam substancialmente conforme o desenho contratual, o ambiente regulatório e a capacidade do setor

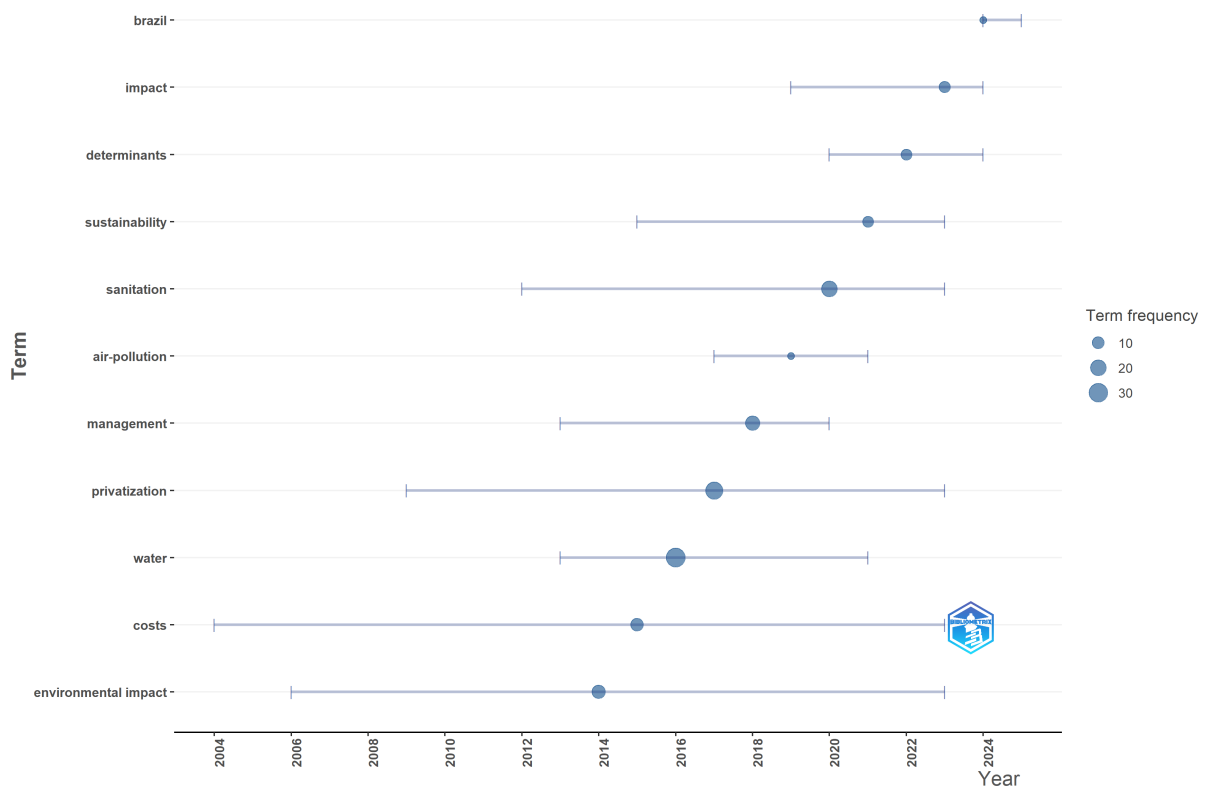


Figura 2.13 – Tópicos de tendência na literatura sobre participação privada no saneamento. O tamanho dos pontos é proporcional à frequência do termo. Fonte: elaboração própria.

público.

Em particular, os autores identificam lacunas críticas sobre o papel do capital humano: como se dá a seleção de trabalhadores em ambos os lados da parceria, como isso afeta o desenho dos contratos, e se parte do impacto das PPPs pode ser atribuída à qualidade dos trabalhadores envolvidos. Essa agenda converge com a necessidade de modelos de financiamento híbridos que combinem recursos públicos, privados e de impacto, em oposição à dicotomia entre provisão pública e privada (Fabre; Straub, 2023).

A análise dos tópicos de tendência (*Trend Topics*), apresentada na Figura 2.13, complementa a leitura da evolução temática ao mostrar quais termos têm concentrado maior atenção ao longo do tempo e, sobretudo, quais estão em ascensão mais recente. O termo “brazil” emerge como o mais recente dos tópicos em tendência, com seu pico de frequência centrado em 2024, o que evidencia o crescente interesse científico pelos desdobramentos da reforma do setor de saneamento brasileiro.

Os termos “impact” e “determinants” também apresentam tendência ascendente recente, o que corrobora a predominância de abordagens econométricas de avaliação causal na produção mais atual. Já “privatization”, “water”, “management” e “sanitation” figuram como termos de frequência mais elevada e de maior longevidade, demonstração de seu papel de pilares temáticos do campo.

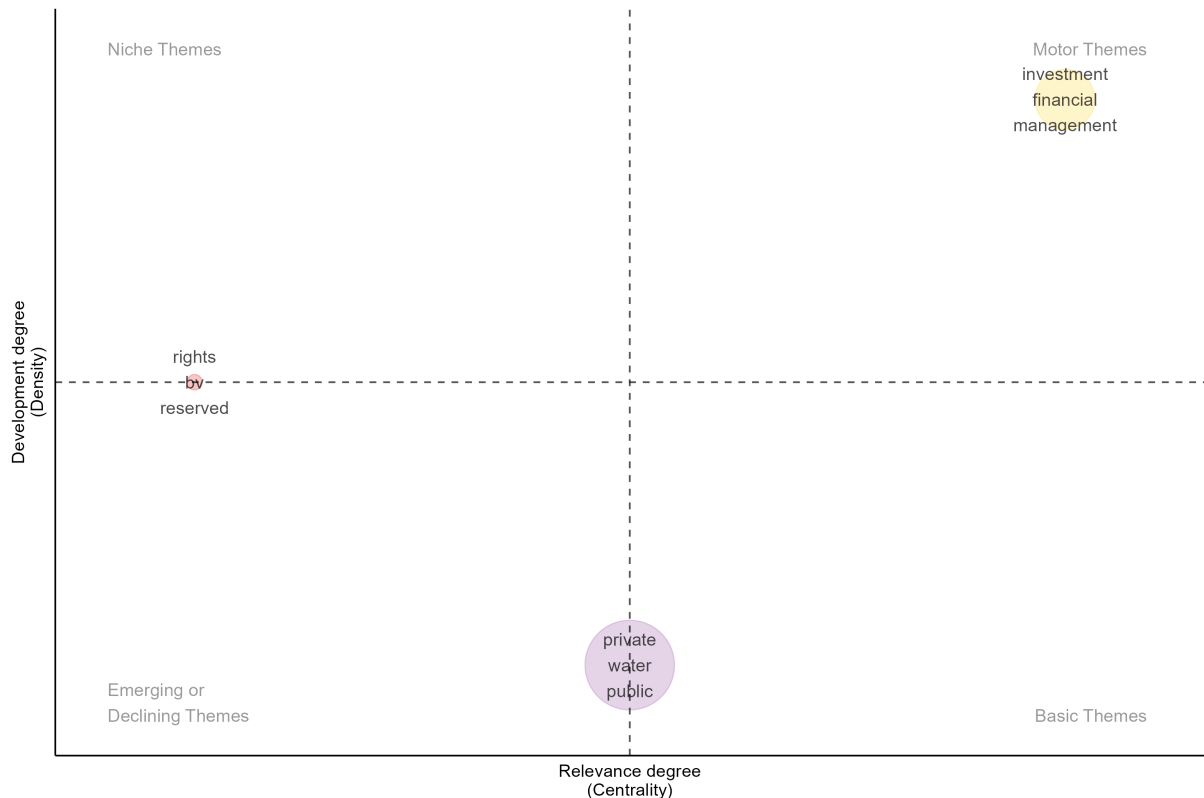


Figura 2.14 – Mapa temático do período 2023–2025 da literatura sobre participação privada no saneamento. Os quadrantes indicam: temas motores (alta centralidade e alta densidade), temas de nicho (baixa centralidade, alta densidade), temas emergentes ou em declínio (baixa centralidade e densidade) e temas básicos ou transversais (alta centralidade, baixa densidade). Fonte: elaboração própria.

2.3.3.1 Mapa temático do período recente e os artigos do quinto período

Para aprofundar a análise do quinto período (2023–2025), foi gerado um mapa temático que posiciona os clusters de palavras-chave em dois eixos: o grau de desenvolvimento interno do tema (*density*, eixo vertical) e o grau de relevância ou centralidade para o campo (*centrality*, eixo horizontal). Essa representação permite classificar os temas em quatro quadrantes interpretativos, conforme a Figura 2.14, e pode ser obtida também via *biblioshiny*.

O mapa temático revela uma estrutura de três posicionamentos distintos para o período mais recente. No quadrante superior direito, dos temas motores (aqueles mais centrais e mais densamente desenvolvidos internamente), situam-se os clusters “investment”, “financial” e “management”. Esse posicionamento indica que o debate mais recente sobre participação privada no saneamento gravita fundamentalmente em torno de questões de financiamento de infraestrutura, viabilidade financeira das concessões e modelos de gestão.

Trata-se de um deslocamento relevante em relação às décadas anteriores, quando os motores do campo eram as questões de regulação tarifária e privatização per se. A centralidade de “investment” e “financial” no período atual é coerente com o desafio global de mobilizar capital privado para atingir as metas de universalização do saneamento previstas nos ODS e, no caso

brasileiro, no NMLSB.

No quadrante inferior central, próximo à linha de base, situa-se o grande cluster “private–water–public”, classificado como tema emergente ou em processo de reconfiguração. Apesar de sua alta frequência acumulada, esse cluster apresenta baixa densidade interna no período mais recente, o que pode indicar uma fragmentação do debate: os artigos que tratam de “private” e “water” e “public” o fazem por ângulos cada vez mais diversos (regulação, impacto distributivo, governança, economia circular, direitos de propriedade), sem convergir para um subcampo coeso.

No quadrante esquerdo, na fronteira dos temas de nicho, figuram “rights” e “reserved”, com densidade mediana mas baixa centralidade, sugerindo que as questões jurídicas de direitos de água e reservas hídricas, embora persistentes, ocupam posição periférica na agenda econômica do campo.

Os artigos identificados como representativos do quinto período são descritos no quadro 2.1. Eles são heterogêneos em termos geográficos e temáticos, mas convergem em torno dos temas motores identificados no mapa: PPPs como modelo de provisão de infraestrutura, financiamento de serviços públicos e gestão sustentável de recursos.

O conjunto de artigos do quinto período ilustra com clareza a dualidade temática identificada no mapa: de um lado, trabalhos com alta citação e foco em avaliação de impacto, como Buntaine *et al.* (2024) (91 citações) e Fabre e Straub (2023) (16 citações), que ancoram os temas motores de “management” e avaliam mecanismos de governança e eficácia das PPPs; de outro, uma série de estudos aplicados a contextos nacionais e setoriais específicos (Zimbábue, Austrália, cidades africanas, China) que exploram dimensões de “investment” e “rights” em configurações institucionais distintas.

Fabre e Straub (2023), em revisão abrangente publicada no *Journal of Economic Literature*, sintetiza as evidências sobre o impacto das PPPs em infraestrutura, saúde e educação, concluindo que os efeitos variam substancialmente com o desenho contratual, o ambiente regulatório e a capacidade do setor público.

O estudo de Buntaine *et al.* (2024), por sua vez, avança na fronteira da governança ambiental ao demonstrar que mecanismos de participação social (via redes sociais, neste caso) podem substituir parcialmente a fiscalização pública insuficiente, reduzindo violações ambientais e melhorando a qualidade da água em mais de 60% nos casos analisados.

Ambos os artigos convergem para uma mensagem central ao campo: a qualidade dos arranjos de governança (sejam contratuais, regulatórios ou participativos) é determinante para os resultados das formas de provisão privada em serviços essenciais.

Já o estudo de Jiang *et al.* (2023) sobre o Rio Chishui aprofunda a análise de PPPs em contextos de compensação ecológica, mostrando como estruturas de incentivos bem desenhadas podem promover cooperação sustentável entre municípios e operadores privados em bacias hidrográficas. Esse achado é particularmente relevante para o Brasil, cujo território é marcado por complexas interdependências hídrico-municipais e cujo NMLSB prevê instrumentos de

Quadro 2.1 – Artigos representativos do quinto período (2023–2025) identificados pelo mapa temático. Brasília, 2026.

Autores	Título / Tema	Periódico	Citações
Buntaine et al. (2024)	Participação cidadã e governança ambiental na China: efeitos das redes sociais sobre violações ambientais e qualidade da água	<i>Am. Econ. Rev.</i>	91
Fabre e Straub (2023)	Revisão abrangente do impacto das PPPs em infraestrutura, saúde e educação	<i>J. Econ. Lit.</i>	16
Jiang et al. (2023)	Cooperação sustentável em PPPs de compensação ecológica de bacias hidrográficas na China (Rio Chishui)	<i>Socio-Econ. Plann. Sci.</i>	15
Jiang, Pei e Zhang (2024)	Crédito verde e integração de firmas nas cadeias globais de produção	<i>China Econ. Rev.</i>	14
Taron et al. (2023)	PPPs para a bioeconomia circular no Sul Global: lições aprendidas	<i>Resour. Recovery Reuse</i>	7
León-Pérez et al. (2024)	Regime jurídico do sargaço em Porto Rico: direitos ambientais e gestão costeira	<i>Marine Policy</i>	4
Atofarati et al. (2025)	Soluções inteligentes e sustentáveis de gestão de resíduos em cidades africanas em urbanização	<i>Utilities Policy</i>	3
Chiswa (2025)	PPPs como catalisadores do desenvolvimento: oportunidades e desafios no Zimbábue	<i>Eur. Procur. PPP Law Rev.</i>	0
Seidl et al. (2025)	Legislação de recarga de aquíferos gerenciados na Austrália: barreiras e lições	<i>Aust. J. Agric. Resour. Econ.</i>	0
Usykova et al. (2025)	Suporte institucional para economia circular no agronegócio	<i>Scientific Horizons</i>	0

Fonte: dados da pesquisa.

regionalização que envolvem consórcios e concessões multimunicipais.

2.3.4 Agenda de pesquisa para o Brasil

A Figura 2.12 e o mapa temático da Figura 2.14 mostram que “private”, “investment” e “rights” são os temas mais proeminentes do período mais recente (2023–2025), com “water” e “public” compondo o entorno do campo. De modo complementar, o quadro 2.1 identifica os dez artigos mais representativos desse período, cuja leitura completa permite mapear as lacunas de pesquisa relevantes para o Brasil. Treze artigos do conjunto total eram elegíveis para leitura integral com foco na agenda brasileira, e nove deles apresentaram sugestões explícitas de pesquisa futura.

A análise desses artigos permitiu identificar sete eixos de lacunas de conhecimento para o Brasil, detalhados a seguir: (i) estudos econométricos de impacto com desenhos causais robustos; (ii) análises de eficiência comparativa entre prestadores públicos e privados; (iii) investigação sobre escala e regionalização; (iv) regulação e captura regulatória; (v) equidade tarifária; (vi) comportamento do consumidor; e (vii) determinantes e *design* da privatização.

(i) Estudos econométricos de impacto

Tourinho *et al.* (2022) realizaram análise de *benchmarking* em municípios brasileiros para dimensões de eficiência e cobertura, encontrando baixa eficiência média para serviços de abastecimento de água e saneamento. Sua sugestão inclui investigar relações de causa e efeito entre eficiência e cobertura: se ações de melhoria de eficiência levam ao aumento da cobertura por meio de superávit financeiro, ou se o aumento da cobertura resulta em ganhos de eficiência ao alavancar economias de escala.

Dassan *et al.* (2024) utilizaram um delineamento quase experimental e encontraram queda inicial de tarifa no momento da concessão, embora os reajustes subsequentes tenham sido superiores aos do setor público nos primeiros quatro anos, além de aumento da cobertura de água, coleta e tratamento de esgoto após a privatização. Os autores usaram informações sobre privatizações em municípios considerando a Lei 11.445 de 2007 e apontam como agenda futura a aplicação da mesma metodologia à reforma de 2020, que institui o NMLSB. Bagnoli *et al.* (2023), em pesquisa sobre propriedade de serviços públicos em países em desenvolvimento, incluindo o Brasil, destaca a necessidade de estudos econométricos para avaliar a causalidade entre escolhas políticas e resultados sociais, apontando a falta de estudos que informem se há viés previsível de acordo com a propriedade, o que poderia ser valioso no contexto de pacotes de reforma.

(ii) Análises de eficiência

Os limites de eficiência das prestadoras de serviços de água e esgoto nas maiores cidades brasileiras foram analisados por de Souza Pereira *et al.* (2022), que encontraram diferenças na provisão de água entre empresas públicas e privadas, sendo as últimas mais eficientes, mas sem diferença significativa no esgoto. Os autores concluem que faltam incentivos para a melhoria de eficiência nesse mercado e apontam como agenda a análise do NMLSB, à época ainda em tramitação.

O trade-off entre custo e qualidade é investigado por Oliveira e Saiani (2022), que encontram elevação da qualidade com redução dos custos nos provedores privados. Os autores questionam, no entanto, se esse padrão se mantém nos prestadores locais e sugerem análises no nível do município, tanto em termos de cobertura quanto de qualidade.

Fabre e Straub (2023), em pesquisa sobre parcerias público-privadas em infraestrutura de água e saneamento, detectaram uma lacuna sobre a importância do capital humano nesse contexto. Levantam questões sobre o processo de seleção de trabalhadores em ambos os lados da parceria, como isso afeta o design dos contratos, e se parte do impacto das PPPs pode ser atribuído à qualidade dos trabalhadores envolvidos.

(iii) Escala e regionalização

A questão da escala é transversal a vários dos estudos analisados, mas emerge com contornos próprios quando se considera o processo de regionalização induzido pelo NMLSB. Tourinho *et al.* (2022) sugerem que o aumento da cobertura pode gerar ganhos de eficiência por meio de economias de escala. Cabe notar que, no estudo de Tourinho, diferenças estatisticamente significativas entre prestadores públicos e privados foram encontradas apenas para entidades de nível regional, e não local, com os próprios autores alertando para cautela na interpretação dado o número reduzido de entidades privadas regionais na amostra. Isso reforça a necessidade de separar empiricamente os efeitos da propriedade dos efeitos da escala operacional, distinção que análises desagregadas no nível municipal ainda não realizaram. Oliveira e Saiani (2022) reforçam essa lacuna ao não encontrarem o mesmo *trade-off* custo-qualidade nos prestadores locais que identificaram nos privados de maior porte, sugerindo que análises desagregadas no nível municipal são necessárias para essa separação. A criação de consórcios e concessões multimunicipais prevista no NMLSB representa, nesse sentido, um experimento natural ainda não explorado pela literatura.

(iv) Regulação e captura regulatória

Aquino e Ledo (2023) desenvolveram e testaram empiricamente um modelo teórico sobre interferência política no processo regulatório, encontrando que ciclos eleitorais municipais são negativamente associados aos preços definidos pelos reguladores. Em termos de agenda, indicam a ampliação do modelo para mais períodos e a estimação de parâmetros estruturais.

Rufino e Cavalcante (2024) debruçaram-se sobre a questão da captura regulatória por empresas privadas, encontrando que companhias privadas de saneamento conseguem reajustes tarifários mais elevados que entidades públicas, resultando em lucros anormais. Esse resultado, combinado com os achados de Aquino e Ledo (2023), sugere que a qualidade do ambiente regulatório é uma variável mediadora central nos efeitos da privatização sobre o bem-estar social, dimensão que merece investigação sistemática no contexto do novo marco regulatório.

(v) Equidade tarifária

Trabalho mais recente, já sobre os impactos do NMLSB, encontrou aumento do peso das tarifas de saneamento sobre o orçamento de famílias mais pobres (Fagundes; Cunha Marques;

Malheiros, 2024). Os autores analisaram dois leilões com base na maior oferta (em Alagoas e no Rio de Janeiro) e indicam como análise futura os novos leilões cujo critério é a menor tarifa. Esse eixo conecta-se diretamente à discussão sobre o design ótimo de leilões: como demonstrado pela experiência argentina analisada por Galiani, Gertler e Schargrotsky (2005), leilões com critério de menor tarifa tendem a ser mais favoráveis às populações de baixa renda do que leilões com critério de maior oferta, o que tem implicações diretas para a avaliação dos leilões realizados sob o NMLSB.

(vi) Comportamento do consumidor

Sousa, Fávero e Fouto (2025) demonstraram que municípios atendidos por prestadoras privadas apresentam menor consumo *per capita* de água entre 2002 e 2019, mas reconhecem não poder distinguir se esse efeito decorre do preço cobrado ou da própria percepção dos usuários sobre o tipo de prestador. Propõem, como agenda, a realização de *surveys* que investiguem diretamente as atitudes e comportamentos dos usuários em relação à titularidade do serviço, dimensão ainda ausente na literatura brasileira e de particular relevância no contexto de expansão da participação privada induzida pelo NMLSB.

(vii) Determinantes e *design* da privatização

Nunes e Marques (2025), ao analisarem os determinantes da privatização no período 2012–2023, encontraram que prestadoras financeiramente mais sólidas têm maior probabilidade de serem privatizadas, contrariando a hipótese de que a concessão seria motivada pela necessidade de corrigir serviços deficientes. Os autores apontam como lacunas o estudo do agrupamento geográfico de *holdings* privadas e a integração explícita do *Multiple Streams Framework* para testar empiricamente o papel das janelas de oportunidade política no processo de concessão.

Chaves e Costa (2025), ao avaliarem os impactos epidemiológicos da privatização entre 1998 e 2021, identificaram resultados heterogêneos entre coortes e destacam os municípios de Campo Grande e Sorriso (MS) como casos de referência para análise de *design* contratual, além de reforçarem a superioridade da privatização plena sobre o modelo híbrido com direitos estatais observado no Tocantins antes de 2014.

Em síntese, as lacunas de conhecimento identificadas na literatura recente sobre o Brasil podem ser sistematizadas conforme o quadro 2.2:

Quadro 2.2 – Agenda de pesquisa para o Brasil sobre participação privada no saneamento. Brasília, 2026.

Eixo	Lacuna identificada	Referência
(i) Estudos econô- métricos	Efeitos causais da privatização pós-NMLSB sobre tarifas, cobertura e qualidade	Dassan <i>et al.</i> (2024), Bagnoli <i>et al.</i> (2023)
(ii) Análise de eficiência	Eficiência em esgotamento sanitário e papel do capital humano nas PPPs	de Souza Pereira <i>et al.</i> (2022), Fabre e Straub (2023)
(iii) Escala e regionalização	Trade-off custo-qualidade nos prestadores locais e efeitos da regionalização	Oliveira e Saiani (2022), Tourinho <i>et al.</i> (2022)
(iv) Regulação e cap- tura	Ciclos eleitorais, captura regulatória e leilões com critérios alternativos	Aquino e Ledo (2023), Ruffino e Cavalcante (2024), Fagundes, Cunha Marques e Malheiros (2024)
(v) Equidade tarifária	Impactos distributivos do NMLSB sobre famílias de baixa renda	Fagundes, Cunha Marques e Malheiros (2024)
(vi) Comportamento do consumidor	Efeito da percepção de titularidade sobre o consumo de água, independentemente do preço cobrado	Sousa, Fávero e Fouto (2025)
(vii) Determinan- tes e <i>design</i> da privatização	Janelas de oportunidade política, agrupamento geográfico de <i>holdings</i> e aprendizado a partir de casos contratuais de referência	Nunes e Marques (2025), Chaves e Costa (2025)

Fonte: dados da pesquisa.

2.4 Discussão

A fundamentação teórica desse campo encontra-se principalmente em Laffont e Tirole (1993), que trata de incentivos na regulação e teoria dos contratos. Duas contribuições são particularmente relevantes: o processo de seleção via leilão da concessionária que ocupará o monopólio natural do saneamento, e os efeitos das estruturas de propriedade alternativas, pública e privada (Laffont; Tirole, 1993).

A rede de co-citação corrobora essa leitura: o cluster central da literatura agrupa, de forma densa, obras sobre privatização (Megginson e Netter (2001); Boycko, Shleifer e Vishny (1996)), teoria da regulação (Laffont e Tirole (1993); Averch e Johnson (1962)), eficiência comparativa (Byrnes, Grosskopf e Hayes (1986); Teeple e Glyer (1987)) e avaliação de impacto (Galiani, Gertler e Schargrodsky (2005); McKenzie e Mookherjee (2003); Estache e Rossi (2002)). Esse conjunto revela que o campo científico tem se estruturado em torno de duas tradições complementares: uma normativa, que discute como contratos e regulação devem ser desenhados para alinhar incentivos; e uma empírica, que avalia *ex-post* os resultados dessas reformas institucionais.

A rede de co-autoria, por sua vez, revela que o campo ainda é relativamente fragmentado. A

ausência de um componente central que conecte os diferentes grupos de pesquisa pode dificultar a acumulação de conhecimento e a replicação de métodos entre contextos. Isso sugere que há espaço para maior colaboração internacional, especialmente entre pesquisadores de países em desenvolvimento (como Brasil, Índia e países africanos) que lidam com desafios similares de acesso e regulação, mas raramente figuram no mesmo cluster de co-autoria.

Sugere-se que o leilão alinhe incentivos entre as partes envolvidas, resultando em eficiência alocativa e minoração dos problemas de assimetria de informação (Laffont; Tirole, 1993). No caso do saneamento, em que existe prazo mais alongado para maturação dos investimentos, faz-se importante que os contratos considerem incentivos para manutenção de qualidade e eficiência ao longo do tempo (Laffont; Tirole, 1993), enquanto minimizam tarifas e garantem acesso universal, objetivo do NMLSB (BRASIL, 2020), conforme verificado em reforma brasileira anterior (Dassan *et al.*, 2024).

Um leilão bem projetado é fundamental para reduzir o risco de seleção adversa, seja baseado em custo mínimo ou tarifa máxima (Laffont; Tirole, 1993). Leilão realizado a partir do NMLSB com base em maior oferta, um critério não ótimo, fez com que as tarifas se elevassem para os mais pobres, reduzindo o bem-estar social (Fagundes; Cunha Marques; Malheiros, 2024). Resultado contrário foi observado na Argentina, em que a concessão foi decidida por leilão de critério menor tarifa e a privatização levou a aumento significativo no acesso aos serviços de água, especialmente entre as populações de baixa renda (Galiani; Gertler; Schargrodsky, 2005).

A metodologia de Galiani, Gertler e Schargrodsky (2005) também é muito influente: os autores utilizaram uma abordagem de diferenças-em-diferenças combinada com pareamento via *propensity score matching* (PSM). O PSM foi empregado para parear os municípios com base na probabilidade de privatizar os serviços de água, reduzindo diferenças sistemáticas entre os municípios tratados e os de controle antes da privatização, o que diminui o viés de seleção e contribui para a identificação da relação causal no que toca ao efeito sobre a mortalidade infantil. Esse desenho metodológico tem sido amplamente replicado em estudos posteriores e é a referência metodológica mais recomendada pela agenda de pesquisa brasileira identificada.

Em relação à privatização, a regulação exerce papel fundamental, na medida em que as empresas privadas tendem a ser mais eficientes que as públicas quando bem reguladas (Laffont; Tirole, 1993). O risco de captura regulatória, porém, é bastante presente nesse cenário, em que empresas privadas podem influenciar a agência reguladora em seu benefício (Laffont; Tirole, 1993), como apontam evidências para o Brasil (Aquino; Ledo, 2023). O risco de práticas predatórias e exclusão social, como reajustes tarifários superiores aos concessionários públicos (Rufino; Cavalcante, 2024) ou que pesem mais sobre os mais pobres (Fagundes; Cunha Marques; Malheiros, 2024), deve ser central nas discussões regulatórias.

O efeito regulatório sobre os diferentes tipos de propriedade das empresas de água e esgoto vem sendo estudado há bastante tempo (Mann; Mikesell, 1976) e, como pode ser visto em trabalhos mais recentes (de Souza Pereira *et al.*, 2022; Oliveira; Saiani, 2022; Fabre; Straub,

2023), não há consenso sobre a origem das diferenças de eficiência: se advêm da propriedade, da escala, da regulação ou mesmo do capital humano. Esse debate permanece em aberto e representa uma das mais profícuas fronteiras de pesquisa do campo.

A análise dos artigos permite concluir que a regulação é o tema mais recorrente nesse campo de estudo, visto que as motivações da produção científica giram em torno da privatização com destaque aos processos ocorridos nos anos 1990 e 2000, e que o Brasil ocupa posição crescentemente relevante nessa literatura, especialmente com o NMLSB.

2.5 Conclusões

Uma revisão bibliométrica é capaz de sintetizar a produção científica de uma área do conhecimento, o que é de grande valia para pesquisadores que desejam avançar a ciência de um tema. Neste estudo foi realizada uma revisão bibliométrica sobre a participação das empresas privadas no saneamento, analisando 392 documentos advindos das bases Scopus e Web of Science, publicados entre 1972 e 2025 em 215 periódicos distintos. A taxa de crescimento anual de 6,34% evidencia o dinamismo do campo, que se acelerou especialmente após 1997, coincidindo com as grandes ondas de privatização global, e voltou a crescer após 2020, impulsionado pelo debate sobre o NMLSB no Brasil.

Os temas mais relevantes, considerando os artigos mais citados, foram privatização e seus efeitos sobre cobertura, eficiência e mortalidade infantil. Já como agenda de pesquisa para o Brasil, os trabalhos mais recentes apontam para sete eixos prioritários: estudos econométricos de impacto sobre os efeitos do NMLSB; análises de eficiência comparativa entre prestadores públicos e privados, incluindo o papel do capital humano nas PPPs; investigação sobre o *trade-off* entre custo e qualidade em nível municipal e os efeitos da regionalização; regulação e captura regulatória; equidade tarifária e impactos distributivos das reformas sobre famílias de baixa renda; comportamento do consumidor em relação à titularidade do serviço; e determinantes e *design* da privatização.

A análise de redes revelou que o campo apresenta fragmentação na colaboração entre grupos de pesquisa e dois clusters intelectuais bem definidos: um central, ancorado em teoria da regulação e estudos de impacto de privatização, e outro especializado em análise de eficiência por métodos de fronteira. A emergência do Brasil como tópico de tendência na produção científica mais recente reforça a oportunidade de contribuição desta tese para a literatura internacional.

Como fragilidades desta investigação, pode-se destacar a inclusão de apenas duas bases, pois apesar de serem bastante relevantes e criteriosas na inclusão de periódicos, não refletem a totalidade do conhecimento produzido sobre esse assunto. Outra fragilidade deve-se à estratégia de busca adotada, cujo objetivo é fornecer uma busca sensibilizada da literatura, mas que restringe os idiomas e direciona os resultados para trabalhos de economia, podendo subrepresentar contribuições interdisciplinares e de ciência política. Por fim, a análise de co-palavras baseada

em resumos não captura nuances presentes nos textos completos, o que representa uma limitação inerente a abordagens bibliométricas em larga escala. Para estudos futuros, recomenda-se a inclusão de outras bases de dados bibliográficos, como o Google Scholar, para aumentar o alcance de literatura não indexada nas bases ora consultadas, bem como a extensão do escopo pela retirada da restrição de idiomas.

3 Eficiência dos prestadores de serviço de saneamento no Brasil: uma análise de fronteira estocástica entre 2015 e 2022

3.1 Introdução

O saneamento básico é um dos setores com maior capacidade de gerar externalidades positivas sobre o desenvolvimento econômico: investimentos em abastecimento de água e esgotamento sanitário elevam o estoque de capital humano por meio da redução de doenças de veiculação hídrica, aumentam a produtividade da força de trabalho e reduzem gastos públicos com saúde (Weil, 2007). Apesar disso, o Brasil apresenta déficits expressivos de cobertura: em 2024, aproximadamente 15% da população ainda não tinham acesso à rede de abastecimento de água tratada e mais de 37% careciam de coleta de esgoto (Brasil. Ministério das Cidades, 2025a; Brasil. Ministério das Cidades, 2025b). Esses números situam o Brasil em posição desfavorável mesmo em relação a países de renda comparável na América Latina (Ferro *et al.*, 2014), e tornam a meta de universalização dos serviços até 2033, estabelecida pelo NMLSB, um dos maiores desafios de política pública da agenda brasileira contemporânea.

A viabilidade econômica dessa universalização depende criticamente da eficiência com que os prestadores de serviços utilizam seus recursos. O setor de saneamento tem características de monopólio natural, altos custos fixos de infraestrutura, baixos custos marginais de atendimento e demanda inelástica (Silva; Ledo, 2023b), o que inibe a concorrência direta e torna o benchmarking regulatório o principal instrumento para induzir ganhos de eficiência (Shleifer, 1985; Tupper; Resende, 2004). Em um ambiente em que o regulador não observa diretamente os custos mínimos de cada prestador, a comparação sistemática de desempenho entre operadores, a regulação por comparação (*yardstick competition*), simula a pressão competitiva e mitiga a assimetria de informação inerente ao monopólio (Shleifer, 1985). Para que esse instrumento funcione, é necessário conhecer com precisão o nível de eficiência de cada operador e os fatores que o determinam.

Uma questão central nesse debate, tanto no Brasil quanto internacionalmente, é se a propriedade das firmas (pública ou privada) afeta sua eficiência. O argumento teórico em favor das empresas privadas baseia-se na maior pressão competitiva, nos incentivos de mercado e na disciplina financeira a que estão sujeitas (Bel; Warner, 2008). O argumento contrário ressalta que, em monopólios naturais sem regulação efetiva, a propriedade privada transfere rendas de monopólio sem necessariamente gerar ganhos de eficiência (Massarutto *et al.*, 2023). Mais de cinco décadas de evidências empíricas internacionais não resolveram essa controvérsia: os

resultados são sistematicamente inconclusivos, variando com o contexto institucional, o período analisado e a metodologia empregada (Walter *et al.*, 2009; Worthington, 2014). Para o Brasil, os estudos disponíveis chegaram consistentemente à mesma conclusão: a propriedade não determina a eficiência quando se controlam as características estruturais do serviço e o ambiente regulatório (Da Silva E Souza; De Faria; Moreira, 2007; Ferro *et al.*, 2014; Silva; Ledo, 2023a).

Este artigo contribui para essa literatura em três dimensões. Primeira, atualiza as estimativas de eficiência para o período 2015–2022. Segunda, estima uma função de custo variável translogarítmica por Análise de Fronteira Estocástica com ineficiência tempo-variante (Battese; Coelli, 1992), especificação mais flexível e robusta que as adotadas na maioria dos estudos anteriores para o Brasil e que, ao contrário das especificações tempo-invariantes predominantes na literatura nacional, permite documentar a trajetória de convergência dos prestadores em direção à fronteira de custo ao longo do período analisado. Terceira, e mais importante, produz a linha de base empírica pré-reforma contra o qual os efeitos do NMLSB poderão ser avaliados em pesquisas futuras, preenchendo uma lacuna metodológica fundamental, uma vez que nenhum estudo havia documentado rigorosamente as condições do setor imediatamente antes da maior reforma regulatória de sua história recente.

Os resultados mostram que a eficiência técnica média dos 256 prestadores de água e esgoto analisados foi de 75,4% no período, com trajetória de crescimento lento e contínuo (de 74,1% em 2015 a 76,7% em 2022), sem ruptura identificável em torno de 2020. A propriedade privada não apresenta efeito estatisticamente significativo sobre os custos em nenhuma das especificações estimadas, resultado esse que é robusto a diferentes janelas temporais. O principal determinante dos custos é a extensão da rede de infraestrutura, com elasticidade de 0,882, o que implica que a expansão necessária para a universalização impõe pressão proporcional sobre os custos operacionais dos prestadores. Esses achados têm implicações diretas para o desenho de contratos de concessão e para a política de consolidação regional prevista no NMLSB.

O restante do artigo está organizado da seguinte forma. A Seção 3.2 apresenta a revisão da literatura sobre eficiência no setor de saneamento, com ênfase na hipótese da propriedade e no papel do ambiente regulatório. A Seção 4.3 descreve a base de dados, a especificação da função de custo translog e o modelo de ineficiência tempo-variante. A Seção 4.4 apresenta os resultados e os testes de robustez. A Seção 4.5 discute os achados à luz da literatura. A Seção 4.6 conclui com implicações de política e agenda de pesquisa.

3.2 Revisão de Literatura

O setor de saneamento básico apresenta características de monopólio natural: altos custos fixos de infraestrutura de rede, baixos custos marginais de atendimento e externalidades positivas sobre a saúde pública que tornam a provisão privada espontânea insuficiente (Silva; Ledo, 2023b). Nesse contexto, a avaliação comparativa de desempenho, ou *benchmarking*, torna-se o principal

instrumento regulatório para mitigar a assimetria de informação inerente entre regulador e regulado, simulando a pressão competitiva ausente em mercados monopolistas (Shleifer, 1985; Tupper; Resende, 2004). A literatura sobre eficiência no setor de água e esgoto, ao longo de mais de cinco décadas, convergiu para uma conclusão contraintuitiva: a propriedade da firma, por si só, não determina sua eficiência. O que importa é o ambiente institucional e regulatório no qual ela opera (Walter *et al.*, 2009; Worthington, 2014; Nyathikala *et al.*, 2023).

3.2.1 A hipótese da propriedade e seus resultados inconclusivos

Um dos primeiros trabalhos a investigar sistematicamente as diferenças de eficiência resultantes da propriedade das firmas na indústria de água foi o de Mann e Mikesell (1976), que examinou empresas norte-americanas e verificou que as públicas apresentam custos de serviço da dívida mais elevados, mas custos operacionais semelhantes aos das privadas. As firmas privadas diferenciavam-se por pagar mais impostos e salários mais altos, bem como por utilizar menos mão de obra; já as públicas tendiam a ter maior volume médio de produção, embora o número total de clientes e a população atendida por área fossem semelhantes nas duas formas organizacionais. O resultado mais relevante desse trabalho, contudo, foi estabelecer os determinantes gerais dos custos unitários e mostrar que as firmas privadas apresentavam custos unitários mais elevados que as públicas, o que contraria as hipóteses derivadas dos trabalhos teóricos da época sobre direitos de propriedade.

Essa inconclusividade inaugural antecipou um padrão que se repetiria sistematicamente nas décadas seguintes. Os primeiros trabalhos com funções de custo estimadas nos EUA chegaram a resultados opostos: Crain e Zardkoohi (1978b) encontraram custos operacionais significativamente mais altos nas públicas, ao passo que Feigenbaum e Teeple (1983b) não identificaram diferença relevante entre as equações de custo dos dois tipos de provedor. Byrnes, Grosskopf e Hayes (1986) e Fox e Hofler (1986) utilizaram, respectivamente, programação matemática e funções de fronteira compostas, sem encontrar diferenças significativas de eficiência técnica ou alocativa entre firmas públicas e privadas. Revisando essa literatura acumulada até o início dos anos 2000, Bel e Warner (2008) concluíram que os estudos baseados em dados transversais do período não sustentam a hipótese de que a privatização reduz custos por meio da competição: os custos dependem mais de características como a natureza do serviço, a dimensão geográfica do mercado e a estrutura industrial do setor, não havendo escolha ótima entre provisão pública e privada. Abbott e Cohen (2009) reforçam essa conclusão ao mostrar que os trabalhos que analisaram a mudança de eficiência trazida pela privatização revelam efeitos limitados ou até mesmo declínio na eficiência média.

Sob abordagens econométricas mais robustas, o quadro não melhorou substancialmente. Bhattacharyya *et al.* (1995) estimaram, com SFA para 257 empresas norte-americanas em 1992, que as firmas privadas eram em média *menos* eficientes que as públicas, embora com maior dispersão. Em sentido oposto, Estache *et al.* (2002) verificaram, com SFA para 50 empresas

asiáticas, ausência de diferença significativa entre os dois tipos de propriedade. Estache e Kouassi (2002), por sua vez, identificaram vantagem das firmas privadas ao estudar empresas africanas por análise de regressão. Para a Espanha, García-Sánchez (2006) concluíram, com DEA, que a natureza jurídica não exercia influência sobre a eficiência, enquanto a densidade populacional e de ligações eram os fatores realmente relevantes.

A inconclusividade se estende a países em desenvolvimento e a períodos mais recentes. Tan (2012) e Hon, Boon e Lee (2016) analisaram o setor de água da Malásia e não encontraram evidências estatisticamente significativas de que as empresas privatizadas superassem as públicas em eficiência, um resultado que os autores interpretam como evidência das limitações intrínsecas da privatização quando desacompanhada de reformas regulatórias estruturais. Suárez-Varela *et al.* (2017) chegaram à mesma conclusão para a Espanha, e Helm (2020) para a Inglaterra. Para a China, Zhao (2020) documentaram melhora na performance do setor associada à participação privada internacional, mas com ganhos de eficiência modestos. Nyathikala *et al.* (2023) estimaram, com SFA para 304 prestadores urbanos da Índia entre 2010 e 2015, que os ganhos de eficiência dependem de mecanismos de governança (como recuperação de custos, qualidade do serviço e redução de perdas) e não do tipo de propriedade. Para o Chile, os trabalhos de Molinos-Senante, Maziotis e Villegas (2022), Maziotis e Molinos-Senante (2022) e Maziotis *et al.* (2025) mostram resultados igualmente contraditórios a depender do modelo e do período: ora as empresas totalmente privadas superam as concessionárias e públicas, ora a empresa pública apresenta maior eficiência técnica, reforçando a conclusão de que o efeito da propriedade é sensível à especificação metodológica e ao contexto institucional.

Para a França, Lannier e Porcher (2014) utilizaram uma abordagem de três estágios combinando DEA e SFA para 177 prestadores em 2009 e encontraram que as empresas sob gestão privada eram, em média, ligeiramente *menos* eficientes que as públicas após controlar por variáveis ambientais. Os autores atribuíram essa diferença a estratégias distintas de gestão de recursos e destacaram a ausência de um regulador nacional como fator limitante para a convergência de desempenho.

Diante desse panorama, as meta-análises de Walter *et al.* (2009) e Worthington (2014) constituem referências metodológicas fundamentais. Ao revisar dezenas de estudos internacionais, os autores concluíram que os principais determinantes de eficiência são economias de escopo e densidade, ao passo que o efeito da propriedade e dos ganhos de escala depende criticamente do contexto institucional. As economias de escala, em particular, foram identificadas em múltiplos contextos nacionais como em Itália, França, Japão e Portugal, mas sua magnitude varia conforme o tamanho das firmas e a organização industrial do setor. Para Portugal, Carvalho e Marques (2016) demonstraram, com SFA bayesiana, a existência de economias de tamanho, de escopo e de integração vertical, sugerindo que fusões entre prestadores gerariam reduções de custo relevantes, o que é um resultado com implicações diretas para países com setores fragmentados como o Brasil.

3.2.2 O papel mediador do ambiente regulatório

A evidência internacional mais persuasiva em favor da privatização não provém de comparações estáticas entre firmas públicas e privadas, mas de estudos dinâmicos que analisam a trajetória das empresas após reformas regulatórias bem estruturadas. O caso inglês e galês é paradigmático: a privatização de 1989, acompanhada da criação do regulador setorial independente OFWAT, gerou ganhos documentados de produtividade. Saal, Parker e Weyman-Jone (2007) estimaram, por SFA para dez empresas no período 1985–2000, que a mudança tecnológica se acelerou após a privatização, embora alertando que empresas excessivamente grandes podem incorrer em deseconomias de escala. Molinos-Senante e Maziotis (2019) documentaram, por metafronteira estocástica para as mesmas empresas, como cada ciclo de revisão tarifária da OFWAT se traduziu em movimentos específicos na fronteira de eficiência, com as empresas de água e esgoto integradas apresentando desempenho superior às empresas exclusivamente de água. Complementarmente, Erbetta e Cave (2007) atribuíram os ganhos de eficiência observados não à privatização em si, mas às mudanças no marco regulatório que a acompanharam, sugerindo que o vetor causal é institucional, não de propriedade.

O contraponto é o caso italiano. Massarutto *et al.* (2023) argumentam que a ausência de uma agência reguladora efetiva no processo de privatização italiano foi determinante para que o setor não apresentasse ganhos consistentes de produtividade após a abertura ao capital privado. O mesmo padrão emerge de Anwandter e Ozuna (2002) para o México: a regulação autônoma, sem reformas competitivas estruturais, não foi suficiente para elevar a eficiência. Aubert e Reynaud (2005) formalizam essa interpretação ao estimar, com SFA para 211 empresas norte-americanas, que a estrutura regulatória à qual a empresa está submetida explica parcialmente seu nível de eficiência. Para a Eslovênia, Filippini, Hrovatin e Zorić (2008) encontraram evidências de deseconomias de escala em firmas de grande porte e mostraram que os escores de eficiência são sensíveis à especificação metodológica. Para países da América Latina, Estache e Trujillo (2003) documentaram crescimento de produtividade em quatro províncias argentinas após a privatização em um contexto em que contratos de concessão regulados substituíram a gestão direta do Estado. Para o Canadá, Price *et al.* (2017) estimaram por SFA para 944 empresas que a qualidade da água bruta eleva os custos e reduz a eficiência, reforçando que variáveis ambientais e estruturais superam a propriedade como determinantes do desempenho.

Tabela 3.1 – Síntese dos principais estudos internacionais sobre eficiência no setor de saneamento.

Autores	Método	Contexto	Principal resultado sobre propriedade/regulação
Crain e Zardkoohi (1978b)	AR	EUA, 1970	Custos maiores nas públicas
Feigenbaum e Teeples (1983b)	AR	EUA, 1970	Sem diferença entre público e privado
Byrnes, Grosskopf e Hayes (1986)	DEA	EUA, 1976	Sem diferença de eficiência técnica
Fox e Hofler (1986)	SFA	EUA, 1981	Sem diferença de eficiência técnica
Bhattacharyya <i>et al.</i> (1995)	SFA	EUA, 1992	Privadas <i>menos</i> eficientes
Estache e Kouassi (2002)	SFA	Ásia, 1995	Sem diferença entre público e privado
Estache <i>et al.</i> (2002)	AR	África, 1995–97	Privadas mais eficientes
Aubert e Reynaud (2005)	SFA	EUA, 1998–2000	Eficiência parcialmente explicada pela estrutura regulatória
Erbetta e Cave (2007)	DEA	Inglaterra/Gales, 1993–2005	Ganhos de eficiência atribuídos às mudanças regulatórias
García-Sánchez (2006)	DEA	Espanha, 1999	Propriedade não afeta eficiência; densidade é determinante
Saal, Parker e Weyman-Jone (2007)	SFA	Inglaterra/Gales, 1985–2000	Mudança tecnológica acelerou após privatização com regulação independente
Filippini, Hrovatin e Zorić (2008)	SFA	Eslovênia, 1997–2003	Eficiência sensível à metodologia; deseconomias em firmas grandes
Lannier e Porcher (2014)	DEA+SFA	França, 2009	Privadas ligeiramente <i>menos</i> eficientes após controle ambiental
Price <i>et al.</i> (2017)	SFA	Canadá, 2011	Variáveis ambientais mais relevantes que propriedade
Molinos-Senante e Maziotis (2019)	SFA	Inglaterra/Gales	Ciclos regulatórios da OFWAT determinam trajetória de eficiência
Carvalho e Marques (2016)	SFA Bayesi- ana	Portugal	Economias de tamanho e escopo; setor fragmentado é ineficiente
Molinos-Senante, Maziotis e Villegas (2022)	SFA	Chile, 2010–2017	Privadas e concessionárias superam pública em produtividade
Maziotis e Molinos-Senante (2022)	SFA	Chile, 2007–2015	Empresa pública supera privadas em eficiência técnica pura
Nyathikala <i>et al.</i> (2023)	SFA	Índia, 2010–2015	Governança e incentivos mais relevantes que propriedade
Massarutto <i>et al.</i> (2023)	SFA	Itália	Privatização sem regulação efetiva não gera ganhos
Maziotis <i>et al.</i> (2025)	SFA	Chile	Resultados inconclusivos entre privadas totais e concessionárias

Nota: AR = Análise de Regressão; DEA = *Data Envelopment Analysis*; SFA = *Stochastic Frontier Analysis*. Fonte: dados da pesquisa.

3.2.3 Evidências para o Brasil

No Brasil, a literatura apresenta resultados igualmente inconclusivos sobre o efeito da propriedade, mas oferece contribuições importantes sobre os determinantes estruturais da eficiência e o papel do ambiente regulatório. Tupper e Resende (2004) utilizaram DEA para vinte empresas no período 1996–2000 e identificaram economias de densidade como o principal ganho potencial do setor, sem diferenças expressivas por tipo de propriedade. Da Silva E Souza, De Faria e Moreira (2007) estimaram uma função de custo translog por SFA para 164 empresas operando em 2002 e não encontraram diferenças significativas de eficiência entre firmas públicas e privadas, verificando que variáveis ambientais (como a densidade da rede e o índice de perdas) afetam substantivamente os escores estimados. O escore médio de ineficiência das firmas públicas (1,452) foi apenas marginalmente superior ao das privadas (1,320), diferença estatisticamente não robusta.

Sabbioni (2008) estimaram SFA para 1.163 observações no período 2000–2004 e identificaram as economias de escala como a principal fonte potencial de redução de custos, sugerindo que a organização regional da prestação é mais relevante que a natureza jurídica da empresa. Esse resultado é corroborado por Ferro *et al.* (2014), que, em um painel de 127 prestadores cobrindo mais de 70% da população urbana brasileira entre 2003 e 2010, encontraram que empresas regionais e microrregionais apresentam custos menores que as municipais. Os autores verificaram também que a propriedade exerce algum papel no nível de custos, mas que a independência administrativa não é relevante quando se controlam outras características estruturais. Além disso, Ferro *et al.* (2014) documentaram redução da ineficiência a uma taxa de 4,9% ao ano no período analisado, consistente com uma trajetória de aprendizado setorial independente de reformas de propriedade.

Motta e Moreira (2006) utilizaram DEA para 104 empresas no período 1998–2002 e concluíram que a natureza jurídica não afeta os ganhos de produtividade, enquanto empresas regionais se beneficiam mais das economias de escala. Barbosa, Lima e Brusca (2016) analisaram a relação entre governança e eficiência relativa intertemporal dos prestadores brasileiros em relação ao déficit de acesso universal, verificando que a privatização não aumentou a eficiência do setor nem gerou desempenho superior ao das concessionárias públicas. Os autores identificaram ainda que subsídios governamentais para investimento de capital estão associados a menores níveis de eficiência; maiores PIB per capita estão correlacionados a maior eficiência; e que economias de escala e de escopo foram encontradas no serviço de esgoto residencial, enquanto economias de densidade foram encontradas no serviço de água.

Silva e Ledo (2023a) estimaram SFA para o período 2015–2019 e encontraram que empresas mistas de capital aberto são mais eficientes que firmas puramente privadas ou puramente públicas, sugerindo que a estrutura de governança corporativa é o mecanismo relevante, não a propriedade formal. Os mesmos autores identificaram que reguladores com abrangência estadual estão associados a menores níveis de ineficiência, e que convênios de cooperação técnica entre agências

reguladoras e outras instituições contribuem para o desempenho dos prestadores. de Souza Pereira *et al.* (2022) estimaram eficiências para as 95 maiores cidades brasileiras entre 2013 e 2018 com DEA e SFA, encontrando resultados diferenciados por serviço: privadas mais eficientes no abastecimento de água, sem diferença no esgotamento sanitário.

Danelon (2020) estimou para o período 1995–2016 que a independência de gestão está correlacionada com maiores níveis de eficiência, e que as empresas com administração pública direta apresentam os menores índices médios. O autor identificou ainda tendência de regressão tecnológica no setor, com os custos mínimos crescendo a 1,13% ao ano em média, resultado que aponta para a necessidade de reformas institucionais capazes de elevar a produtividade. Esses resultados convergem para a conclusão de Sampaio e Sampaio (2020) de que o ambiente institucional complexo do Brasil, marcado por superposição de competências municipais e estaduais, ausência histórica de regulador federal efetivo e contratos de taxa de retorno, é o principal obstáculo à eficiência do setor.

3.2.4 O Novo Marco Legal e a condição necessária para ganhos de eficiência

A promulgação da Lei nº 14.026/2020 alterou substancialmente esse ambiente institucional (Freire, 2025; Chaves *et al.*, 2024). A lei estabeleceu metas de universalização até 2033 (99% de cobertura de água e 90% de esgoto), consolidou o papel da ANA como supervisora regulatória do setor, exigiu contratos com cláusulas de desempenho mensuráveis em substituição a acordos informais historicamente praticados, e expandiu os incentivos à prestação regionalizada e às parcerias público-privadas (Freire, 2025). Ao criar mecanismos de regulação por comparação (Shleifer, 1985; Tupper; Resende, 2004), o NMLSB estabeleceu, pela primeira vez no Brasil, as condições institucionais que a literatura aponta como necessárias para que a privatização ou a concessão gere ganhos de eficiência: regulação independente, contratos com metas e benchmarking sistemático entre prestadores (Danelon, 2020; Silva; Ledo, 2023a).

A hipótese que motiva este trabalho, derivada dessa revisão, é que o efeito da propriedade sobre a eficiência é condicionado pelo ambiente regulatório. No período aqui analisado (2015–2022), espera-se que a diferença de eficiência entre prestadores públicos e privados seja reduzida ou estatisticamente não significativa, consistente com a literatura prévia ao marco regulatório de 2020 (Da Silva E Souza; De Faria; Moreira, 2007; Ferro *et al.*, 2014; Silva; Ledo, 2023a). A documentação rigorosa dessa linha de base pré-reforma, com metodologia de fronteira estocástica translog em painel, constitui a contribuição empírica central deste artigo: oferecer o ponto de referência contra o qual os efeitos do Novo Marco Legal poderão ser avaliados em estudos futuros (Freire, 2025).

Tabela 3.2 – Síntese dos principais estudos sobre eficiência no setor de saneamento brasileiro.

Autores	Método	Contexto	Principal resultado sobre propriedade/regulação
Tupper e Resende (2004)	DEA	Brasil, 1996–2000	Economias de densidade; sem diferença por propriedade
Motta e Moreira (2006)	DEA	Brasil, 1998–2002	Propriedade não afeta produtividade; empresas regionais mais eficientes
Da Silva E Souza, De Faria e Moreira (2007)	SFA	Brasil, 2002	Sem diferença público/privado; variáveis ambientais relevantes
Sabbioni (2008)	SFA	Brasil, 2000–2004	Economias de escala como principal fonte de ganhos
Ferro <i>et al.</i> (2014)	SFA	Brasil, 2003–2010	Empresas regionais mais eficientes; ineficiência cai 4,9% a.a.
Danelon (2020)	SFA	Brasil, 1995–2016	Independência de gestão correlacionada à eficiência
Silva e Ledo (2023a)	SFA	Brasil, 2015–2019	Empresas mistas de capital aberto mais eficientes; regulação importa
de Souza Pereira <i>et al.</i> (2022)	SFA+DEA	Brasil, 2013–2018	Privadas mais eficientes em água; sem diferença em esgoto
Freire (2025)	SFA	Brasil, 2013–2023	Redução de ineficiência em novas prestadoras pós-Marco Legal

Nota: AR = Análise de Regressão; DEA = *Data Envelopment Analysis*; SFA = *Stochastic Frontier Analysis*. Fonte: dados da pesquisa.

3.3 Metodologia

3.3.1 Base de dados

Os dados utilizados provêm do SNIS, base de acesso público organizada pelo Ministério das Cidades, que reúne informações anuais sobre volumes produzidos, custos operacionais, preços de insumos, extensão de redes e características institucionais dos serviços de água e esgoto em todo o território brasileiro.

A amostra cobre o período de 2015 a 2022, compreendendo oito anos de observação de prestadores que ofertam água e esgoto simultaneamente. Três critérios motivam esse recorte temporal: (i) a disponibilidade consistente do consumo de energia elétrica desagregado por sistema (AG028 e ES028), cuja cobertura se consolida a partir de 2015; (ii) a exigência de painel balanceado do estimador de fronteira estocástica com ineficiência tempo-variante (Battese; Coelli, 1992), que requer a presença de cada prestador em todos os períodos; e (iii) as informações de 2023 e 2024 não foram incorporadas devido à grande quantidade de valores ausentes decorrentes da mudança de base para o SINISA.

Partiu-se do universo de prestadores declarantes e aplicaram-se filtros sequenciais de consistência: exclusão de observações com valores nulos ou negativos em custos, volumes e preços; exclusão de prestadores com classificações inconsistentes de tipo de serviço; e exclusão

de observações com custo residual de outros insumos negativo, situação que ocorre quando subsídios ou receitas financeiras superam os demais componentes do custo variável. Após a filtragem, mantiveram-se apenas os prestadores com série completa de oito anos, resultando em um painel balanceado de 256 prestadores e 2.048 observações. A sensibilidade dos resultados a esse critério é avaliada na seção de testes de robustez, com janelas de seis e cinco anos.

Todas as variáveis monetárias foram deflacionadas pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA), com base em dezembro de 2015, de modo a garantir comparabilidade intertemporal dos valores reais (Meireles, 2018).

3.3.2 Fundamentos teóricos: função de custo variável

A estrutura teórica do modelo segue a tradição de dualidade aplicada à análise de produção em indústrias de rede, cujas raízes remontam a Christensen, Jorgenson e Lau (1973) e foram sistematizadas por Paul (1999). O ponto de partida é o problema de minimização de custo dos prestadores de saneamento: firmas que operam em regime de monopólio natural, atendem demanda inelástica e, portanto, têm como objetivo comportamental a minimização de custo para um nível dado de produto (Kumbhakar; Lovell, 2000; Ferro *et al.*, 2014).

A indústria de saneamento é intensiva em capital físico, com a maior parte da infraestrutura instalada sob o solo (redes de distribuição de água, coletores de esgoto, estações de tratamento), o que implica elevados custos de ajustamento e incapacidade de expansão no curto prazo (Torres; Paul, 2006; Sousa; Torres, 2022). Por essa razão, adota-se uma função de custo variável (ou restrita), na qual o capital físico é tratado como insumo quasi-fixado: as firmas minimizam o custo dos insumos variáveis dado o nível de capital instalado, os preços de insumos, os produtos e as características ambientais (Paul, 1999; Torres; Paul, 2006). Formalmente, a função de custo variável pode ser escrita como:

$$CV = CV(Y, P, \bar{K}, Z), \quad (3.1)$$

onde CV é o custo variável mínimo, Y é o vetor de produtos, P é o vetor de preços dos insumos variáveis, $\bar{K}$ é o capital quasi-fixado e Z é um vetor de variáveis de controle estrutural e institucional. Essa especificação com capital quasi-fixado foi adotada por Torres e Paul (2006) para os EUA e por Sousa e Torres (2022) para o Brasil, ambos utilizando dados de prestadores de água e esgoto.

Para que a função (3.1) seja economicamente bem comportada, três condições de regularidade devem ser satisfeitas (Christensen; Jorgenson; Lau, 1973; Christensen; Greene, 1976):

- (i) Homogeneidade de grau 1 nos preços: se todos os preços dos insumos aumentam na mesma proporção $\lambda > 0$, o custo mínimo deve aumentar na mesma proporção: $CV(Y, \lambda P, \bar{K}, Z) = \lambda \cdot CV(Y, P, \bar{K}, Z)$. Essa condição implica, na especificação translog, as restrições paramétricas $\sum_j \beta_j = 1$, $\sum_j \beta_{jk} = 0$ para todo k , e $\sum_j \beta_{jy} = 0$ (Christensen; Jorgenson; Lau, 1973; Kumbhakar; Lovell, 2000; Freire, 2025).

- (ii) Monotonicidade nos preços: o custo não deve cair quando um preço aumenta ($\partial CV / \partial P_j \geq 0$), condição verificada pelo sinal positivo das participações estimadas nos custos.
- (iii) Concavidade nos preços: a matriz hessiana do custo em relação aos preços deve ser negativa semidefinida, condição verificada empiricamente após a estimação.

As condições (ii) e (iii) dependem dos parâmetros estimados e são verificadas *ex post*. A condição (i), por ser puramente paramétrica, pode ser imposta durante a estimação por meio de reparametrização (Christensen; Greene, 1976; Filippini; Hrovatin; Zorić, 2008; Ferro *et al.*, 2014).

3.3.3 Especificação: função de custo translogarítmica

3.3.3.0.1 Forma funcional

A forma funcional adotada é a translog, introduzida por Christensen, Jorgenson e Lau (1973) como aproximação de segunda ordem de Taylor em logaritmos de qualquer função de custo diferenciável, sem impor restrições sobre as elasticidades de substituição entre insumos ou sobre o padrão de economias de escala. A superioridade da translog sobre a Cobb-Douglas (que restringe essas elasticidades a serem constantes e iguais para firmas de qualquer porte) foi confirmada pelo primeiro grande estudo empírico usando a forma funcional, de Christensen e Greene (1976) para a indústria de energia elétrica norte-americana, e é corroborada para o setor de saneamento brasileiro pelo teste de razão de verossimilhança estimado neste trabalho (LR = 651,92; $p < 0,001$).

3.3.3.0.2 Variável dependente

O Custo Variável é definido como as Despesas de Exploração (DEX) reportadas ao SNIS (FN015), que incluem pessoal próprio, energia elétrica, produtos químicos, serviços de terceiros, água importada e esgoto exportado. Essa variável captura o conjunto dos custos operacionais efetivamente incorridos pelo prestador no processo de produção dos serviços de água e esgoto, sendo a medida padrão de custo variável adotada na literatura brasileira do setor (Da Silva E Souza; De Faria; Moreira, 2007; Ferro *et al.*, 2014; Silva; Ledo, 2023a).

A variável resultante equivale à soma das despesas com os três insumos variáveis da especificação: pessoal ($P_L \cdot L$), energia ($P_E \cdot E$) e outros insumos residuais ($P_O \cdot O$), conforme Torres e Paul (2006) e Vitor de Carvalho Sousa e de Oliveira Torres (2022).

3.3.3.0.3 Imposição de homogeneidade por normalização

A condição (i) é imposta por reparametrização: dividindo-se o custo variável e todos os preços de insumos pelo preço da energia elétrica (P_E), que serve de numerário. Esse procedimento

transforma as três restrições paramétricas em identidades algébricas satisfeitas automaticamente, sem necessidade de otimização com restrições durante a estimação por máxima verossimilhança (Christensen; Jorgenson; Lau, 1973; Christensen; Greene, 1976; Kumbhakar; Lovell, 2000). A técnica foi introduzida por Christensen e Greene (1976) na estimação de funções de custo translog para a indústria de energia elétrica e é padrão na literatura de custos em indústrias de rede, incluindo o setor de água e esgoto (Filippini; Hrovatin; Zorić, 2008; Ferro *et al.*, 2014).

A necessidade da normalização é confirmada empiricamente neste trabalho: na especificação sem numerário, a soma dos coeficientes de primeira ordem dos três preços estimados pelos dados é 0,683, um desvio de 31,7% em relação ao valor unitário exigido pela teoria, tornando indefensável qualquer especificação sem a restrição imposta.

3.3.3.0.4 A equação estimada

A função de custo variável translogarítmica, com custos e preços normalizados por P_E e todas as variáveis explicativas contínuas centralizadas na média amostral antes da estimação, é:

$$\begin{aligned}
 \ln\left(\frac{CV_{it}}{P_{E,it}}\right) = & \beta_0 + \beta_1 \ln\left(\frac{P_{L,it}}{P_{E,it}}\right) + \beta_2 \ln\left(\frac{P_{O,it}}{P_{E,it}}\right) + \beta_3 \ln Y_{A,it} + \beta_4 \ln Y_{E,it} + \beta_5 \ln \bar{K}_{it} \\
 & + \frac{1}{2}\beta_{11} \left[\ln\left(\frac{P_L}{P_E}\right)\right]^2 + \frac{1}{2}\beta_{22} \left[\ln\left(\frac{P_O}{P_E}\right)\right]^2 + \beta_{12} \ln\left(\frac{P_L}{P_E}\right) \ln\left(\frac{P_O}{P_E}\right) \\
 & + \frac{1}{2}\beta_{33} (\ln Y_A)^2 + \frac{1}{2}\beta_{44} (\ln Y_E)^2 + \beta_{34} \ln Y_A \ln Y_E + \frac{1}{2}\beta_{55} (\ln K)^2 \\
 & + \beta_{13} \ln\left(\frac{P_L}{P_E}\right) \ln Y_A + \beta_{14} \ln\left(\frac{P_L}{P_E}\right) \ln Y_E + \beta_{23} \ln\left(\frac{P_O}{P_E}\right) \ln Y_A + \beta_{24} \ln\left(\frac{P_O}{P_E}\right) \ln Y_E \\
 & + \beta_{15} \ln\left(\frac{P_L}{P_E}\right) \ln \bar{K} + \beta_{25} \ln\left(\frac{P_O}{P_E}\right) \ln \bar{K} + \beta_{35} \ln Y_A \ln \bar{K} + \beta_{45} \ln Y_E \ln \bar{K} \\
 & + \delta' \mathbf{z}_{it} + v_{it} + u_{it},
 \end{aligned} \tag{3.2}$$

onde os índices $i = 1, \dots, 256$ e $t = 2015, \dots, 2022$ identificam, respectivamente, o prestador e o ano.

Os coeficientes de primeira ordem $\hat{\beta}_1$ e $\hat{\beta}_2$ representam, avaliados na média amostral (ponto de centralização), as participações dos insumos trabalho e outros no custo variável. A participação implícita da energia é $1 - \hat{\beta}_1 - \hat{\beta}_2$, que resulta positiva e corresponde a aproximadamente 12,8% do custo, valor coerente com a literatura setorial (Ferro *et al.*, 2014; Da Silva E Souza; De Faria; Moreira, 2007). Os coeficientes $\hat{\beta}_3$, $\hat{\beta}_4$ e $\hat{\beta}_5$ são as elasticidades de custo em relação ao volume de água, ao volume de esgoto e à extensão da rede (capital quasi-fixe), respectivamente, também avaliadas no ponto médio da amostra.

Os termos de segunda ordem capturam substituição entre insumos ($\hat{\beta}_{11}$, $\hat{\beta}_{22}$, $\hat{\beta}_{12}$), curvatura nas relações produto–custo ($\hat{\beta}_{33}$, $\hat{\beta}_{44}$, $\hat{\beta}_{34}$, $\hat{\beta}_{55}$) e interações entre preços, produtos e capital ($\hat{\beta}_{13}$ a $\hat{\beta}_{45}$). A simetria de Slutsky é imposta por construção ao restringir $\beta_{ij} = \beta_{ji}$ para todos os pares

de preços. A centralização na média amostral garante que os termos de primeira ordem possam ser interpretados diretamente como elasticidades no ponto médio, sem depender do ponto de avaliação escolhido (Christensen; Greene, 1976; Coelli *et al.*, 2005).

O vetor $\mathbf{z}_{it}$ reúne as variáveis de controle estrutural e institucional descritas na Subseção 3.3.5: o logaritmo natural da densidade da rede ($\ln Dens$), o logaritmo natural da taxa de micromedicação ($\ln Micro$), o índice de perdas em nível ($Loss$), o logaritmo natural da cobertura de esgoto ($\ln Sew$) e as variáveis dummies de tratamento de esgoto, gestão independente, propriedade privada, prestador local, múltiplos prestadores e região geográfica.

Os termos estocásticos $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$ e $u_{it} \geq 0$ compõem o erro composto da Análise de Fronteira Estocástica: v_{it} captura ruído aleatório simétrico (erros de medição, choques exógenos não controlados), enquanto u_{it} é o termo de ineficiência de custo, que mede o quanto o custo observado excede o custo mínimo da fronteira estocástica, dado o nível de produto, os preços e o capital instalado do prestador.

3.3.4 Modelo de ineficiência e estratégia de estimação

A estimação segue o modelo de Battese e Coelli (1992) para fronteiras estocásticas com ineficiência tempo-variante em dados de painel, implementado via Máxima Verossimilhança pelo pacote `frontier` do R (Coelli; Henningsen, 2020). A escolha da SFA em detrimento da DEA fundamenta-se em três razões: (i) a SFA decompõe o resíduo em ineficiência e ruído aleatório, separando erros de medição e choques exógenos da ineficiência genuína (Kumbhakar; Parmeter; Zelenyuk, 2021); (ii) a estrutura paramétrica permite a estimação de elasticidades de custo e economias de escala com inferência estatística formal; e (iii) o valor estimado de $\gamma = 0,935$ confirma que 93,5% da variância do resíduo composto é atribuível à ineficiência, validando a abordagem de fronteira frente à regressão ordinária (Coelli *et al.*, 2005; Kumbhakar; Parmeter; Zelenyuk, 2021).

No modelo de Battese e Coelli (1992), a ineficiência de custo evolui ao longo do tempo de forma sistemática segundo:

$$u_{it} = \exp[-\eta (t - T)] u_i, \quad u_i \sim \mathcal{N}^+(\mu, \sigma_u^2), \quad (3.3)$$

onde T é o último período do painel e η é um parâmetro escalar estimado pelos dados: $\eta > 0$ implica que a ineficiência decresce ao longo do tempo (aprendizado setorial), $\eta < 0$ implica que cresce, e $\eta = 0$ colapsa na especificação tempo-invariante de Battese e Coelli (1992). A escolha do modelo tempo-variante sobre o tempo-invariante é confirmada por teste de razão de verossimilhança (LR = 49,87; $p < 0,001$).

Os parâmetros de variância identificáveis são:

$$\sigma^2 = \sigma_u^2 + \sigma_v^2, \quad \gamma = \frac{\sigma_u^2}{\sigma^2}, \quad \lambda = \frac{\sigma_u}{\sigma_v}. \quad (3.4)$$

O escore de eficiência técnica de custo de cada prestador em cada período é obtido como a esperança condicional (Coelli *et al.*, 2005):

$$ET_{it} = E[\exp(-u_{it}) \mid \varepsilon_{it}] \in (0, 1], \quad (3.5)$$

onde $\varepsilon_{it} = v_{it} + u_{it}$ é o resíduo composto observado. Um valor de $ET_{it} = 1$ indica que o prestador opera sobre a fronteira de custo mínimo; valores menores indicam ineficiência de custo na proporção $1 - ET_{it}$.

Para seleção de especificação, foram estimados quatro modelos: Cobb-Douglas tempo-invariante (CD-TI), Cobb-Douglas tempo-variante (CD-TV), Translog tempo-invariante (TL-TI) e Translog tempo-variante (TL-TV). O modelo selecionado, Translog Tempo-Variante, apresenta os melhores critérios de informação (AIC = -3245,32; BIC = -3037,20) e é superior a todos os modelos concorrentes pelos testes LR reportados na Tabela 3.8.

3.3.5 Variáveis

A Tabela 3.3 descreve todas as variáveis da especificação com a correspondência aos campos do SNIS.

Tabela 3.3 – Descrição das variáveis e correspondência com o SNIS.

Símbolo	Variável	SNIS	Descrição
<i>Custo (variável dependente)</i>			
CV_{it}	Custo Variável (R\$)	FN015	Despesas com pessoal, energia e outros (produtos químicos, serviços de terceiros, água importada e esgoto exportado); deflacionado pelo IPCA (base: dez./2015)
<i>Produtos (outputs)</i>			
Y_A	Volume de água (mil m ³ /ano)	AG006	Volume anual de água disponível para consumo
Y_E	Volume de esgoto (mil m ³ /ano)	ES005+ES006+ES012	Volume coletado, tratado e exportado bruto
<i>Preços dos insumos variáveis</i>			
P_L	Preço do trabalho (R\$/empregado)	FN010/FN026	Despesa com pessoal próprio dividida pelo número de empregados
P_E	Preço da energia (R\$/kWh)	FN013/(AG028+ES028)	Despesa com energia dividida pelo consumo total; serve de <i>numerário</i> para imposição de homogeneidade
P_O	Preço de outros insumos (R\$/km)	(CV–FN010–FN013)/ (AG005+ES004)	Custo residual do CV após deduzir trabalho e energia, dividido pela extensão da rede como proxy das quantidades dos demais insumos; seguindo Da Silva E Souza, De Faria e Moreira (2007) e Ferro <i>et al.</i> (2014)
<i>Capital quasi-fixo</i>			
$\bar{K}$	Extensão da rede (km)	AG005+ES004	Comprimento total das redes de água e esgoto; proxy do estoque de capital instalado, tratado como insumo quasi-fixo seguindo Torres e Paul (2006) e Sousa e Torres (2022)
<i>Controles estruturais (z_{it})</i>			
$Dens$	Densidade (lig/km)	AG002/AG005	Ligações ativas por km de rede de água; captura economias de densidade (Ferro <i>et al.</i> , 2014; Torres; Paul, 2006)
$Micro$	Micromedicação	AG004/AG002	Proporção de ligações micromedidas; proxy de qualidade do controle de perdas
$Loss$	Índice de perdas (%)	IN013	Percentual de água produzida não faturada
Sew	Cobertura de esgoto (%)	ES026/GE06B	Proporção da população urbana atendida com coleta de esgoto
<i>Dummies institucionais</i>			
$Treat$	Tratamento de esgoto	—	1 se proporção tratada > 20%
$Indep$	Gestão independente	Natureza Jurídica	1 se não é administração pública direta (Ferro <i>et al.</i> , 2014; Silva; Ledo, 2023a)
$Priv$	Propriedade privada	Natureza Jurídica	1 se empresa privada
$Local$	Prestador local	Abrangência	1 se abrangência local (municipal)
$Multi$	Múltiplos prestadores	Código município	1 se o município possui mais de um prestador na amostra
<i>Dummies regionais (ref.: Centro-Oeste)</i>			
NO, NE, SE, SO	Regiões	Estado	Norte, Nordeste, Sudeste e Sul

Nota: Valores monetários em reais constantes de dez./2015 (deflator: IPCA). Todas as variáveis contínuas da função de custo foram centralizadas na média amostral antes da estimação.

Duas decisões metodológicas merecem destaque adicional. Primeira, a extensão da rede

$\bar{K}$) é o quasi-fixo natural do setor: ela representa o estoque de capital instalado que determina a capacidade operacional e não pode ser ajustado no curto prazo, justificando seu tratamento como argumento da função de custo variável separado dos preços de insumos (Torres; Paul, 2006; Paul, 1999; Sousa; Torres, 2022). Segunda, o preço de outros insumos (P_O) é calculado como custo residual dividido pela extensão da rede, procedimento que captura o custo médio por km dos insumos não observados individualmente (químicos, materiais, terceiros) e que foi adotado originalmente por Da Silva E Souza, De Faria e Moreira (2007) e replicado por Ferro *et al.* (2014) com os dados do SNIS.

3.4 Resultados

3.4.1 Amostra e estatísticas descritivas

A Figura 3.1 apresenta a distribuição geográfica dos 254 prestadores que compõem o painel balanceado, com círculos proporcionais à população média atendida com abastecimento de água no período 2015–2022. A amostra cobre aproximadamente 61,5% da população brasileira e reflete a concentração histórica da prestação de serviços de saneamento: a Região Sudeste responde por 163 dos 254 prestadores (64,2% da amostra) e concentra os maiores operadores do país, incluindo a Companhia de Saneamento Básico do Estado de São Paulo (SABESP), identificada pelo maior círculo verde. As regiões Norte e Nordeste, que acumulam os maiores déficits de cobertura, são representadas por poucos prestadores de grande porte, o que também é reflexo de que o painel balanceado, ao exigir série completa de oito anos, seleciona naturalmente os operadores com maior estabilidade institucional e capacidade de reporte ao SNIS.

A Tabela 3.4 traz as estatísticas descritivas das variáveis em valores originais, desagregadas por tipo de prestador. A heterogeneidade da amostra é marcante em ambos os grupos: entre os prestadores não privados ($N = 1.731$), o custo variável médio é de R\$ 125,2 milhões, mas o desvio padrão supera R\$ 544 milhões, evidenciando a convivência de pequenos prestadores municipais com grandes companhias estaduais. Os prestadores privados ($N = 301$) apresentam escala média menor, com custo variável médio de R\$ 54,7 milhões, embora também exibam elevada dispersão.

O mesmo padrão de assimetria de escala se observa nos produtos: o volume médio de água dos prestadores não privados (50,3 mil m³/ano) é aproximadamente o dobro do verificado entre os privados (24,9 mil m³/ano), diferença que se repete no produto esgoto (32,9 mil contra 16,9 mil m³/ano) e na extensão da rede (3.178 km contra 1.878 km). Esse resultado é consistente com o fato de que as companhias estaduais, predominantemente públicas, operam em escala regional. Em termos de eficiência operacional, os prestadores privados apresentam índice médio de perdas de 23,2%, valor inferior ao dos não privados (30,5%), sugerindo melhor desempenho operacional nessa dimensão. A taxa de micromedição reforça essa leitura: 99% entre os privados ante 95% entre os não privados, indicando maior cobertura de medição individual no primeiro

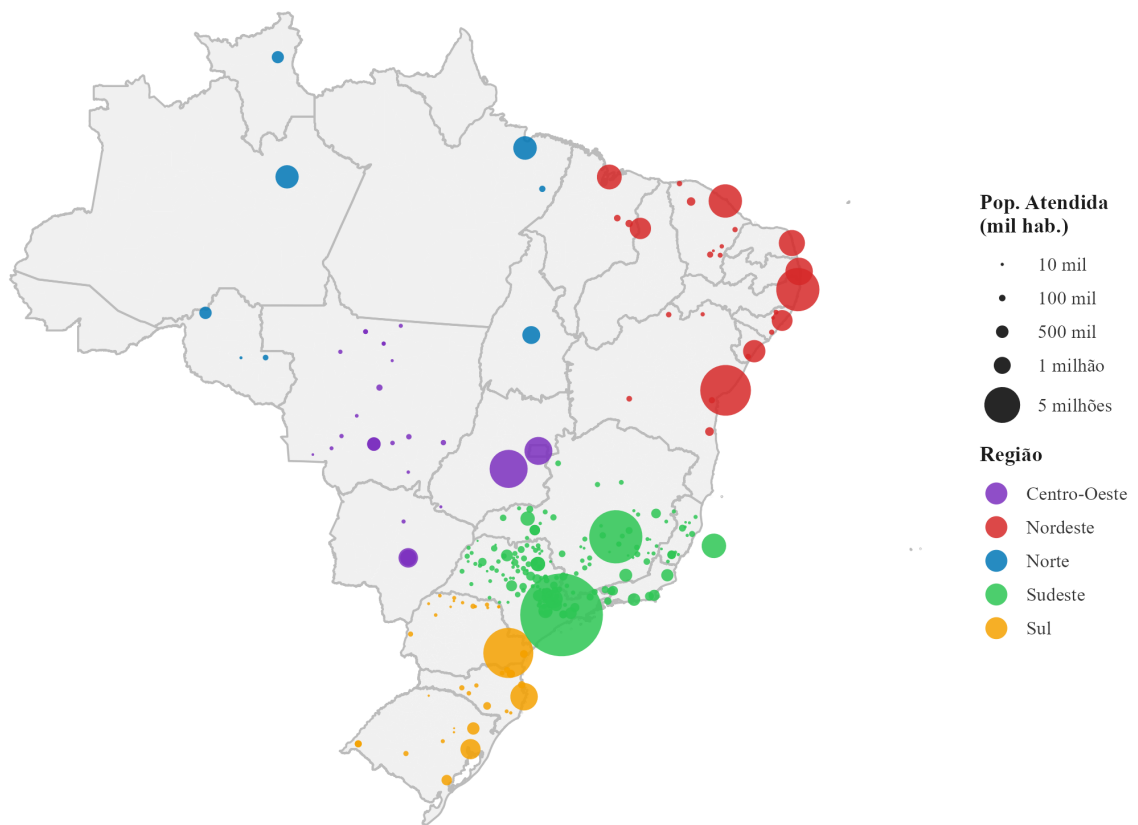


Figura 3.1 – Distribuição geográfica da amostra: prestadores por região e população atendida (2015–2022).
 Nota: Cada círculo representa um prestador posicionado no centroide do município-sede. Tamanho proporcional à média da população atendida com abastecimento de água (AG001) no período. Fonte: SNIS (2015–2022); elaboração própria.

grupo. O preço de outros insumos, por sua vez, é substancialmente mais elevado entre os prestadores privados (R\$ 14.683/km contra R\$ 9.237/km), o que pode refletir diferenças na estrutura de terceirização ou no mix de serviços contratados. Os preços do trabalho e da energia são relativamente similares entre os grupos, com leve vantagem dos privados no preço da energia (R\$ 468/kWh contra R\$ 517/kWh).

Tabela 3.4 – Estatísticas Descritivas das Variáveis por Tipo de Prestador (Valores Originais)

Variável	Privado (N = 301)				Não Privado (N = 1731)			
	Média	DP	Mín.	Máx.	Média	DP	Mín.	Máx.
Custo Variável (R\$ 1.000)	54685472.73	185125604.96	424130.20	2220551194.73	125160101.91	544449623.65	68340.30	7607659810.52
Produto Água (1.000 m ³ /ano)	24881.65	69119.23	195.10	740678.90	50284.35	219524.47	50.00	2906681.86
Produto Esgoto (1.000 m ³ /ano)	16902.67	55184.51	95.20	675304.02	32949.34	170749.05	5.60	2518654.02
Extensão da Rede — quasi-fixo (km)	1878.16	6891.72	28.68	84494.40	3177.94	13124.27	6.00	155138.10
Preço do Trabalho (R\$/empregado)	54827.24	19602.91	6154.19	131219.62	55558.10	34169.28	5663.36	403528.01
Preço da Energia (R\$/kWh)	468.29	129.45	141.28	1937.16	516.73	339.88	5.16	4625.37
Preço de Outros Insumos (R\$/km)	14683.08	8510.87	2449.00	49433.57	9237.48	8375.09	103.71	57884.66
Densidade da Rede (lig/km)	71.93	18.84	26.87	132.74	75.95	32.58	4.06	285.80
Micromedição (%)	0.99	0.02	0.84	1.00	0.95	0.13	0.10	1.00
Índice de Perdas (%)	23.17	14.38	0.00	67.42	30.46	16.85	0.00	98.38

Nota: N = 2032 observações de 254 prestadores ao longo de 8 anos (2015–2022). Custo Variável = FN015. Produtos em 1.000 m³/ano. Extensão da Rede ($\bar{K}$) é o capital quasi-fixo. Fonte: dados da pesquisa.

A Tabela 3.5 apresenta as mesmas variáveis após as transformações logarítmicas e a

normalização pelo preço da energia necessárias à estimação da fronteira. As diferenças entre grupos observadas em valores originais se mantêm, porém atenuadas pela compressão logarítmica. O custo variável normalizado é em média ligeiramente superior entre os prestadores privados (10,39 contra 9,95), o que à primeira vista pode parecer contraintuitivo, mas reflete a composição da amostra: os prestadores privados apresentam menor dispersão (DP de 1,72 contra 2,13), sugerindo maior homogeneidade interna do grupo. Os produtos água e esgoto também são em média maiores em log para os privados, embora a diferença seja modesta (0,19 e 0,30, respectivamente), o que contrasta com a diferença mais expressiva observada nos valores originais e indica que a assimetria de escala é impulsionada por grandes outliers no grupo não privado.

A variável de preço relativo de outros insumos exibe a maior diferença entre grupos: 3,31 para privados contra 2,60 para não privados, com dispersão também mais elevada no grupo não privado (DP de 1,15 contra 0,69), reforçando a heterogeneidade nas estruturas de custo dos prestadores públicos. O logaritmo da micromedicação é próximo de zero para os privados (0,003) e ligeiramente negativo para os não privados (-0,054), consistente com as taxas de 99% e 95% observadas nos valores originais. O índice de perdas, não logaritimizado, mantém a diferença já destacada: 23,2% entre privados e 30,5% entre não privados.

A Tabela 3.6 descreve as variáveis institucionais e regionais. Entre as variáveis institucionais, 85,9% das observações correspondem a prestadores com gestão independente e 14,8% são de propriedade privada. Destaca-se que 90,6% das observações são de prestadores de abrangência local, o que é relevante para a interpretação dos resultados, dado que prestadores regionais operam sob lógicas de escala e governança distintas. Apenas 2,4% das observações envolvem municípios com múltiplos prestadores, minimizando potenciais problemas de sobreposição de serviço. A concentração no Sudeste (64,2% das observações) reforça a necessidade de controles regionais na especificação, especialmente considerando que a região Norte, com apenas 3,2% das observações, é justamente aquela com menor cobertura de saneamento no país.

Tabela 3.5 – Estatísticas Descritivas das Variáveis por Tipo de Prestador (Transformação Logarítmica)

Variável	Privado (N = 301)				Não Privado (N = 1731)			
	Média	DP	Mín.	Máx.	Média	DP	Mín.	Máx.
$\ln(CV/P_E)$	10.3877	1.7223	6.7541	15.2381	9.9547	2.1330	4.3879	16.8162
$\ln(Y_{\text{água}})$	8.8398	1.5358	5.2735	13.5153	8.6461	1.8563	3.9120	14.8825
$\ln(Y_{\text{esgoto}})$	8.5023	1.6401	4.5560	13.4229	8.2050	1.8932	1.7228	14.7392
$\ln(K)$	6.4334	1.3791	3.3562	11.3444	6.1593	1.6614	1.7918	11.9521
$\ln(P_L/P_E)$	4.7111	0.5110	2.3994	5.6932	4.6655	0.7368	1.5197	8.5831
$\ln(P_O/P_E)$	3.3069	0.6945	1.5712	5.0998	2.5950	1.1457	-2.1628	7.3291
$\ln(\text{Densidade})$	4.2409	0.2694	3.2911	4.8884	4.2362	0.4695	1.4012	5.6553
$\ln(\text{Micromedicação})$	0.0031	0.0235	-0.1629	0.0100	-0.0538	0.2277	-2.2097	0.0100
Índice de Perdas (%)	23.1732	14.3841	0.0000	67.4200	30.4635	16.8525	0.0000	98.3800

Nota: Todas as variáveis, exceto o Índice de Perdas, estão em forma logarítmica. Custos e preços normalizados pelo preço da energia (P_E) como numerário. Variáveis explicativas centralizadas na média amostral antes da estimação. Fonte: dados da pesquisa.

Tabela 3.6 – Distribuição das Variáveis Categóricas

Variável	Observações	
	N	Percentual (%)
Variáveis Institucionais		
Tratamento de Esgoto (>20%)	1651	81.25
Gestão Independente	1745	85.88
Propriedade Privada	301	14.81
Prestador Local	1840	90.55
Município com Múltiplos Prestadores	48	2.36
Variáveis Regionais		
Região Norte	64	3.15
Região Nordeste	216	10.63
Região Sudeste	1304	64.17
Região Sul	280	13.78
Região Centro-Oeste	168	8.27

Nota: Total de observações: 2032. Fonte: dados da pesquisa.

3.4.2 Estimativas da fronteira de custo

A Tabela 3.7 contém os resultados do modelo Translog Tempo-Variante. Os parâmetros de variância confirmam a adequação da abordagem de fronteira estocástica: $\gamma = 0,9346$ indica que 93,5% da variância do termo composto é atribuível à ineficiência (u_{it}), e não ao ruído aleatório (v_{it}), e a razão $\lambda = \sigma_u/\sigma_v = 3,78$ reforça que a assimetria do resíduo é estatisticamente relevante. O modelo foi confirmado como superior à Cobb-Douglas pelo teste LR (LR = 651,92; $p < 0,001$) e à especificação tempo-invariante (LR = 49,87; $p < 0,001$).

A tabela apresenta três especificações do modelo, estimadas com o mesmo conjunto de parâmetros da fronteira translog mas diferindo nos controles incluídos no vetor $\mathbf{z}_{it}$. A especificação (1), completa, inclui os quatro controles estruturais contínuos (densidade da rede, micromedição, índice de perdas e cobertura de esgoto) e as quatro dummies regionais, além das cinco dummies institucionais presentes em todas as colunas. A especificação (2) remove os controles estruturais e mantém apenas as dummies regionais e institucionais, permitindo avaliar se os coeficientes institucionais são sensíveis à omissão de variáveis de contexto operacional. A especificação (3) remove também as dummies regionais, restando apenas as variáveis institucionais como controles além dos termos da fronteira. Essa progressão de especificações serve a dois propósitos: verificar a robustez dos coeficientes de interesse (em particular o da propriedade privada) à inclusão ou exclusão de controles, e isolar o papel das variáveis estruturais na determinação dos custos. Os coeficientes reportados a seguir referem-se à especificação (1), que é a preferida por controlar adequadamente por heterogeneidade operacional e regional.

As elasticidades de custo de primeira ordem, avaliadas no ponto médio da amostra, revelam que o trabalho ($\hat{\beta}_1 = 0,431$) e os outros insumos ($\hat{\beta}_2 = 0,438$) são os componentes mais relevantes

Tabela 3.7 – Resultados da Análise de Fronteira Estocástica Modelo Translog Tempo-Variante.

Variável	(1)		(2)		(3)	
	Estimativa	EP	Estimativa	EP	Estimativa	EP
Intercepto	7.9483***	(0.1137)	9.5038***	(0.1063)	9.7522***	(0.0985)
$\ln(P_L/P_E)$	0.4307***	(0.0077)	0.4045***	(0.0086)	0.4081***	(0.0087)
$\ln(P_O/P_E)$	0.4383***	(0.0062)	0.4585***	(0.0069)	0.4578***	(0.0071)
$\ln(Y_{\text{água}})$	0.1139***	(0.0181)	0.1991***	(0.0151)	0.2090***	(0.0155)
$\ln(Y_{\text{esgoto}})$	-0.0028	(0.0122)	0.0881***	(0.0116)	0.0859***	(0.0118)
$\ln(\bar{K})$	0.8698***	(0.0185)	0.6232***	(0.0143)	0.6159***	(0.0147)
$\frac{1}{2}[\ln(P_L/P_E)]^2$	0.1141***	(0.0108)	0.0852***	(0.0117)	0.0900***	(0.0118)
$\frac{1}{2}[\ln(P_O/P_E)]^2$	0.1523***	(0.0070)	0.1464***	(0.0077)	0.1491***	(0.0079)
$\ln(P_L/P_E) \times \ln(P_O/P_E)$	-0.1203***	(0.0072)	-0.1059***	(0.0079)	-0.1096***	(0.0080)
$\frac{1}{2}[\ln(Y_{\text{água}})]^2$	0.0657**	(0.0223)	-0.0030	(0.0238)	-0.0011	(0.0252)
$\frac{1}{2}[\ln(Y_{\text{esgoto}})]^2$	-0.0102	(0.0099)	0.0164	(0.0105)	0.0168	(0.0105)
$\ln(Y_{\text{água}}) \times \ln(Y_{\text{esgoto}})$	-0.0379***	(0.0102)	-0.0450***	(0.0107)	-0.0463***	(0.0109)
$\frac{1}{2}[\ln(\bar{K})]^2$	0.0216	(0.0187)	-0.0149	(0.0205)	-0.0098	(0.0206)
$\ln(P_L/P_E) \times \ln(Y_{\text{água}})$	0.0051	(0.0111)	0.0019	(0.0121)	0.0012	(0.0125)
$\ln(P_L/P_E) \times \ln(Y_{\text{esgoto}})$	0.0367***	(0.0080)	0.0550***	(0.0085)	0.0540***	(0.0086)
$\ln(P_O/P_E) \times \ln(Y_{\text{água}})$	-0.0425***	(0.0097)	-0.0305**	(0.0105)	-0.0294**	(0.0110)
$\ln(P_O/P_E) \times \ln(Y_{\text{esgoto}})$	-0.0044	(0.0060)	-0.0262***	(0.0062)	-0.0263***	(0.0063)
$\ln(P_L/P_E) \times \ln(\bar{K})$	-0.0619***	(0.0115)	-0.0762***	(0.0124)	-0.0754***	(0.0126)
$\ln(P_O/P_E) \times \ln(\bar{K})$	0.0558***	(0.0102)	0.0596***	(0.0109)	0.0592***	(0.0112)
$\ln(Y_{\text{água}}) \times \ln(\bar{K})$	-0.0432*	(0.0187)	0.0350	(0.0200)	0.0356	(0.0205)
$\ln(Y_{\text{esgoto}}) \times \ln(\bar{K})$	0.0374**	(0.0122)	0.0219	(0.0132)	0.0174	(0.0131)
Controles Estruturais						
$\ln(\text{Densidade})$	0.3756***	(0.0193)	—	—	—	—
$\ln(\text{Micromedicação})$	0.0677	(0.0387)	—	—	—	—
Índice de Perdas	-0.0011***	(0.0003)	—	—	—	—
Cobertura de Esgoto	-0.0004	(0.0004)	—	—	—	—
Dummies Institucionais						
Tratamento de Esgoto	-0.0278*	(0.0140)	-0.0717***	(0.0151)	-0.0777***	(0.0152)
Gestão Independente	0.1021***	(0.0248)	0.1547***	(0.0234)	0.1356***	(0.0241)
Propriedade Privada	-0.0268	(0.0237)	0.0221	(0.0246)	-0.0197	(0.0252)
Prestador Local	0.0274	(0.0645)	-0.1995	(0.1157)	-0.2994**	(0.1160)
Múltiplos Prestadores	0.3160***	(0.0552)	0.3337***	(0.0913)	0.1637*	(0.0666)
Dummies Regionais (ref.: Centro-Oeste)						
Norte	0.3219***	(0.0607)	0.3400***	(0.0902)	—	—
Nordeste	0.1034*	(0.0422)	0.2217***	(0.0466)	—	—
Sudeste	0.1481***	(0.0314)	0.1788***	(0.0322)	—	—
Sul	0.1062**	(0.0390)	0.0614	(0.0438)	—	—
Parâmetros de Variância						
σ^2	0.1126		0.1814		0.2114	
γ	0.9346		0.9510		0.9578	
$\lambda = \sigma_u/\sigma_v$	3.7807		4.4032		4.7620	
σ_u	0.3244		0.4153		0.4500	
σ_v	0.0858		0.0943		0.0945	
Estatísticas do Modelo						
Log-Verossimilhança	1648.49		1425.34		1405.45	
AIC	-3222.98		-2784.67		-2752.89	
BIC	-3015.16		-2599.32		-2590.00	
Eficiência Técnica Média	0.7542		0.6898		0.6675	

Notas: Erros padrão entre parênteses. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$. Variável dependente: $\ln(CV/P_E)$. CV = FN015. Preços normalizados pelo preço da energia (P_E) como numerário. Todas as variáveis explicativas contínuas foram centralizadas na média amostral. Especificação (1): modelo completo com controles estruturais e dummies regionais. Especificação (2): sem controles estruturais, mantém dummies regionais. Especificação (3): sem controles estruturais e sem dummies regionais. As variáveis institucionais são mantidas em todas as especificações. Painel balanceado com 254 prestadores e 8 anos (2015–2022). Fonte: elaboração própria.

da estrutura de custos dos prestadores brasileiros, com a energia respondendo pelo complemento ($1 - 0,431 - 0,438 = 0,131$), dado que os preços estão normalizados e a homogeneidade impõe que as participações somem 1. Esses valores são coerentes com os encontrados por Ferro *et al.* (2014) e Da Silva E Souza, De Faria e Moreira (2007) para o mesmo setor.

A elasticidade de custo em relação ao volume de água ($\hat{\beta}_3 = 0,114$) é positiva e significativa: um aumento de 1% no volume de água produzida e distribuída está associado a um aumento estimado de 0,101% no custo variável, na média. Esse resultado indica que a expansão da produção de água eleva os custos de forma menos que proporcional, sugerindo economias de escala no abastecimento. O volume de esgoto ($\hat{\beta}_4 = -0,0017$), por sua vez, não apresenta efeito significativo sobre os custos no ponto médio, resultado que pode refletir as especificidades do esgotamento sanitário brasileiro, onde o custo do serviço é frequentemente subsidiado cruzado pelo abastecimento de água. A extensão da rede ($\hat{\beta}_5 = 0,870$), o capital quasi-fixo, apresenta a maior elasticidade de todas as variáveis, confirmando que a infraestrutura instalada é o principal determinante do nível de custos dos prestadores e que a expansão da rede, necessária para a universalização, terá impacto expressivo sobre os custos operacionais.

Entre as variáveis de controle estrutural, a densidade da rede ($\hat{\beta}_{Dens} = 0,376$; $p < 0,001$) apresenta o sinal positivo esperado: prestadores que operam em redes mais densas (mais ligações por km) incorrem em maiores custos operacionais de manutenção e atendimento, embora também se beneficiem de economias de densidade no longo prazo. O índice de perdas mostra coeficiente negativo e significativo ($\hat{\beta}_{Loss} = -0,001$; $p < 0,001$), indicando que prestadores com maiores perdas tendem a apresentar menores custos variáveis correntes, o que é um resultado à primeira vista contraintuitivo, mas consistente com a literatura (Da Silva E Souza; De Faria; Moreira, 2007): investimentos em redução de perdas requerem desembolsos de capital que não integram o custo variável, de modo que prestadores que postergam esse investimento exibem custos correntes mais baixos a um custo de qualidade e eficiência futura.

No que diz respeito às variáveis institucionais, destaca-se que a propriedade privada não apresenta coeficiente estatisticamente significativo ($\hat{\beta}_{Priv} = -0,027$; $p = 0,26$), resultado que, à luz da revisão de literatura, confirma a hipótese de que no período analisado, a diferença de custos entre prestadores públicos e privados é nula quando se controla pela estrutura do serviço, pela região e pelas características operacionais. Ao contrário, a gestão independente eleva significativamente os custos ($\hat{\beta}_{Indep} = 0,102$; $p < 0,001$), o que reflete os maiores custos de estrutura administrativa e de pessoal das autarquias e empresas de economia mista em relação à administração pública direta. Os municípios com múltiplos prestadores apresentam custos sistematicamente maiores ($\hat{\beta}_{Multi} = 0,316$; $p < 0,001$), possivelmente por conta da sobreposição de infraestruturas. As dummies regionais confirmam que todos os prestadores fora do Centro-Oeste (região de referência) incorrem em custos estruturais mais elevados, com o Norte apresentando a maior diferença ($\hat{\beta}_{Norte} = 0,322$; $p < 0,001$).

Tabela 3.8 – Estatísticas de Eficiência Técnica por Modelo.

Modelo	N	Média	Mediana	DP	Mín.	Máx.
CD Tempo-Invariante	2032	0.7389	0.7304	0.1276	0.4444	0.9898
CD Tempo-Variante	2032	0.7396	0.7302	0.1283	0.4329	0.9902
TL Tempo-Invariante	2032	0.7514	0.7477	0.1262	0.4409	0.9908
TL Tempo-Variante	2032	0.7542	0.7488	0.1272	0.4111	0.9920

Nota: Eficiência varia de 0 (completamente ineficiente) a 1 (plenamente eficiente).

Fonte: elaboração própria.

Tabela 3.9 – Eficiência Técnica por Região — Modelo Translog Tempo-Variante.

Região	Observações	Prestadores	Eficiência Média
Norte	64	8	0.8410
Centro-Oeste	168	21	0.8037
Sudeste	1304	163	0.7499
Sul	280	35	0.7398
Nordeste	216	27	0.7345

Fonte: elaboração própria.

3.4.3 Eficiência técnica média e comparação de modelos

A Tabela 3.8 compara os escores de eficiência técnica estimados pelos quatro modelos alternativos. A eficiência média varia de 0,738 (Cobb-Douglas Tempo-Invariante) a 0,754 (Translog Tempo-Variante), e a hierarquia entre modelos é consistente: a especificação translog supera a Cobb-Douglas e a versão tempo-variante supera a tempo-invariante em todos os critérios. A eficiência mediana do modelo selecionado é de 0,750, indicando que metade dos prestadores poderia reduzir seus custos em pelo menos 25,0% sem alterar seus níveis de produto, caso alcançasse a fronteira de melhor prática. A variação dos escores (mínimo de 0,411; máximo de 0,992) aponta para dispersão expressiva, consistente com a heterogeneidade estrutural da amostra.

A distribuição regional da eficiência, que consta na Tabela 3.9, revela um padrão interessante de ser analisado. Os prestadores do Norte apresentam a maior eficiência média (0,841), seguidos pelo Centro-Oeste (0,804), enquanto Nordeste (0,734) e Sul (0,740) registram os menores escores. Esses resultados não contradizem os coeficientes regionais positivos estimados na fronteira (Tabela 3.7), que indicam custos estruturalmente mais elevados fora do Centro-Oeste: eficiência técnica mede o afastamento de cada prestador em relação à sua fronteira condicional, não o nível absoluto de custos. Cabe notar, no entanto, que o Norte é representado por apenas 8 prestadores na amostra, provavelmente aqueles de maior capacidade de informe continuado ao SNIS na região, o que sugere seleção positiva (ou viés de sobrevivência) acentuada. A comparação mais informativa para fins regulatórios é a intra-regional, e não entre regiões com tamanhos amostrais tão distintos.

Tabela 3.10 – Evolução da Eficiência Técnica ao Longo do Tempo — Modelo Translog Tempo-Variante.

Ano	Eficiência Média	Eficiência Mediana	N
2015	0.7414	0.7344	254
2016	0.7451	0.7385	254
2017	0.7488	0.7425	254
2018	0.7525	0.7464	254
2019	0.7561	0.7503	254
2020	0.7597	0.7542	254
2021	0.7633	0.7580	254
2022	0.7668	0.7617	254

Fonte: elaboração própria.

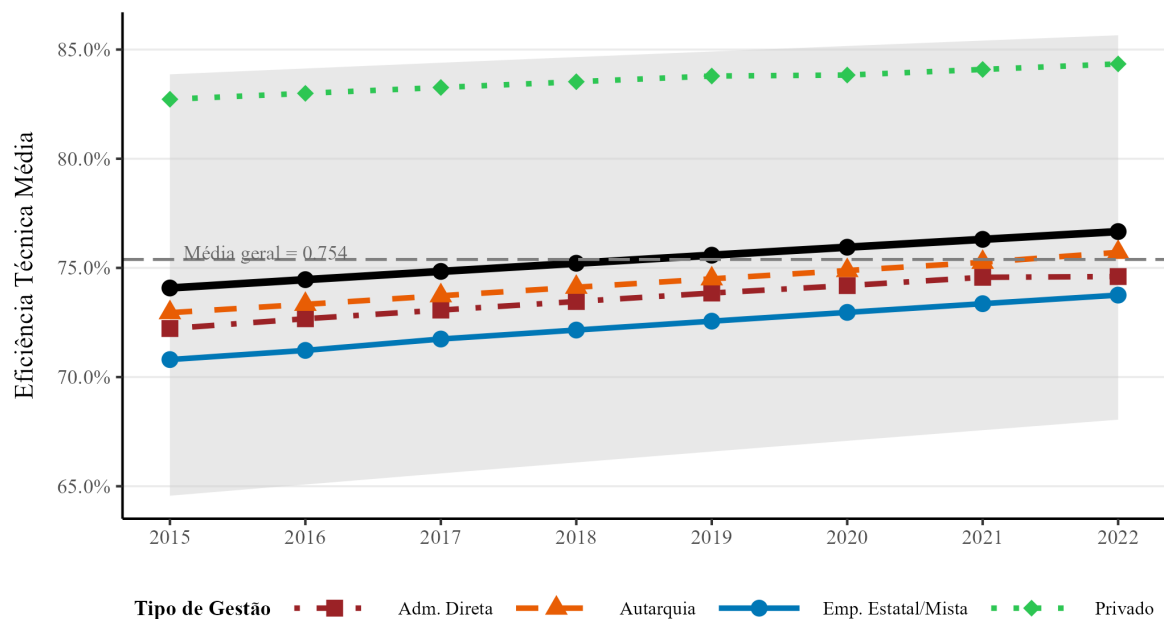


Figura 3.2 – Evolução da eficiência técnica média por tipo de gestão – Modelo Translog Tempo-Variante (2015-2022).

Nota: Linha preta espessa: média geral da amostra. Faixa cinza: intervalo interquartil (P25–P75). Linha tracejada cinza: média do período completo (0,754). Fonte: elaboração própria.

A evolução temporal, apresentada na Tabela 3.10 e ilustrada na Figura 3.2, revela tendência de crescimento contínuo da eficiência média ao longo do período, de 0,741 em 2015 a 0,767 em 2022, um ganho acumulado de 2,6 pontos percentuais em oito anos. Esse padrão é capturado pelo parâmetro $\eta > 0$ do modelo tempo-variante de Battese e Coelli (1992), que implica que os prestadores da amostra, em média, aprenderam a operar mais próximo da fronteira de custo ao longo do tempo. O ritmo de melhoria é modesto mas consistente, e não apresenta ruptura no período, em particular, não há aceleração identificável nos anos mais próximos de 2020, quando o NMLSB foi promulgado.

A Figura 3.2 acrescenta uma dimensão importante ao revelar a heterogeneidade do desempenho por tipo de gestão. Os prestadores privados mantêm eficiência consistentemente

superior à média ao longo de todo o período (cerca de 83–85%), enquanto as empresas estatais e mistas situam-se sistematicamente abaixo da média geral (70–74%). Autarquias e administrações diretas apresentam desempenho inferior à média geral durante a maior parte do período, convergindo para ela apenas no final do horizonte analisado. Esse gradiente entre tipos de gestão, contudo, não se traduz em significância estatística no coeficiente de propriedade privada da fronteira de custo, pois a vantagem das privadas é absorvida parcialmente pela estrutura operacional dessas empresas (que tendem a atender mercados mais densos e com menores perdas) e pelos controles regionais e de produto incluídos na especificação.

A Tabela 3.11 e a Tabela 3.12 apresentam os cinco prestadores com maior e menor eficiência média, respectivamente, no período. Entre os mais eficientes, predominam empresas estatais de pequeno e médio porte: o Departamento de Esgoto e Água de Guaíra (SP), com escore médio de 0,991, e a Companhia de Água e Esgoto do Ceará (CAGECE), com 0,983, sugerindo que a eficiência de custo não é monopólio de grandes operadores nem de empresas privadas. Entre os menos eficientes, a Prefeitura Municipal de Doutor Ricardo (RS) apresenta o menor escore médio da amostra (0,434), seguida pelo Serviço de Água, Esgoto e Limpeza Pública de Muriaé (MG), e pela Prefeitura Municipal de Governador Lindenberg (ES). A Companhia de Saneamento Ambiental do Maranhão (CAEMA), que atende mais de 2,3 milhões de habitantes, figura em 4º lugar com escore de 0,482, evidenciando que ineficiência e grande porte podem coexistir.

Tabela 3.11 – Os cinco prestadores com maior eficiência técnica média (2015–2022).

Pos.	Prestador	UF	Região	Tipo de Gestão	Pop. Atendida (hab.)	Ef. Média	Ef. Mín.	Ef. Máx.
1º	Departamento de Esgoto e Água de Guaíra	SP	Sudeste	Emp. Estatal	40.385	0.9915	0.9909	0.9920
2º	Serviço Autônomo Municipal de Água e Esgoto	PR	Sul	Emp. Estatal	4.674	0.9896	0.9889	0.9902
3º	Prefeitura Municipal de São José da Bela Vista	SP	Sudeste	Adm. Direta	8.541	0.9877	0.9869	0.9885
4º	Serviço Autônomo de Água e Esgoto	SP	Sudeste	Emp. Estatal	81.096	0.9857	0.9848	0.9866
5º	Companhia de Água e Esgoto do Ceará	CE	Nordeste	Emp. Estatal	4.315.775	0.9834	0.9824	0.9844

Nota: Eficiência técnica estimada pelo Modelo Translog Tempo-Variante (SFA). Ef. Média = média da eficiência técnica anual no período 2015–2022. População atendida com abastecimento de água (AG001), média do período. Fonte: elaboração própria.

Tabela 3.12 – Os cinco prestadores com menor eficiência técnica média (2015–2022).

Pos.	Prestador	UF	Região	Tipo de Gestão	Pop. Atendida (hab.)	Ef. Média	Ef. Mín.	Ef. Máx.
1º	Prefeitura Municipal de Doutor Ricardo	RS	Sul	Adm. Direta	840	0.4340	0.4111	0.4567
2º	Serviço de Água, Esgoto e Limpeza Pública	MG	Sudeste	Emp. Estatal	5.344	0.4393	0.4165	0.4620
3º	Prefeitura Municipal de Governador Lindenberg	ES	Sudeste	Adm. Direta	4.767	0.4478	0.4251	0.4703
4º	Companhia de Saneamento Ambiental do Maranhão	MA	Nordeste	Emp. Estatal	2.303.555	0.4824	0.4602	0.5045
5º	Serviço Autônomo Municipal de Água e Esgoto	PR	Sul	Emp. Estatal	7.545	0.4845	0.4623	0.5065

Nota: Eficiência técnica estimada pelo Modelo Translog Tempo-Variante (SFA). Ef. Média = média da eficiência técnica anual no período 2015–2022. População atendida com abastecimento de água (AG001), média do período. Fonte: elaboração própria.

3.4.4 Testes de robustez

A exigência de painel balanceado com oito anos de dados completos adiciona potencial viés de sobrevivência: permanecem na amostra apenas os prestadores mais estáveis institucionalmente e com maior capacidade de reporte ao SNIS, o que pode superestimar ou subestimar a eficiência média do setor. Para avaliar a sensibilidade dos resultados a essa restrição, reestimou-se o modelo com janelas temporais menores, sendo a primeira de 2017–2022 (seis anos, 308 prestadores) e a segunda de 2018–2022 (cinco anos, 339 prestadores), que incorporam prestadores ausentes da amostra principal e ampliam a cobertura populacional de 61,5% para 66,1% e 67,5%, respectivamente.

Os resultados da Tabela 3.13 revelam estabilidade notável dos coeficientes da fronteira. Os parâmetros dos preços dos insumos, das elasticidades de produto e das interações principais mantêm sinais, magnitudes e significâncias estatísticas muito próximas às do modelo base em todas as especificações. Os parâmetros de variância também permanecem estáveis (γ varia de 0,935 a 0,944; λ de 3,78 a 4,12), confirmando que a dominância da ineficiência sobre o ruído aleatório não é próprio da janela temporal escolhida. A ausência de significância estatística da propriedade privada se mantém nas três especificações, reforçando a robustez desse achado central.

Tabela 3.13 – Testes de Robustez: Modelo Translog Tempo-Variante com Diferentes Amostras

Variável	2015-2022		2017-2022		2018-2022	
	Coef.	EP	Coef.	EP	Coef.	EP
Intercepto	7.9483***	0.1137	8.5089***	0.1064	8.5341***	0.1025
$\ln(P_L/P_E)$	0.4307***	0.0077	0.4188***	0.0081	0.4103***	0.0094
$\ln(P_O/P_E)$	0.4383***	0.0062	0.4508***	0.0067	0.4515***	0.0073
$\ln(Y_{\text{água}})$	0.1139***	0.0181	0.1473***	0.0163	0.1630***	0.0162
$\ln(Y_{\text{esgoto}})$	-0.0028	0.0122	0.0353**	0.0123	0.0741***	0.0128
$\ln(K)$	0.8698***	0.0185	0.7862***	0.0170	0.7489***	0.0183
$\frac{1}{2}[\ln(P_L/P_E)]^2$	0.1141***	0.0108	0.0873***	0.0105	0.0652***	0.0120
$\frac{1}{2}[\ln(P_O/P_E)]^2$	0.1523***	0.0070	0.1454***	0.0068	0.1331***	0.0072
$\ln(P_L/P_E) \times \ln(P_O/P_E)$	-0.1203***	0.0072	-0.1050***	0.0070	-0.0900***	0.0076
$\frac{1}{2}[\ln(Y_{\text{água}})]^2$	0.0657**	0.0223	-0.0115	0.0158	0.0668***	0.0096
$\frac{1}{2}[\ln(Y_{\text{esgoto}})]^2$	-0.0102	0.0099	0.0043	0.0098	0.0300**	0.0093
$\ln(Y_{\text{água}}) \times \ln(Y_{\text{esgoto}})$	-0.0379***	0.0102	-0.0409***	0.0090	-0.0439***	0.0083
$\frac{1}{2}[\ln(K)]^2$	0.0216	0.0187	-0.0468*	0.0188	-0.0835***	0.0229
$\ln(P_L/P_E) \times \ln(Y_{\text{água}})$	0.0051	0.0111	0.0095	0.0113	-0.0123	0.0118
$\ln(P_L/P_E) \times \ln(Y_{\text{esgoto}})$	0.0367***	0.0080	0.0186*	0.0081	0.0120	0.0090
$\ln(P_O/P_E) \times \ln(Y_{\text{água}})$	-0.0425***	0.0097	-0.0338***	0.0085	-0.0091	0.0079
$\ln(P_O/P_E) \times \ln(Y_{\text{esgoto}})$	-0.0044	0.0060	-0.0079	0.0054	-0.0202***	0.0060
$\ln(P_L/P_E) \times \ln(K)$	-0.0619***	0.0115	-0.0559***	0.0119	-0.0182	0.0133
$\ln(P_O/P_E) \times \ln(K)$	0.0558***	0.0102	0.0604***	0.0095	0.0279**	0.0104
$\ln(Y_{\text{água}}) \times \ln(K)$	-0.0432*	0.0187	0.0299	0.0156	0.0133	0.0143
$\ln(Y_{\text{esgoto}}) \times \ln(K)$	0.0374**	0.0122	0.0363**	0.0118	0.0203	0.0123
Controles Estruturais						
$\ln(\text{Densidade})$	0.3756***	0.0193	0.2476***	0.0169	0.2073***	0.0183
$\ln(\text{Micromedicação})$	0.0677	0.0387	0.0830**	0.0309	0.0394	0.0364
Índice de Perdas	-0.0011***	0.0003	-0.0014***	0.0003	-0.0015***	0.0004
Cobertura de Esgoto	-0.0004	0.0004	-0.0016***	0.0004	-0.0015***	0.0004
Dummies Institucionais						
Tratamento de Esgoto	-0.0278*	0.0140	-0.0430**	0.0148	-0.0725***	0.0168
Gestão Independente	0.1021***	0.0248	0.0970***	0.0222	0.2330***	0.0364
Propriedade Privada	-0.0268	0.0237	-0.0205	0.0228	-0.0203	0.0247
Prestador Local	0.0274	0.0645	0.0186	0.0668	-0.0159	0.0501
Múltiplos Prestadores	0.3160***	0.0552	0.3109***	0.0587	0.2293***	0.0654
Dummies Regionais (ref.: Centro-Oeste)						
Norte	0.3219***	0.0607	0.2202***	0.0592	0.1618**	0.0576
Nordeste	0.1034*	0.0422	0.1096**	0.0397	-0.0499	0.0379
Sudeste	0.1481***	0.0314	0.1638***	0.0283	0.1082***	0.0311
Sul	0.1062**	0.0390	0.1038**	0.0345	0.0814*	0.0339
Parâmetros de Variância						
σ^2	0.1126		0.1217		0.1531	
γ	0.9346		0.9404		0.9443	
λ	3.7807		3.9738		4.1179	
Características da Amostra						
Observações	2032		1848		1695	
Prestadores	254		308		339	
Anos	8		6		5	
Cobertura Pop. (%)	61.5		66.1		67.5	
Log-Verossimilhança	1648.49		1414.04		1086.51	
Estatísticas de Eficiência						
Eficiência Média	0.7542		0.7576		0.7321	
Eficiência Mediana	0.7488		0.7601		0.7268	

Notas: Erros padrão entre parênteses. * p < 0,05; ** p < 0,01; *** p < 0,001. $\sigma^2 = \sigma_u^2 + \sigma_v^2$; $\gamma = \sigma_u^2/\sigma^2$; $\lambda = \sigma_u/\sigma_v$. Variável dependente: $\ln(\text{Custo Variável} / P_{\text{Energia}})$. CV = FN015. Todos os modelos utilizam painéis balanceados com série temporal completa para cada prestador.

As eficiências médias das amostras alternativas (0,758 para 2017–2022 e 0,732 para 2018–2022) são comparáveis à do modelo base (0,754), sem diferenças que alterem as conclusões qualitativas. As Figuras 3.3 e 3.4 mostram que o padrão de crescimento temporal da eficiência e a hierarquia entre tipos de gestão se mantêm robustos nas amostras alternativas, ainda que com algumas diferenças de nível. Na amostra 2018–2022, a administração pública direta apresenta queda abrupta para cerca de 64–66%, possivelmente refletindo a entrada de prestadores municipais mais fragilizados que só completam cinco anos contínuos de dados completos, e que estariam capturando a cauda inferior da distribuição de eficiência do setor. Esse resultado indica que as estimativas do modelo base podem, de fato, subestimar levemente a ineficiência do setor como um todo.

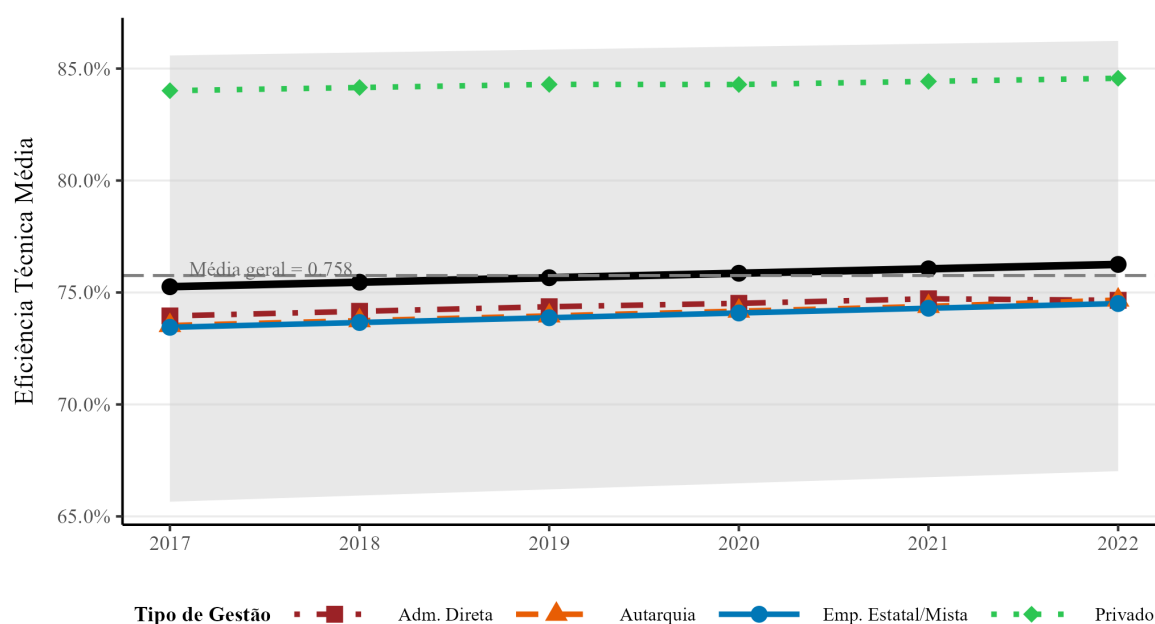


Figura 3.3 – Evolução da eficiência técnica média por tipo de gestão — Amostra de robustez 2017–2022 (308 prestadores)

Nota: Modelo Translog Tempo-Variante. Eficiência média geral = 0,759. Fonte: elaboração própria.

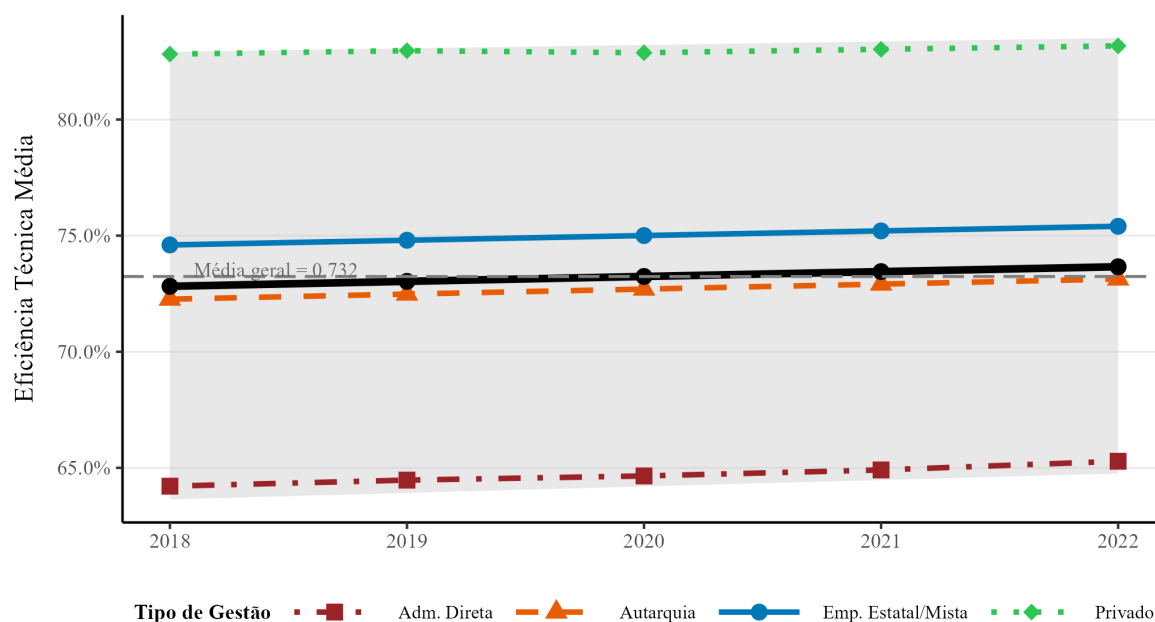


Figura 3.4 – Evolução da eficiência técnica média por tipo de gestão — Amostra de robustez 2018–2022 (339 prestadores)

Nota: Modelo Translog Tempo-Variante. Eficiência média geral = 0,732. Fonte: elaboração própria.

3.5 Discussão

3.5.1 Nível de eficiência: posicionamento na literatura brasileira

A eficiência técnica média estimada de 0,754 para o período 2015–2022 situa-se entre os valores encontrados pelos principais estudos anteriores para o Brasil. Ferro *et al.* (2014) estimaram, com SFA translog para 127 operadores no período 2003–2010, uma eficiência geral de 67,5%, substancialmente inferior à aqui obtida. Essa diferença é esperada e reflete dois fatores distintos: primeiro, o período de análise uma vez que os anos de 2015–2022 representam uma etapa de maior maturidade institucional do setor, posterior à Lei de Diretrizes Nacionais do Saneamento Básico (Lei nº 11.445/2007) e à expansão do SNIS como instrumento de monitoramento; segundo, a restrição de painel balanceado adotada neste estudo seleciona prestadores com maior estabilidade operacional e capacidade de reporte, o que pode introduzir viés de sobrevivência positivo, como discutido na seção de robustez.

O estudo de Danelon (2020) encontrou eficiência média de 84,7% para o período 1995–2016, valor superior ao aqui estimado. Contudo, esse resultado é metodologicamente explicado: o autor utilizou o estimador de Kumbhakar, Wang e Horncastle (2015), que desagrega a ineficiência em componentes persistente e transitória e, ao controlar heterogeneidades latentes, tende a produzir escores de eficiência mais elevados do que modelos que não fazem essa decomposição (Danelon, 2020). A comparação entre os dois resultados não implica contradição, mas diferença de abordagem: o modelo aqui adotado, de Battese e Coelli (1992), é mais parcimonioso e mais

frequentemente utilizado em análises regulatórias de benchmarking, conforme recomendado por Ferro *et al.* (2014).

Já de Souza Pereira *et al.* (2022) analisaram as 95 maiores cidades brasileiras entre 2013 e 2018 com DEA e SFA. Embora não reportem uma eficiência média única comparável, identificam que a distribuição de escores é heterogênea e que o setor de esgoto apresenta tendência de melhoria ao longo do tempo, resultado consistente com a trajetória crescente documentada neste estudo (de 0,741 em 2015 para 0,767 em 2022). A tendência de aprendizado setorial verificada neste artigo é também compatível com os 4,9% de redução anual de ineficiência documentados por Ferro *et al.* (2014) no período anterior, muito embora o ritmo aqui seja mais modesto, o que pode indicar que os ganhos mais fáceis já foram capturados e que a trajetória de convergência para a fronteira está se tornando mais difícil à medida que os prestadores se aproximam de seus limites operacionais.

Vale destacar que nenhum dos estudos brasileiros revisados documentou sistematicamente a trajetória temporal da eficiência com modelo de ineficiência tempo-variante para o período posterior a 2010. Ferro *et al.* (2014) estimam a taxa de redução da ineficiência como parâmetro do modelo, mas não reportam os escores anuais individuais. Danelon (2020) e Silva e Ledo (2023a) trabalham com especificações que capturam heterogeneidade latente, mas sem a dimensão de aprendizado sistemático ao longo do tempo que o modelo de Battese e Coelli (1992) permite identificar. A Tabela 3.10 e a Figura 3.2 constituem, portanto, o primeiro registro sistemático da trajetória anual de eficiência dos prestadores brasileiros no período 2015–2022, oferecendo uma linha de base temporal que vai além do escore médio reportado pela literatura anterior.

3.5.2 Propriedade e gestão: resultado nulo e suas interpretações

O achado central deste artigo, a ausência de efeito estatisticamente significativo da propriedade privada sobre os custos ($\hat{\beta}_{Priv} = -0,027$; $p = 0,26$), é plenamente convergente com a literatura brasileira acumulada. Da Silva E Souza, De Faria e Moreira (2007) estimaram, com SFA translog para 164 empresas em 2002, que o escore médio de ineficiência das firmas públicas (1,452) era apenas marginalmente superior ao das privadas (1,320), diferença não robusta estatisticamente. Motta e Moreira (2006) chegaram à mesma conclusão com DEA para 104 empresas entre 1998 e 2002, verificando que a natureza jurídica não afeta os ganhos de produtividade. Silva e Ledo (2023a) confirmaram, para o período mais recente (2015–2019), que a propriedade formal, seja ela pública ou privada, não determina a eficiência, sendo o mecanismo relevante a estrutura de governança corporativa: empresas mistas de capital aberto (SABESP, COPASA e SANEPAR) superam tanto as puramente públicas quanto as puramente privadas.

Aparentemente divergente é o resultado de Ferro *et al.* (2014), que encontrou significância estatística da propriedade privada sobre o nível de custos. Contudo, o próprio estudo ressalva que é a independência administrativa, e não a propriedade em si, o fator mais relevante quando se controlam outras características estruturais. Este artigo corrobora esse ponto: a dummy de gestão

independente é positiva e significativa ($\hat{\beta}_{Indep} = 0,102$; $p < 0,001$), indicando que autarquias e empresas estatais têm estruturas de custo mais elevadas que a administração pública direta, não porque sejam menos eficientes operacionalmente, mas porque incorrem em maiores gastos de estrutura administrativa e de pessoal, compatíveis com sua maior autonomia gerencial e orçamentária. Danelon (2020) documenta exatamente esse padrão: a independência de gestão está correlacionada com maiores níveis de eficiência técnica, mas as empresas com administração pública direta apresentam os menores custos estruturais absolutos porque sua escala e capacidade administrativa são menores.

O gradiente observado na Figura 3.2, em que prestadores privados apresentam eficiência consistentemente superior à média (cerca de 83–85%), enquanto empresas estatais situam-se abaixo (70–74%), poderia sugerir, numa análise descritiva, uma vantagem associada à propriedade privada. Contudo, os resultados das três especificações estimadas na Tabela 3.7 revelam que essa vantagem não se sustenta quando controlado adequadamente pelas características estruturais do serviço.

Na especificação (2), que exclui os controles estruturais (densidade, micromedição, perdas e cobertura de esgoto) mas mantém as dummies regionais, o coeficiente da propriedade privada é positivo ($\hat{\beta}_{Priv} = 0,0221$), embora não significativo ($p = 0,41$). Isso indica que, na ausência dos controles, os prestadores privados apresentam custos ligeiramente maiores que os públicos, um resultado que contraria a hipótese de superioridade privada. Ao se incluir os controles estruturais na especificação (1), o coeficiente torna-se negativo ($\hat{\beta}_{Priv} = -0,027$), mas permanece não significativo ($p = 0,26$). Na especificação mais parcimoniosa (3), sem controles e sem regionais, o coeficiente volta a ser negativo, mas igualmente não significativo ($\hat{\beta}_{Priv} = -0,0197$; $p = 0,32$).

Esse padrão de instabilidade do coeficiente em sinal e magnitude, mas consistentemente não significativo em todas as especificações, sugere que o efeito da propriedade privada sobre os custos é, na melhor das hipóteses, frágil e sensível à especificação. A interpretação mais plausível, consistente com a literatura brasileira (Da Silva E Souza; De Faria; Moreira, 2007; Motta; Moreira, 2006; Silva; Ledo, 2023a), é que a aparente vantagem das privadas observada na Figura 3.2 decorre não da propriedade em si, mas das condições estruturais mais favoráveis em que operam: mercados mais densos, menores índices de perdas e maior cobertura de micromedição. Uma vez que esses fatores são controlados, qualquer diferença residual entre tipos de propriedade desaparece estatisticamente.

Esse resultado é particularmente relevante para o debate regulatório: a privatização, por si só, não gera ganhos de eficiência mensuráveis. O que importa não é se o prestador é público ou privado, mas se ele opera em um ambiente regulatório que oferece incentivos para a redução de perdas, a expansão da micromedição e a densificação da rede. Esse argumento é corroborado pela especificação (2): ao remover os controles estruturais, o coeficiente da propriedade privada torna-se positivo, sugerindo que, em mercados menos favoráveis, os prestadores privados podem ter desempenho pior que os públicos.

3.5.3 O papel do ambiente regulatório e as implicações do NMLSB

A ausência de ruptura na trajetória de eficiência em torno de 2020, verificada tanto na Tabela 3.10 quanto na Figura 3.2, é um resultado esperado dado o desenho da pesquisa, e não representa evidência de ineficácia do NMLSB. O período de análise termina em 2022, apenas dois anos após a promulgação da Lei nº 14.026/2020, e os efeitos regulatórios sobre a eficiência tipicamente se materializam com defasagem considerável, conforme documentado por Saal, Parker e Weyman-Jone (2007) para o caso inglês: os ganhos pós-privatização com regulação da OFWAT levaram vários anos para se manifestar nas estimativas de fronteira. No caso brasileiro, Freire (2025) já identifica, para o período 2013–2023, uma redução de ineficiência em prestadoras criadas após a promulgação da lei, o que segundo os autores é um resultado que aponta para efeitos iniciais positivos, embora com aumento de ineficiência em 2023, atribuído à instabilidade normativa gerada pelos Decretos 11.466 e 11.467 daquele ano.

A conclusão mais relevante desta pesquisa para o debate regulatório não é sobre a eficácia do NMLSB, mas sobre as condições que lhe precederam. Durante o período 2015–2022, o setor operou em um ambiente de regulação fragmentada, com agências infranacionais de capacidade técnica e abrangência heterogêneas e contratos predominantemente de taxa de retorno, que oferecem poucos incentivos à eficiência produtiva (Sampaio; Sampaio, 2020). Silva e Ledo (2023a) documentam que reguladores com abrangência estadual estão associados a menores níveis de ineficiência, e que convênios de cooperação técnica entre agências reguladoras e outras instituições contribuem para o desempenho dos prestadores. O ambiente que prevaleceu nesse período é precisamente o que o NMLSB buscou transformar, ao centralizar a supervisão regulatória na ANA, exigir contratos com metas de desempenho mensuráveis e criar condições para a regulação por comparação (Shleifer, 1985; Tupper; Resende, 2004).

Nesse sentido, os resultados aqui reportados devem ser interpretados como a linha de base a partir da qual os efeitos do NMLSB serão mensuráveis. A eficiência média de 75,4% e a margem de ineficiência residual de 24,6% representam o ponto de partida da reforma: um setor em trajetória de melhoria lenta e espontânea, sem a pressão de contratos baseados em desempenho. Massarutto *et al.* (2023) mostram, para a Itália, que a privatização sem regulação efetiva não gera ganhos de produtividade. Saal, Parker e Weyman-Jone (2007) e Molinos-Senante e Maziotis (2019) demonstram, para a Inglaterra, que os ganhos vieram da regulação independente, não da propriedade em si. O Brasil, ao buscar com o NMLSB as condições institucionais que a literatura identifica como necessárias (regulação independente, contratos com metas e benchmarking sistemático), estabeleceu o arcabouço para que os ganhos documentados no caso inglês possam se repetir no contexto brasileiro. A avaliação empírica dessa hipótese, contudo, requer dados confiáveis do período pós-2022, que só estarão disponíveis nos próximos anos.

Finalmente, o resultado de que o custo da infraestrutura instalada, medido pela elasticidade da extensão da rede ($\hat{\beta} = 0,870$), é o principal determinante dos custos dos prestadores traz uma implicação direta para a política de universalização: a expansão da rede necessária para

alcançar as metas do NMLSB até 2033 impõe pressão significativa sobre os custos operacionais dos prestadores, especialmente nos municípios menores e nas regiões Norte e Nordeste, que partem de menor cobertura. Esse resultado é consistente com Sabbioni (2008) e Danelon (2020), que identificaram nas economias de escala a principal fonte potencial de redução de custos no setor brasileiro. A prestação regionalizada incentivada pela nova legislação, que permite a um único operador atender blocos de municípios, é, portanto, não apenas uma alternativa de governança, mas uma condição de viabilidade econômica para a universalização em contextos de baixa densidade e escala insuficiente, resultado análogo ao documentado por Torres e Paul (2006) para o setor americano.

3.6 Conclusões

Este artigo estimou a eficiência técnica de custo de 256 prestadores brasileiros de serviços de água e esgoto no período 2015–2022, utilizando uma função de custo variável translogarítmica estimada por Análise de Fronteira Estocástica com ineficiência tempo-variante (Battese; Coelli, 1992). O objetivo foi produzir uma linha de base empírica pré-reforma do setor, o ponto de referência do qual os efeitos do Novo Marco Legal do Saneamento (Lei nº 14.026/2020) poderão ser mensurados à medida que mais dados do período pós-reforma se tornem disponíveis.

O primeiro resultado central é que a eficiência técnica média do setor no período foi de 75,4%, com mediana de 74,9% e dispersão expressiva (mínimo de 41,1%; máximo de 99,2%). A trajetória foi de crescimento contínuo, de 74,1% em 2015 a 76,7% em 2022, um ganho acumulado de 2,6 pontos percentuais em oito anos, consistente com a trajetória de aprendizado setorial documentada por Ferro *et al.* (2014) para o período anterior. A margem de ineficiência residual de aproximadamente 24,6% indica que, em média, os prestadores poderiam reduzir seus custos operacionais em cerca de um quarto sem alterar o nível de produção, caso operassem sobre a fronteira de melhor prática.

Em termos monetários, a margem de ineficiência de 24,6% aplicada ao custo variável total da amostra (R\$ 233,1 bilhões ao longo do painel 2015–2022) representa uma economia potencial de aproximadamente R\$ 57,3 bilhões no período, ou cerca de R\$ 7,2 bilhões por ano. Esse potencial de ganho representa o principal argumento econômico para políticas de benchmarking regulatório.

O segundo resultado central, e o mais relevante para o debate sobre privatização, é a ausência de efeito estatisticamente significativo da propriedade privada sobre os custos ($\hat{\beta} = -0,027$; $p = 0,26$), resultado robusto às especificações alternativas com janelas de cinco e seis anos. Esse achado se soma a uma série de evidências convergentes para o Brasil (Da Silva E Souza; De Faria; Moreira, 2007; Motta; Moreira, 2006; Silva; Ledo, 2023a) e consolida a conclusão de que, no período anterior à plena implementação do NMLSB, a forma de propriedade por si só não determina a eficiência dos prestadores. O que determina é o ambiente institucional no

qual eles operam: os resultados mostram que a extensão da rede ($\hat{\beta} = 0,870$), as condições de densidade e micromedição, e a estrutura de gestão (independente versus administração pública direta) têm impacto muito superior ao da natureza jurídica sobre os custos. Esse resultado tem implicação direta para o desenho da política regulatória: instrumentos que promovam a regulação por comparação (Shleifer, 1985), como os previstos no NMLSB, são mais promissores para elevar a eficiência do que mudanças de propriedade desacompanhadas de reformas contratuais e regulatórias.

O terceiro resultado a ser destacado é o de que a extensão da rede de água e esgoto é a variável com maior elasticidade de custo ($\hat{\beta} = 0,870$), confirmando que a infraestrutura instalada é o principal determinante do custo operacional dos prestadores brasileiros. Essa elasticidade próxima de 1 implica que a expansão da rede necessária para atingir as metas de universalização do NMLSB (99% de cobertura de água e 90% de esgoto até 2033) impõe pressão proporcional sobre os custos variáveis dos prestadores. Em municípios pequenos e nas regiões Norte e Nordeste, onde a escala atual é insuficiente para gerar economias de densidade, essa pressão pode comprometer a viabilidade financeira da universalização sem os ganhos de escala que a prestação regionalizada pode proporcionar (Sabbioni, 2008; Danelon, 2020). Assim como para o caso americano (Torres; Paul, 2006), o incentivo à consolidação regional previsto na nova legislação brasileira não é apenas uma escolha de governança, mas uma condição econômica para a sustentabilidade do setor.

Implicações para a política regulatória

Os resultados têm três implicações diretas para os reguladores e formuladores de política. Primeira: a heterogeneidade de eficiência documentada (escores variam de 42% a 99% entre prestadores) cria condições para uma regulação por comparação efetiva, em que os prestadores menos eficientes possam ser identificados, comparados com pares estruturalmente semelhantes e pressionados a convergir para a fronteira de melhor prática. A função de custo translog aqui estimada oferece a estrutura metodológica para esse benchmarking, controlando por produto, preços de insumos e características ambientais que estão fora do controle do gestor. Segunda: os resultados sugerem que o foco regulatório no período pós-NMLSB deve incidir não sobre a forma de propriedade, mas sobre os determinantes estruturais identificados como relevantes: redução de perdas, expansão da micromedição, densificação da rede e fortalecimento da capacidade de gestão dos prestadores locais. Terceira: a tendência de crescimento da eficiência sem ruptura em 2020 reforça que os efeitos do NMLSB ainda não eram perceptíveis ao final do período analisado, o que é esperado dado o tempo necessário para a internalização de contratos com metas de desempenho. Os reguladores devem acompanhar a trajetória de eficiência nos anos subsequentes, usando a linha de base aqui estabelecida como referência.

Limitações e agenda de pesquisa

Este estudo apresenta três limitações que devem orientar pesquisas futuras. Primeira: a exigência de painel balanceado introduz potencial viés de sobrevivência, excluindo os prestadores com menor capacidade de reporte continuado ao SNIS, tipicamente os menores e mais fragilizados, que são justamente os mais relevantes para a agenda de universalização. Métodos de imputação de dados ou estimadores para painéis não balanceados podem reduzir esse viés em pesquisas futuras. Segunda: o modelo estima apenas ineficiência técnica de custo, deixando de fora a ineficiência alocativa, ou seja, a possibilidade de que os prestadores estejam usando a combinação errada de insumos dados os preços relativos. A estimação conjunta da ineficiência técnica e alocativa, embora metodologicamente mais exigente, produziria uma medida mais completa do potencial de ganho de eficiência. Terceira: o período de análise termina em 2022, o que impede qualquer inferência sobre os efeitos do NMLSB, a avaliação empírica desses efeitos requer dados de mais longo prazo, quando as mudanças contratuais e regulatórias terão tido tempo suficiente para se materializar nas decisões operacionais dos prestadores.

A agenda de pesquisa mais imediata é precisamente a avaliação dos efeitos do NMLSB sobre a trajetória de eficiência, utilizando o ponto de base aqui estabelecido como contrafactual. Nesse sentido, este estudo contribui para a literatura ao documentar, com metodologia rigorosa e amostra representativa, o estado do setor imediatamente antes da maior reforma regulatória da sua história recente.

4 Participação privada no mercado de água reduz tarifas? Evidências de uma reforma legal brasileira

4.1 Introdução

A posição majoritária da literatura científica é a de que empresas privadas são mais eficientes e rentáveis do que suas concorrentes de propriedade estatal (Megginson; Netter, 2001; Galiani; Gertler; Schargrodsky, 2005; Clarke; Kosec; Wallsten, 2009; Cetrulo; Marques; Malheiros, 2019), de modo que um dos mecanismos investigados para a melhora dos indicadores de acesso ao saneamento é o aumento de eficiência por meio da privatização dos serviços. No Brasil, contudo, os resultados são inconclusivos (Cetrulo; Marques; Malheiros, 2019), e o debate ganhou nova dimensão com a aprovação do NMLSB, Lei n.º 14.026, de 15 de julho de 2020, que atribuiu à ANA a competência regulatória do setor e estabeleceu metas de universalização, redução de perdas e modicidade tarifária (BRASIL, 2020).

O efeito institucional do NMLSB sobre a estrutura do mercado de saneamento foi imediato e expressivo. A Figura 4.1 apresenta o número anual de concessões municipais de saneamento entre 1994 e 2024. No período anterior à reforma, o volume de concessões oscilou próximo à média histórica de 23 por ano, com dois picos isolados em 1999 e 2015. A partir de 2021, primeiro ano completo de vigência do novo marco regulatório, o número de concessões saltou para mais de 80, atingindo o pico de 320 em 2023, mais de dez vezes a média histórica. Parte desse número se deve à regularização de concessões existentes, não somente criação de novas, pois muitos contratos já existiam de forma irregular. Esse padrão é consistente com a hipótese de que o NMLSB funcionou como um choque exógeno de política, criando incentivos e condições institucionais que aceleraram a transição de municípios para o modelo de provisão privada.

Uma consequência operacional desse processo já é observável nos dados agregados. A Figura 4.2 apresenta a evolução do índice incidência da despesa de pessoal próprio nas despesas de exploração de água (IFA2008) para os grupos de municípios tratados (concessão a partir de 2020) e de controle (nunca tratados), entre 2017 e 2024. No período pré-tratamento, as trajetórias dos dois grupos são paralelas e próximas, com o grupo tratado partindo de um nível ligeiramente superior. A partir de 2020, as trajetórias divergem de forma pronunciada: enquanto o grupo de controle mantém queda gradual e moderada, os municípios concedidos experimentam redução abrupta e contínua, chegando a valores próximos de 30% em 2024, ante cerca de 55% em 2019. Essa divergência antecipa o resultado central do estudo e sugere que as concessões promoveram uma reestruturação expressiva da composição de custos operacionais, com substituição de pessoal

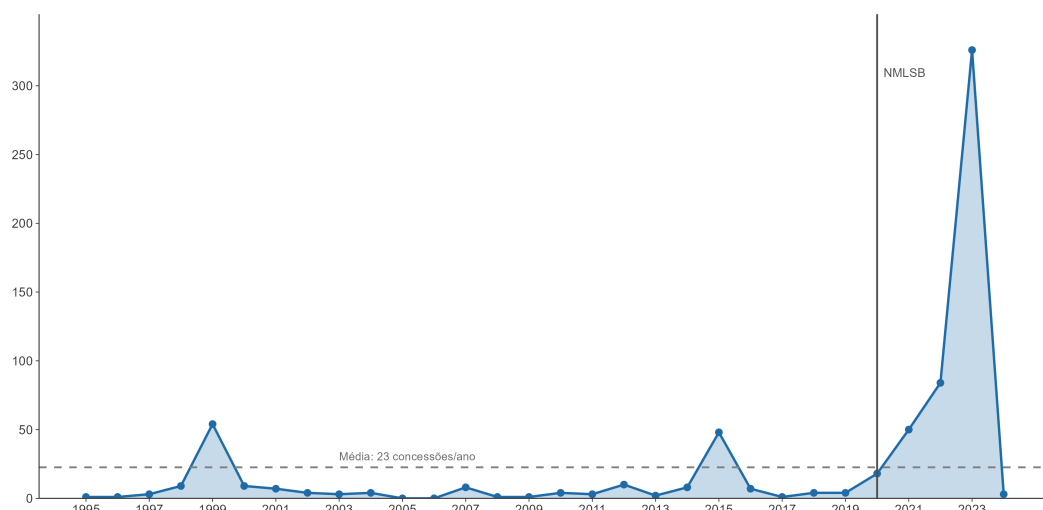


Figura 4.1 – Número anual de concessões municipais de serviços de saneamento. Brasil, 1994–2024.

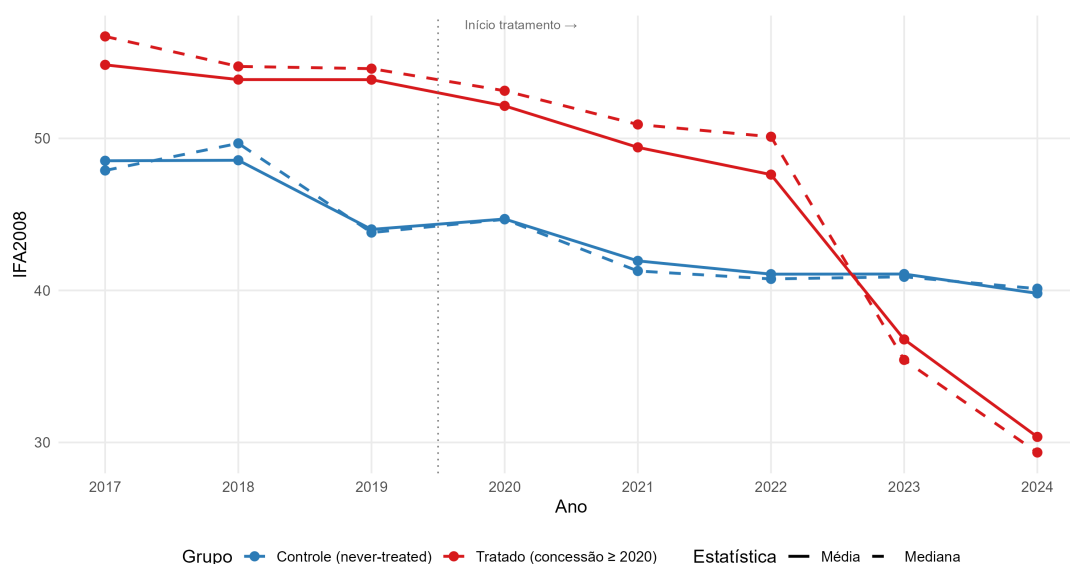


Figura 4.2 – Tendência do índice IFA2008 (incidência da despesa de pessoal próprio nas despesas de exploração de água) para municípios tratados e controles. Brasil, 2017–2024.

próprio por terceirização ou automação.

A questão tarifária é central nesse debate. Os incentivos para as companhias estatais estabelecerem uma cobrança compatível com o nível de investimentos que o setor requer são baixos, uma vez que tarifas mais elevadas geram perda de acessibilidade e insatisfação popular (Clarke; Kosec; Wallsten, 2009). Tal situação configura um trade-off entre acesso e sustentabilidade financeira: na ausência de uma tarifa adequada, torna-se difícil expandir a cobertura, prejudicando a universalização.

A questão da acessibilidade tarifária é particularmente aguda no Brasil. Martínez-espiñeira e Pérez-urdales (2025) estimam que os domicílios do quintil mais pobre gastam, em média, 6,7% da renda apenas com água encanada, o que é mais do que o dobro do limiar internacional de 3%

frequentemente utilizado como referência de acessibilidade. Quando se incorporam os custos de enfrentamento decorrentes do uso de fontes alternativas, como água mineral e caminhão-pipa, esse índice sobe para 10,2%, evidenciando que as métricas convencionais subestimam o verdadeiro ônus financeiro das famílias mais vulneráveis. Esse cenário reforça a centralidade da modicidade tarifária no debate sobre o modelo de provisão dos serviços de saneamento. Os primeiros leilões realizados sob o NMLSB, baseados no critério de maior outorga, agravaram essa tensão ao transferir para os usuários o custo de prêmios bilionários, com impactos tarifários de 13,2% a 46,3% nas concessões de Alagoas e Rio de Janeiro (Fagundes; Cunha Marques; Malheiros, 2024).

Por outro lado, as firmas privadas também enfrentam desincentivos ao investimento de longo prazo, típico do saneamento, em razão da desconfiança popular na privatização da água e da ameaça de expropriação de ativos em caso de reversão do modelo (Clarke; Kosec; Wallsten, 2009). A literatura sugere que direitos de propriedade mais bem definidos podem superar essa barreira: uma reforma anterior da legislação brasileira que clarificou os direitos dos municípios sobre suas companhias de saneamento quase dobrou o investimento dessas companhias (Kresch, 2020). O NMLSB pode ser interpretado como um aprofundamento dessa lógica, ao estabelecer regras mais claras para contratos de concessão e criar uma autoridade regulatória nacional com poder de normatização.

Diante desse contexto, o objetivo deste artigo é estimar os efeitos causais das concessões plenas de abastecimento de água celebradas a partir do NMLSB sobre um conjunto amplo de resultados municipais. Especificamente, investigam-se os efeitos sobre tarifas, cobertura de atendimento, eficiência operacional, qualidade do serviço e morbidade hospitalar por doenças relacionadas ao saneamento. Para tanto, adota-se o estimador de diferenças-em-diferenças escalonado proposto por Callaway e Sant’Anna (2021), combinado com pareamento por escore de propensão, em um painel de municípios brasileiros para o período 2017–2024.

Os resultados indicam que as concessões produziram ganhos operacionais mensuráveis no curto prazo, especialmente na redução de despesas com pessoal e na expansão da macromedição, mas não se traduziram em melhora dos indicadores de qualidade do serviço nem em efeitos detectáveis sobre saúde pública. Os efeitos tarifários são heterogêneos entre coortes e dependem criticamente do modelo de leilão adotado, com o critério de maior outorga associado a aumentos e o de menor tarifa associado a reduções. Esses achados posicionam o NMLSB como uma reforma com potencial real, mas cujos resultados dependem de condições institucionais que ainda estão em processo de consolidação.

O artigo está organizado da seguinte forma. A seção 4.2 apresenta a revisão da literatura sobre privatização de serviços de saneamento e os métodos de avaliação empregados na área. A seção 4.3 descreve a estratégia empírica, incluindo a construção dos indicadores e o desenho quase-experimental. A seção 4.4 apresenta os resultados. A seção 4.5 discute os achados à luz da literatura e de suas implicações para a política regulatória. A seção 4.6 conclui.

4.2 Revisão da literatura

Considerando o acesso ao saneamento sob duas formas, a acessibilidade ou modicidade da tarifa e a cobertura do serviço, como variável de interesse na análise das privatizações das companhias de saneamento, a literatura registra resultados em variadas direções (Fabre; Straub, 2023).

Os impactos econômicos e sociais da privatização das companhias de água e esgoto em diversos países do mundo foram verificados por Lobina e Hall (2000). Um desses impactos é o aumento das tarifas percebido logo após a mudança na provisão do serviço público. Os autores elencam acréscimos nas tarifas de 175% na Hungria, 100,7% seguido de 39,8% para a água e 66,6% para o esgoto da República Checa e 400% para consumidores industriais das Filipinas. Além disso, no Reino Unido houve falhas graves de operadoras privadas relacionadas à política de dividendos em detrimento dos investimentos em infraestrutura.

Pesquisas domiciliares de vários anos da Argentina, Bolívia e Brasil antes e depois da introdução da participação privada no setor de saneamento foram analisadas por Clarke, Kosec e Wallsten (2009) especificamente sobre o número de conexões de água e esgoto e as comparou com municípios que não privatizaram seus serviços. Os achados deste trabalho foram o aumento do número de ligações pós privatização, porém também foi verificado similar expansão nas cidades que permaneceram com a provisão pública no saneamento, de modo que a evidência não indica a privatização como responsável pelo aumento na cobertura de água e esgoto. Entretanto, um resultado bastante interessante foi obtido quando os autores analisaram por renda e os resultados se mantiveram nos domicílios mais pobres, sugerindo que a privatização não os prejudicou (Clarke; Kosec; Wallsten, 2009).

A Bolívia apresentou aumento no acesso pós privatização. Houve a transferência para a iniciativa privada do serviço em algumas cidades e o acesso à água tratada foi comparado com cidades semelhantes que permaneceram com a provisão pública (Hailu; Osorio; Tsukada, 2012). Os resultados mostram que no curto prazo (até cinco anos) não há diferença significativa no acesso, mas que no longo prazo (mais de nove anos) é mais provável que um domicílio tenha acesso numa cidade “privatizada” do que em uma que não privatizou o serviço (Hailu; Osorio; Tsukada, 2012). Em se tratando de tarifa, as conclusões foram prejudicadas pela ausência de dados de despesa domiciliar anteriores à privatização, o que impediu uma comparação direta antes e depois da reforma (Hailu; Osorio; Tsukada, 2012).

Na Colômbia, não houve aumento no acesso ao saneamento pós-privatização, nem efeito sobre as tarifas na média, mas para os quintis mais baixos de renda houve elevação de preços (Barrera-Osorio; Olivera; Ospino, 2009). Em relação à qualidade, verificou-se melhora com repercussões na saúde de crianças tanto da zona urbana como rural. A privatização foi acompanhada da eliminação de subsídios cruzados, o que contribuiu para o aumento de preços, a redução do acesso na zona rural e, paradoxalmente, a melhora na frequência do serviço para

pessoas mais pobres nos municípios privatizados (Barrera-Osorio; Olivera; Ospino, 2009).

O caso da Argentina é particularmente interessante em termos de acesso, pois houve redução significativa da mortalidade infantil principalmente em áreas pobres cujo acesso aumentou mais que nas demais áreas (Galiani; Gertler; Schargrodsky, 2005). No entanto, a questão da regulação tarifária permaneceu um problema já que, por exemplo, em Buenos Aires reduziu-se a tarifa em 26% em virtude do leilão mas que em 1 ano um reajuste já foi autorizado em 13%, além de cobrança de instalação com valor inacessível para os mais vulneráveis (Galiani; Gertler; Schargrodsky, 2005). Em outra cidade argentina, Tucumán, as tarifas de água dobraram logo após a concessão, ao mesmo tempo em que a empresa concessionária deixou de cumprir o programa de investimentos previsto (Lobina; Hall, 2000).

Com dados no nível do domicílio na Malásia, verificou-se que a privatização teve impacto indefinido no acesso e na acessibilidade (Lee, 2011). Não houve melhora em termos de acesso tampouco efeito negativo sobre as tarifas, o que os autores atribuem à forte presença governamental tanto na oferta como na regulação tarifária do setor (Lee, 2011).

No Brasil, estudos indicam que a privatização pode ter efeitos benéficos na saúde (Saiani; Azevedo, 2018; Oliveira; Saiani, 2022; Chaves; Costa, 2025). O modo de provisão privado local diminuiu as taxas de morbidade em doenças relacionadas ao saneamento em grupos etários vulneráveis, já o modelo híbrido (privado com direitos estaduais) não mostrou evidência robusta de deterioração da qualidade (Saiani; Azevedo, 2018). Chaves e Costa (2025) atualizam essa análise para o período 1998–2021 com o estimador de DiD escalonado de Callaway–Sant’Anna, obtendo resultados mistos: enquanto alguns municípios apresentaram redução nas taxas de morbidade, outros não exibiram mudança ou até registraram aumento, o que ressalta o caráter dependente do contexto local nos desfechos da privatização. Um caso emblemático é o de Tocantins, onde a transição de um modelo híbrido público-privado para a propriedade privada plena resultou em redução significativa da morbidade por doenças ligadas ao saneamento, especialmente nos grupos etários mais vulneráveis (Chaves; Costa, 2025).

Ainda, foram encontradas evidências da existência de trade-off custo-qualidade (sendo a qualidade medida como proxy pela morbidade hospitalar) nas provisões pública e híbrida, com o trade-off sendo atenuado na provisão híbrida (sociedades de economia mista e empresas públicas) (Oliveira; Saiani, 2022). Já na provisão privada, os resultados sugerem ausência do trade-off, com possível reversão (menor custo e maior qualidade) (Oliveira; Saiani, 2022).

Com dados de 2007 a 2018 do Brasil, constatou-se que as tarifas médias reais definidas por concessões privadas são inicialmente mais baixas do que as de entidades públicas (Dassan *et al.*, 2024). No entanto, as tarifas aumentaram mais nos primeiros quatro anos após os municípios optarem por operadores privados do que nos municípios com operadores públicos, e a cobertura de água, a coleta e o tratamento de esgoto também cresceram mais entre os prestadores privados do que entre os públicos (Dassan *et al.*, 2024).

No que diz respeito aos determinantes da privatização no Brasil, Nunes e Marques (2025)

investigam, com análise de duração em tempo discreto aplicada a painel de 5.397 municípios entre 2012 e 2023, quais fatores no nível da prestadora estão associados à maior probabilidade de privatização dos serviços de água e esgoto. Contrariamente à hipótese de que municípios com maior necessidade de investimento, menor eficiência ou pior qualidade de serviço seriam mais propensos a privatizar, o principal achado é que a condição financeira da prestadora é o fator mais relevante: operadoras mais lucrativas e com tarifas mais altas têm maior probabilidade de ser privatizadas, sugerindo que a atratividade da concessão para o setor privado é um elemento central no processo. Variáveis municipais, como menor saúde fiscal, maior população, governo não de esquerda e proximidade de municípios já privatizados, também se mostraram associadas positivamente à privatização, em linha com a literatura (Nunes; Marques, 2025).

4.2.1 Métodos de avaliação de privatizações

Diversos métodos são empregados na literatura científica com o intuito de estimar os efeitos da privatização de serviços públicos, e no caso do saneamento o mesmo é verificado. Um desses métodos é o estudo de caso, no qual é realizada uma análise detalhada de um ou poucos casos específicos, como a concessão na Região Metropolitana de Maceió (Carvalho, 2022) ou a experiência de países específicos como Reino Unido, França (Dore; Kushner; Zumer, 2004), Chile (Molinos-Senante; Sala-Garrido, 2015), ou México (Megginson; Netter, 2001). Essa metodologia pode incluir pesquisa documental (Carvalho, 2022). Existem ainda as revisões de literatura e surveys empíricos, nos quais são realizadas a compilação e síntese dos resultados de múltiplos estudos para identificar padrões ou lacunas (Cetrulo; Marques; Malheiros, 2019; Lima; Brochado; Marques, 2021).

No entanto, são os métodos econométricos que apresentam maior destaque nos trabalhos que estimam o efeito médio da privatização. Estudos mais gerais sobre eficiência ou fatores que a influenciam utilizam análises de regressão, como Mínimos Quadrados Ordinários ou *Ordinary Least Squares* (OLS), ou Tobit (Lee, 2011). Alguns métodos em econometria merecem ênfase, o primeiro deles é o Diferença em Diferenças ou *Difference-in-Differences* (DiD), que compara as mudanças nos resultados ao longo do tempo entre municípios ou regiões que privatizaram (grupo de tratamento) e aqueles que não o fizeram (grupo de controle). É frequentemente usado em conjunto com técnicas de pareamento, como o PSM, para criar grupos de controle comparáveis em estudos observacionais.

O estudo de Galiani, Gertler e Schargrotsky (2005) sobre a Argentina utilizou métodos não experimentais que mimetizam ensaios controlados randomizados para identificar o efeito médio da privatização na mortalidade infantil. Saiani e Azevedo (2018) e Dassan *et al.* (2024) também usam o DiD com PSM para avaliar os efeitos da privatização no Brasil. Chaves e Costa (2025), por sua vez, empregam o estimador de DiD escalonado proposto por Callaway e Sant'Anna (2021), que lida com a heterogeneidade dos efeitos do tratamento em contextos em que diferentes unidades são tratadas em momentos distintos, permitindo uma análise mais robusta no caso

brasileiro.

Dados em Painel (Fixed Effects/Within Estimator) é uma abordagem econométrica que analisa dados de múltiplos municípios ao longo do tempo, permitindo controlar por características não observadas e fixas no tempo que possam influenciar tanto a provisão quanto os resultados. Os estudos de Oliveira (2008) e Oliveira e Saiani (2022) sobre o Brasil utilizam regressões em painel de municípios com dados anuais, no primeiro de 2001 a 2003 com métodos de efeitos fixos e aleatórios, e no segundo, de 2008 a 2019 empregando o método de efeitos fixos com erros-padrão robustos. Nunes e Marques (2025) adotam análise de duração em tempo discreto para modelar o momento em que cada município transita para a provisão privada, abordagem adequada para capturar fatores variantes no tempo associados à decisão de privatizar.

Por fim, há técnicas usadas especificamente para medir a eficiência dos operadores e que nessa seara servem para comparar a eficiência de empresas públicas e privadas. Entre os estudos para países em desenvolvimento, a Análise Envoltória de Dados (Data Envelopment Analysis ou DEA) é a mais utilizada, mas há também bastante frequência de estudos com base em Análise de Fronteira Estocástica (Stochastic Frontier Analysis ou SFA) (Cetrulo; Marques; Malheiros, 2019).

4.3 Metodologia

4.3.1 Construção dos indicadores

A base de dados utilizada neste estudo compreende informações provenientes de duas versões complementares do sistema federal de informações sobre saneamento. Para o período 2017-2022, utilizou-se o SNIS, cujos microdados estão disponíveis ao nível do prestador de serviço. Para os anos de 2023 e 2024, a versão é o SINISA, que disponibiliza uma Base Municipal Consolidada com indicadores já calculados no nível do município. Ambas as bases foram complementadas com dados fiscais, demográficos e de infraestrutura dos municípios provenientes do IBGE e do Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro (SICONFI), bem como com informações da Associação e Sindicato Nacional das Concessionárias Privadas de Serviços Públicos de Água e Esgoto (ABCON SINDCON) (ABCON SINDCON, 2024).

A estrutura do SNIS impõe uma limitação metodológica relevante: os dados brutos são reportados ao nível do prestador, de modo que municípios com mais de um prestador atuante possuem múltiplas observações por ano. Conforme argumentado por Oliveira e Saiani (2022) e seguido por Chaves e Costa (2025), a regionalização incentivada pelo NMLSB reforça a necessidade de construir indicadores consolidados por município. Para construir indicadores comparáveis em nível municipal, os microdados dos prestadores foram consolidados por município e ano mediante a soma das variáveis quantitativas subjacentes, preservando a lógica das fórmulas oficiais do glossário técnico do SNIS (BRASIL. Ministério das Cidades., 2023). Em seguida,

os indicadores foram recalculados a partir dessas variáveis consolidadas, gerando medidas padronizadas que, para municípios com prestador único, são numericamente equivalentes aos índices originais do SNIS. A qualidade dos indicadores recalculados foi verificada por comparação com os valores originalmente publicados no SNIS para municípios com prestador único, com resultados reportados na Tabela A.2 (Apêndice A). Os indicadores que não atingiram correlação mínima de 0,95 ou RMSE relativo superior a 20% foram excluídos da análise.

A harmonização entre as duas bases é necessária porque o SINISA adota uma nova nomenclatura de indicadores (códigos IAG, de Indicadores de Abastecimento de Água e IFA, de Indicadores Financeiros e Administrativos) que corresponde, com pequenas diferenças de cálculo, aos indicadores tradicionais do SNIS (códigos IN). A compatibilização foi realizada a partir da tabela oficial de equivalências publicada pelo próprio SINISA, que mapeia cada indicador da nova nomenclatura ao respectivo indicador correspondente no SNIS. A Tabela A.1 (Apêndice A) apresenta essa correspondência para os indicadores utilizados no estudo.

Os indicadores resultantes estão organizados em quatro grupos temáticos: (i) atendimento, que mensura a cobertura da população e dos domicílios com rede de abastecimento de água (ex.: IAG0001/IN055); (ii) estruturais e operacionais, que capturam aspectos como hidrometração, perdas na distribuição e consumo per capita; (iii) qualidade, que refletem paralisações, interrupções e tempo de reparo; e (iv) administrativo-financeiro, que compreendem receitas operacionais, despesas de exploração e indicadores de eficiência (ex.: IFA1001/IN005, IFA2008/IN035).

Após o cálculo, foram aplicados filtros de consistência para remoção de outliers: valores percentuais fora do intervalo [0, 100] e valores estritamente negativos para variáveis de volume e custo foram substituídos por missings. O painel final cobre o período 2017-2024, abrangendo todos os municípios brasileiros com informações disponíveis em ambas as fontes.

4.3.2 Estratégia empírica

Este estudo adota uma abordagem quase-experimental, inspirada nas estratégias propostas por Galiani, Gertler e Schargrodsky (2005), Saiani e Azevedo (2018), Dassan *et al.* (2024), e alinhada ao desenho metodológico recente de Chaves e Costa (2025), para estimar os efeitos causais da privatização no setor de saneamento. A identificação do efeito causal baseia-se no estimador de Diferenças em Diferenças para tratamentos escalonados proposto por Callaway e Sant'Anna (2021). Esse estimador é preferível ao modelo de efeitos fixos de dois sentidos (*two-way fixed effects*, TWFE) convencional em contextos com adoção escalonada do tratamento, pois o TWFE pode produzir estimativas enviesadas quando os efeitos do tratamento são heterogêneos entre grupos e ao longo do tempo (Baker *et al.*, 2025).

Grupos de tratamento e controle

O grupo tratado é composto pelos municípios que celebraram contratos de concessão plena de serviços de abastecimento de água com operadores privados a partir de 2020, ano de publicação do Novo Marco Legal do Saneamento Básico (Lei n.º 14.026/2020). Define-se G_i como o primeiro ano em que o município i foi submetido ao tratamento, de modo que os municípios tratados pertencem a diferentes *coortes* ou grupos $g \in \mathcal{G}$ conforme o ano de início da concessão. Municípios que realizaram concessões em qualquer modalidade ou período distintos dos critérios acima, como parceiras público-privadas ou concessões anteriores a 2020, são excluídos da amostra para evitar contaminação do grupo de controle.

O grupo de controle é composto exclusivamente por municípios que nunca concederam os serviços de água a operadores privados em qualquer modalidade ao longo de todo o período amostral (nunca tratado). A adoção do grupo nunca tratado como contrafactual, em vez de municípios ainda não tratados, é conservadora e evita que estimativas de períodos pós-tratamento de outros grupos contaminem a comparação (Callaway; Sant’Anna, 2021).

Parâmetros de interesse

O efeito médio do tratamento sobre os tratados ou *Average Treatment Effect on the Treated* (ATT) é estimado separadamente para cada combinação de coorte g e período t :

$$ATT(g,t) = \mathbb{E} [Y_t(g) - Y_t(0) \mid G_i = g]$$

onde $Y_t(g)$ é o resultado potencial do município se tratado a partir de g , e $Y_t(0)$ é o resultado contrafactual na ausência de tratamento. Os $ATT(g,t)$ são então agregados em três formas complementares: (i) um efeito médio (*overall*), que pondera os efeitos por coorte e período na proporção de municípios tratados em cada combinação; (ii) um efeito por coorte, que captura o impacto médio para cada grupo de municípios com o mesmo ano de adoção; e (iii) um estudo de evento (*event-study*), que examina a dinâmica do efeito em função do tempo relativo ao início da concessão.

Hipóteses de identificação

A identificação se baseia em duas hipóteses principais. Primeiro, a hipótese de tendências paralelas condicional: na ausência de tratamento, os desfechos dos municípios tratados e do grupo nunca tratado teriam seguido trajetórias paralelas, condicional a um conjunto de covariadas observadas medidas antes do tratamento (Callaway; Sant’Anna, 2021). Segundo, a hipótese de não-antecipação: o comportamento dos municípios não é afetado pela expectativa de tratamento futuro, de modo que os efeitos potenciais são nulos em períodos pré-tratamento.

A hipótese de tendências paralelas condicional é sustentada por dois argumentos. Do ponto de vista institucional, a decisão de conceder os serviços de água a operadores privados a partir

de 2020 é atribuída ao ambiente regulatório e incentivos criados pelo NMLSB, corroborada pela evidência do número de concessões (Figura 4.1), sendo portanto um choque de política exógeno ao desempenho histórico dos indicadores de saneamento. Do ponto de vista empírico, o pareamento por escore de propensão restringe a amostra a municípios com características estruturais comparáveis em 2017, reduzindo a probabilidade de que diferenças pré-existent nas trajetórias dos desfechos sejam o mecanismo gerador das diferenças observadas após o tratamento (Baker *et al.*, 2025). A hipótese é testada empiricamente por meio dos coeficientes pré-tratamento do estudo de evento, que nos períodos anteriores a g devem ser estatisticamente indistinguíveis de zero, ou seja, nulos.

Estimação

Os parâmetros $ATT(g, t)$ são estimados pelo método *doubly robust* (DR) (DR), que combina uma regressão do desfecho com uma ponderação por probabilidade inversa. Esse procedimento é consistente se ao menos um dos dois modelos auxiliares (o modelo para a trajetória do desfecho no grupo de controle ou o modelo para o *propensity score*) for corretamente especificado (Callaway; Sant’Anna, 2021; Baker *et al.*, 2025). Os erros-padrão são clusterizados ao nível do município. A estimação foi realizada com o pacote `did` para R (Callaway; Sant’Anna, 2021).

4.3.2.1 Pareamento por escore de propensão (*Propensity Score Matching*)

O método de pareamento por escore de propensão visa construir um grupo de controle adequado, composto por unidades não tratadas que possuam características observáveis semelhantes às unidades tratadas no período anterior à intervenção. Essa abordagem mitiga o viés de seleção e permite que as diferenças nos desfechos entre os grupos, após o tratamento, sejam atribuídas com maior credibilidade à intervenção (Rosenbaum; Rubin, 1983).

O escore de propensão é definido como a probabilidade condicional de receber o tratamento dado um vetor de covariáveis observáveis pré-tratamento, $P(D = 1 | X)$. A utilização do escore como função de balanceamento reduz a dimensionalidade do problema, mantendo o equilíbrio das covariáveis entre os grupos (Rosenbaum; Rubin, 1983). O parâmetro de interesse é o efeito médio do tratamento sobre os tratados (ATT):

$$\tau_{ATT} = E[Y(1) | D = 1] - E[Y(0) | D = 1] \quad (4.1)$$

Sob as hipóteses de seleção em observáveis ($Y(0), Y(1) \perp D | X$) e de suporte comum ($0 < P(D = 1 | X) < 1$), o ATT pode ser identificado por meio do escore de propensão (Rosenbaum; Rubin, 1983):

$$\tau_{ATT}^{PSM} = E_{P(X)|D=1} \{E[Y(1) | D = 1, P(X)] - E[Y(0) | D = 0, P(X)]\} \quad (4.2)$$

A seleção das variáveis para compor o escore seguiu os critérios de Caliendo e Kopeinig (2008): devem ser mensuradas antes da intervenção e não podem ser afetadas pelo tratamento.

Para além desse critério, a escolha do conjunto de covariadas envolve uma tensão conhecida entre parcimônia e completude: especificações mais amplas tendem a produzir escores mais precisos, mas reduzem o tamanho da amostra pareada; especificações mais restritas preservam observações, porém podem omitir determinantes relevantes da decisão de privatização. Para lidar com essa tensão de forma sistemática, foram testadas três estratégias de seleção de covariadas, cujos resultados de balanceamento são comparados na Tabela 4.1.

Tabela 4.1 – Comparação das 3 estratégias de seleção de covariadas para PSM.

Estratégia	Nº Var.	Tratados	Controles	Total	SMD Médio	SMD Máx.
Completo	10	418	418	836	0,0744	0,2296
Rigorouso	3	434	434	868	0,0361	0,0510
Flexível	6	424	424	848	0,0278	0,0639

Fonte: resultados da pesquisa

A estratégia completa emprega dez variáveis (população total, PIB per capita, passivo não circulante, ligações ativas, ligações micromedidas, controle de perdas, volume faturado, volume consumido e duas dummies de abrangência, local e microrregional) e produz uma diferença média padronizada ou *Standardized Mean Difference* (SMD) de 0,074. A estratégia rigorosa, com apenas três variáveis (população, PIB per capita e passivo não circulante), eleva o tamanho da amostra para 868 observações e reduz o SMD médio para 0,036, mas ao custo de não controlar pela estrutura operacional e de abrangência dos prestadores. A estratégia flexível, com seis variáveis (PIB per capita, passivo não circulante, ligações ativas micromedidas, volume faturado por dia e as duas dummies de abrangência), alcança o menor SMD médio (0,028) com uma amostra de 848 observações, combinando o melhor balanceamento geral com cobertura satisfatória das dimensões fiscais, operacionais e institucionais relevantes para a decisão de concessão.

Com base nesses resultados, foi adotada a estratégia flexível como especificação principal. Essa escolha é respaldada por dois critérios complementares. O primeiro, seguindo Caliendo e Kopeinig (2008), é que as variáveis devem ser determinantes observáveis da probabilidade de concessão: o PIB per capita e o passivo não circulante refletem a capacidade fiscal municipal, e o tipo de abrangência do prestador capta a escala operacional que influencia a decisão de privatizar, conforme documentado por Bel (2020) e Saiani e Azevedo (2018). O segundo critério, enfatizado por Callaway e Sant’Anna (2021) (Remark 7), é que as covariadas pré-tratamento devem ser associadas à evolução do desfecho $Y(0)$ na ausência do tratamento: municípios com maior volume faturado, mais ligações micromedidas e maior passivo não circulante em 2017 apresentavam trajetórias estruturalmente distintas de expansão e eficiência operacional, independentemente da decisão de conceder. Ao condicionar nesses fatores, torna-se mais plausível que tratados e controles teriam seguido trajetórias paralelas nos desfechos de interesse na ausência da concessão.

Segundo Callaway e Sant’Anna (2021) (Remark 7), são excluídas covariadas que possam

ter sido afetadas pelo tratamento; todas as variáveis descritas no quadro 4.1 foram mensuradas em 2017, período de referência pré-tratamento, e o pareamento foi realizado pelo algoritmo de vizinho mais próximo sem reposição, implementado no pacote `MatchIt` para R (Ho *et al.*, 2011).

Quadro 4.1 – Variáveis utilizadas no PSM — Estratégia Flexível (2017). Brasil, 2026.

Descrição	Fonte
Abrangência do prestador: local (dummy)	SINISA/SNIS
Abrangência do prestador: microrregional (dummy)	SINISA/SNIS
Produto Interno Bruto per capita (R\$ correntes)	IBGE
Passivo não circulante (R\$ correntes)	SICONFI/STN
Ligações ativas micromedidas (unidades)	PNSB/IBGE
Volume faturado por dia (m ³ /dia)	PNSB/IBGE

Fonte: dados da pesquisa

4.4 Resultados

4.4.1 Pareamento e seleção da amostra

A estimação do PSM resultou em uma amostra de 424 municípios tratados e 424 no grupo de controle, na qual as diferenças entre as variáveis utilizadas não são mais significativas, com base em um teste T de diferenças de médias. A qualidade do balanceamento obtido pela estratégia escolhida é reportada na Tabela 4.2. Antes do pareamento, os grupos de tratamento e controle diferem significativamente em abrangência (local e microrregional) e PIB per capita. Após o pareamento, nenhuma dessas variáveis apresenta diferença estatisticamente significativa ao nível de 5%.

Tabela 4.2 – Diferença de médias entre grupos tratamento e controle para amostras sem pareamento e com pareamento (2017). Brasil, 2026.

Variável	Tratados	Controle	Teste t	Valor-p	Significância	Amostra
Abrangência: local	424	3285	-8,606	0,0000	***	Completa
PIB per capita	424	3285	7,450	0,0000	***	Completa
Passivo não circulante	424	3285	1,212	0,2263	ns	Completa
Ligações ativas micromedidas	424	3285	0,811	0,4178	ns	Completa
Abrangência: microrregional	424	3285	-2,451	0,0143	*	Completa
Volume faturado por dia (m ³)	424	3285	1,071	0,2850	ns	Completa
Abrangência: local	424	424	0,000	1,0000	ns	Pareada
PIB per capita	424	424	-0,810	0,4184	ns	Pareada
Passivo não circulante	424	424	0,990	0,3225	ns	Pareada
Ligações ativas micromedidas	424	424	0,534	0,5935	ns	Pareada
Abrangência: microrregional	424	424	NaN	NaN	ns	Pareada
Volume faturado por dia (m ³)	424	424	1,032	0,3025	ns	Pareada

Nota: Significância: *** p<0,001; ** p<0,01; * p<0,05. Fonte: Resultados da pesquisa.

A amostra resultante de municípios está representada na Figura 4.3, em que os municípios em vermelho são os que concederam a prestação do serviço de saneamento a partir de 2020, ano de entrada em vigor do NMLSB (grupo de tratamento), e em azul são os que permaneceram públicos (grupo de controle) e são comparáveis em termos das variáveis utilizadas no PSM. Observa-se que há uma distribuição aparentemente homogênea entre as regiões, o que depõe a favor da qualidade do pareamento.

As estatísticas descritivas da amostra utilizada nas estimações são apresentadas na Tabela 4.3, separadas por grupo de tratamento. Os indicadores financeiros com dimensão monetária (IFA1001, IFA2002, IFA2004, IFA2005, IFA2006 e IFA2007) foram deflacionados pelo IPCA (Meireles, 2018) e estão expressos em reais de dezembro de 2024, de modo que as comparações entre anos e grupos refletem diferenças reais de preços e custos, expurgado o efeito inflacionário. Alguns padrões merecem destaque antes da análise causal. Em média, os municípios tratados apresentam maior receita operacional direta (IFA1001: R\$ 11,65/m³ contra R\$ 6,64/m³ no controle) e maiores despesas de exploração (IFA2004: R\$ 9,70/m³ contra R\$ 5,35/m³), o que é consistente com o perfil de municípios de maior porte e com prestadores que já operavam em escala regional antes da concessão.

Chama atenção que as despesas médias com pessoal próprio (IFA2006) são praticamente equivalentes entre os grupos, R\$ 229.495 contra R\$ 228.434 por empregado ao ano em valores reais de 2024, sugerindo que a diferença de custo operacional entre os grupos não passa pelo nível salarial, mas possivelmente pela escala e composição da força de trabalho. Em contrapartida, os municípios tratados exibem menor cobertura de abastecimento (IAG0001: 63,7% contra 74,6% no controle) e maiores perdas de faturamento (IAG2012: 38,3% contra 27,3%), sugerindo que a decisão de conceder foi motivada, ao menos em parte, por deficiências operacionais preexistentes. A elevada dispersão observada em alguns indicadores, como perdas por ligação (IAG2015)

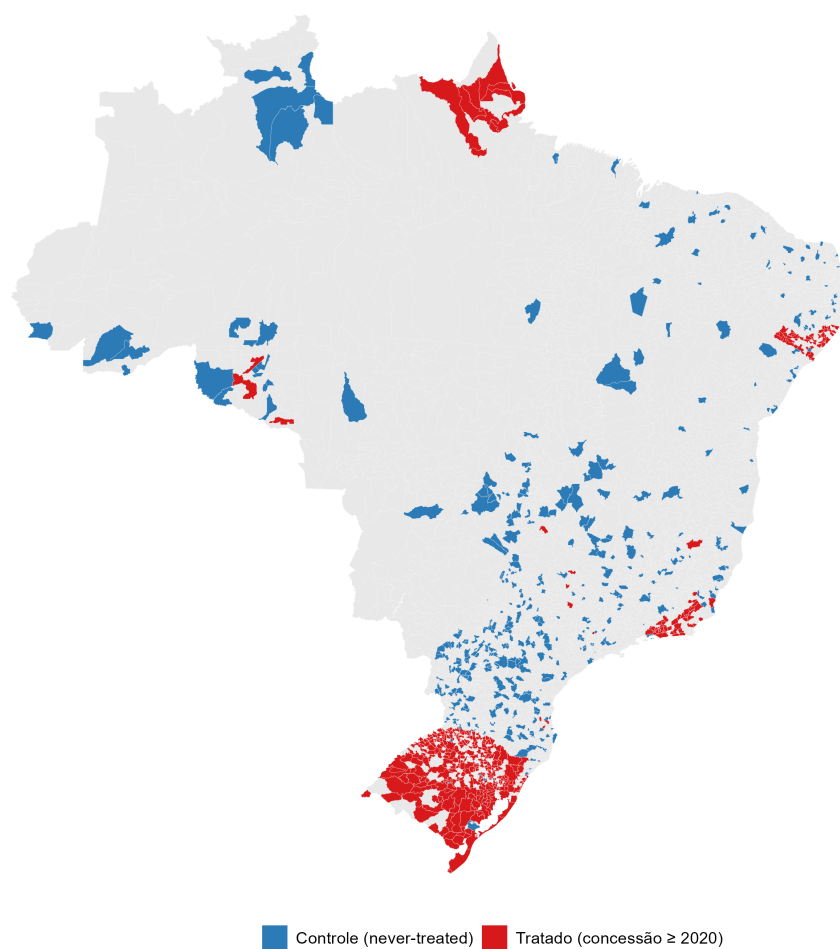


Figura 4.3 – Amostra pareada de municípios tratados e controles. Brasil, 2026. Fonte: Resultados da pesquisa.

e economias atingidas por paralisações (IAG3001), reflete a heterogeneidade estrutural dos municípios brasileiros e reforça a necessidade do estimador de Callaway e Sant’Anna (2021), que acomoda efeitos heterogêneos entre coortes e ao longo do tempo.

Tabela 4.3 – Estatísticas descritivas dos indicadores da amostra 2017-2024. Brasil, 2026.

Indicador	Controle (never-treated)				Tratado (concessão ≥ 2020)			
	N	Média	Mediana	DP	N	Média	Mediana	DP
IAG0001	3370	74.601	79.676	23.001	3381	63.682	63.920	23.316
IAG0002	2944	94.025	100.000	12.758	2952	91.036	100.000	17.560
IAG1001	2951	16.619	14.447	10.679	2968	16.124	14.135	10.303
IAG1003	2950	95.582	100.000	14.598	2968	93.827	100.000	18.468
IAG2001	3340	64.971	70.239	19.149	3388	57.211	60.897	20.999
IAG2002	3347	93.512	99.998	17.040	3383	90.967	99.570	21.251
IAG2003	3282	80.826	100.000	36.693	3158	38.061	15.441	41.544
IAG2004	2938	10.432	9.920	4.507	2935	9.112	8.930	2.576
IAG2005	2893	11.448	11.034	3.319	2874	10.808	9.255	7.145
IAG2006	2949	146.733	140.730	53.174	2967	149.516	142.708	50.640
IAG2008	2916	17.045	14.351	9.865	2967	20.424	15.329	39.431
IAG200G	2115	10.673	10.046	3.575	2119	9.998	9.088	4.022
IAG2010	3339	69.216	72.157	15.005	3384	63.300	64.845	17.061
IAG2011	3112	72.708	76.145	17.721	3281	61.662	63.400	18.085
IAG2012	3112	27.292	23.855	17.721	3281	38.338	36.600	18.085
IAG2013	3339	30.629	27.723	15.045	3384	36.348	34.782	16.931
IAG2015	2915	220.279	138.087	279.324	2960	393.794	218.895	1574.627
IAG2023	3342	2.956	0.785	53.424	3391	0.879	0.663	1.109
IAG3001	1436	3426.244	1183.835	16707.156	3000	2532.473	1148.775	6660.946
IAG3003	3362	48.547	16.305	82.486	3351	9.754	0.499	34.123
IAG3004	1209	28.257	10.652	76.492	2582	12.122	9.939	12.476
IFA1001	3317	6.643	6.687	4.441	3285	11.645	11.959	11.271
IFA1003	3365	2.908	2.207	3.137	3379	3.855	3.104	3.590
IFA1004	2515	7.312	3.764	11.332	3052	10.673	2.729	16.497
IFA1005	3319	83.166	57.377	96.026	3366	94.161	46.231	131.717
IFA2001	3372	103.137	100.715	38.301	3392	103.234	98.367	53.314
IFA2002	3367	6.723	6.102	4.694	3392	12.575	10.882	13.030
IFA2004	3370	5.345	4.663	3.575	3392	9.704	8.682	8.362
IFA2005	2951	689.693	622.463	398.676	2968	1122.532	995.481	919.945
IFA2006	2807	229495.020	200823.266	151128.768	2884	228434.022	223465.480	128977.319
IFA2007	3320	0.996	0.807	3.126	3364	1.402	0.893	12.148
IFA2008	3375	43.723	43.334	14.146	3392	47.360	49.982	17.710
IFA200G	2539	60.983	62.237	11.535	2544	62.638	64.145	14.789
IFA2010	3375	17.432	15.694	10.337	3392	14.656	11.320	12.403
IFA2011	2523	70.928	71.218	15.882	2419	70.836	72.764	17.773
IFA2012	3248	36.313	32.947	18.129	3189	39.269	36.630	21.280
IFA2013	3165	47.681	45.523	17.517	3084	48.400	46.123	20.245
IFA2014	3350	7.101	5.202	9.085	3308	4.338	1.270	11.526

Nota: DP = desvio padrão. Descrição e unidade dos indicadores — IAG0001: Atendimento da população total com rede de abastecimento de água (Percentual); IAG0002: Atendimento da população urbana com rede de abastecimento de água (Percentual); IAG1001: Extensão de rede de distribuição de água por ligação (m/ligação); IAG1003: Incidência da hidrometração de água (Percentual); IAG2001: Micromedição de água em relação ao volume disponibilizado para distribuição (Percentual); IAG2002: Micromedição do volume de água consumido (Percentual); IAG2003: Macromedição do volume de água de entrada no sistema de distribuição (Percentual); IAG2004: Consumo micromedido de água por economia (m³/economia/mês); IAG2005: Consumo de água faturado por economia (m³/economia/mês); IAG2006: Consumo total médio per capita de água (l/hab/dia); IAG2008: Volume de água disponibilizado para distribuição por economia (m³/economia/mês); IAG200G: Consumo total médio de água por economia (m³/economia/mês); IAG2010: Consumo de água em relação ao volume disponibilizado para distribuição (Percentual); IAG2011: Volume faturado de água em relação ao volume de entrada no sistema de distribuição (Percentual); IAG2012: Perdas de faturamento de água (Percentual); IAG2013: Perdas totais de água na distribuição (Percentual); IAG2015: Perdas totais de água por ligação (l/lig/dia); IAG2023: Consumo médio de energia elétrica no serviço de abastecimento de água (KWh/m³); IAG3001: Economias ativas de água atingidas por paralisações (economias/paralisação); IAG3003: Tempo médio de reparo de vazamentos de água (horas/reparo); IAG3004: Duração média das paralisações (horas/paralisação); IFA1001: Receita operacional direta média de usuários de água (R\$/m³); IFA1003: Participação da receita operacional indireta na receita operacional total de água (percentual); IFA1004: Evasão de receitas do serviço de água (percentual); IFA1005: Dias de faturamento comprometidos com contas a receber de água (dias); IFA2001: Desempenho financeiro do serviço de abastecimento de água (Percentual); IFA2002: Despesa total média de água incluindo tributos (R\$/m³); IFA2004: Despesa de exploração média de água (R\$/m³); IFA2005: Despesa de exploração média de água por economia ativa (R\$/economia.ano); IFA2006: Despesa média com pessoal próprio do serviço de abastecimento de água (R\$/empregado.ano); IFA2007: Despesa média de energia elétrica do serviço de abastecimento de água (R\$/KWh); IFA2008: Incidência da despesa de pessoal próprio nas despesas de exploração de água (Percentual); IFA200G: Incidência da despesa de pessoal total nas despesas de exploração de água (Percentual); IFA2010: Incidência da despesa de energia elétrica nas despesas de exploração de água (Percentual); IFA2011: Margem da despesa de exploração de água (Percentual); IFA2012: Margem da despesa com pessoal próprio do serviço de abastecimento de água (Percentual); IFA2013: Margem da despesa com pessoal total do serviço de abastecimento de água (Percentual); IFA2014: Margem da despesa com serviço da dívida de água (Percentual). Fonte: Resultados da pesquisa.

4.4.2 Efeitos sobre tarifas

Os resultados para receita operacional direta média de usuários de água por m^3 , expressa em reais de dezembro de 2024 (IFA1001) são apresentados na Tabela 4.4. O efeito médio não é estatisticamente significativo ($\text{ATT} = 0,815$; $p = 0,255$), o que indica ausência de impacto médio robusto da concessão plena sobre a tarifa real de água. Contudo, há heterogeneidade relevante entre coortes: a coorte de 2020 apresenta redução significativa de R\$ $1,69/\text{m}^3$ ($p = 0,002$); a coorte de 2021 registra aumento de R\$ $3,96/\text{m}^3$ ($p = 0,012$), o maior efeito pontual observado; e as coortes de 2022 e 2024 exibem reduções significativas de R\$ $2,81/\text{m}^3$ e R\$ $1,18/\text{m}^3$, respectivamente. Considerando que a receita média dos municípios tratados no período pré-concessão (2017 a 2019) era de R\$ $11,20/\text{m}^3$, as variações significativas representam entre -25% e +35% desse nível de referência, o que evidencia efeitos de magnitude economicamente relevante, ainda que em direções opostas entre coortes. A coorte de 2023 não apresenta efeito distinguível de zero. A alternância de sinais sugere que os efeitos tarifários dependem fortemente do momento da concessão e das condições contratuais específicas de cada lote, de modo que o efeito médio nulo esconde dinâmicas opostas entre grupos.

Tabela 4.4 – Efeito médio e por coortes da concessão plena sobre receita operacional direta média de usuários de água (IFA1001). Brasil, 2026.

Coorte	ATT	Erro Padrão	Valor-p	IC 95%
2020	-1,6901**	0,5356	0,0016	[-2,7398; -0,6403]
2021	3,9610*	1,5748	0,0119	[0,8744; 7,0476]
2022	-2,8147*	1,1333	0,0130	[-5,0360; -0,5935]
2023	1,4817	1,1455	0,1959	[-0,7636; 3,7270]
2024	-1,1808*	0,4879	0,0155	[-2,1370; -0,2246]
Efeito médio	0,8150	0,7166	0,2554	[-0,5896; 2,2195]

Notas: Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$; . $p < 0,10$. Fonte: Resultados da pesquisa.

O estudo de eventos da Figura 4.4 é consistente com esse padrão. No período pré-tratamento, os coeficientes oscilam próximos de zero sem tendência sistemática, o que sustenta a plausibilidade da hipótese de tendências paralelas. Após o tratamento, os efeitos são inicialmente modestos e imprecisos, com ampliação dos intervalos de confiança a partir do período 2, refletindo a heterogeneidade entre coortes e a menor quantidade de municípios observados nos períodos mais distantes do tratamento.

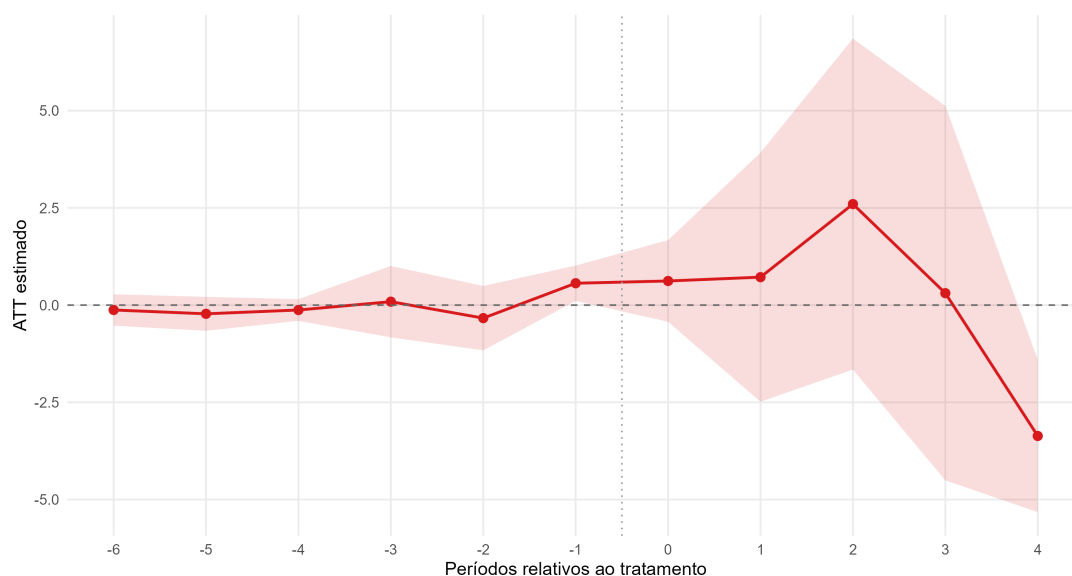


Figura 4.4 – Estudo de eventos da receita operacional direta média de usuários de água (IFA1001). Brasil, 2026. Fonte: Resultados da pesquisa.

4.4.3 Efeitos sobre atendimento

Os resultados para o índice de atendimento da população total com rede de abastecimento de água (IAG0001) são apresentados na Tabela 4.5. O efeito médio é positivo e altamente significativo ($ATT = 4,78$ p.p.; $p < 0,001$). Considerando que a cobertura média dos municípios tratados no período pré-concessão era de 63,0%, esse efeito representa um acréscimo de aproximadamente 7,6% sobre o nível de referência, o que significa uma magnitude substantiva para um serviço de infraestrutura com elevados custos de expansão de rede. Os efeitos por coorte reforçam esse padrão: as coortes de 2021, 2022 e 2023 exibem ganhos significativos de 10,74, 7,65 e 2,93 p.p., respectivamente.

O resultado da coorte de 2020 é o mais intrigante: apesar de dispor do maior horizonte pós-tratamento e, portanto, do maior tempo para que investimentos em rede se traduzam em cobertura efetiva, é a única coorte sem efeito distinguível de zero. Uma interpretação alternativa para os ganhos observados nas coortes posteriores é que o NMLSB estabelece metas de universalização, criando incentivo direto para que as concessionárias reportem avanços de cobertura. Como os dados do SINISA são autodeclarados pelo próprio prestador, parte do efeito estimado pode refletir mudança no comportamento de reporte após a assunção do contrato, amplificada pela revisão cadastral de ligações e economias que tipicamente acompanha a transição entre operadores. A ausência de efeito na coorte de 2020 seria consistente com essa hipótese: contratos celebrados no início do marco regulatório, ainda sem pressão consolidada de reporte de metas e em meio à pandemia de COVID-19, teriam menor incentivo e capacidade operacional para alterações cadastrais imediatas. O coeficiente negativo e significativo da coorte de 2024 ($-2,74$ p.p.) reflete apenas um período pós-tratamento e pode captar efeito transitório da fase inicial do contrato.

Em conjunto, esses resultados sugerem que os ganhos de cobertura reportados merecem ser interpretados com reserva, sendo necessária validação por fontes independentes para distinguir expansão real de rede de artefatos de reporte.

Tabela 4.5 – Efeito médio e por coortes da concessão plena sobre o índice de atendimento da população total com rede de abastecimento de água (IAG0001). Brasil, 2026.

Coorte	ATT	Erro Padrão	Valor-p	IC 95%
2020	0,4163	4,8155	0,9311	[-9,0221; 9,8547]
2021	10,7408***	2,7788	0,0001	[5,2945; 16,1872]
2022	7,6523***	1,3950	0,0000	[4,9180; 10,3866]
2023	2,9305***	0,6727	0,0000	[1,6120; 4,2490]
2024	-2,7413***	0,7957	0,0006	[-4,3008; -1,1817]
Efeito médio	4,7813***	0,7097	0,0000	[3,3903; 6,1722]

Notas: Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$; . $p < 0,10$. Fonte: Resultados da pesquisa.

O estudo de eventos da Figura 4.5 oferece suporte adicional à identificação causal. Os coeficientes pré-tratamento estão consistentemente próximos de zero sem tendência ascendente ou descendente clara, o que valida a hipótese de tendências paralelas. A partir do período 0, inicia-se uma trajetória de crescimento até o período 3, com intervalos de confiança que excluem zero a partir do período 1. A inflexão negativa no período 4 acompanha o padrão da coorte de 2024 e merece acompanhamento com dados futuros.

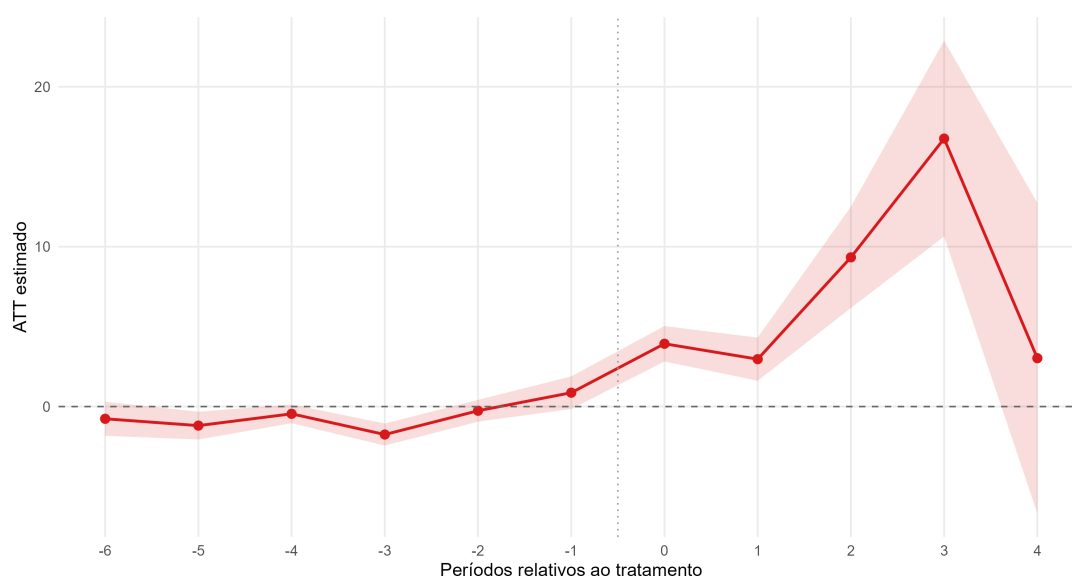


Figura 4.5 – Estudo de eventos do índice de atendimento da população total com rede de abastecimento de água (IAG0001). Brasil, 2026. Fonte: Resultados da pesquisa.

4.4.4 Efeitos sobre eficiência

Os resultados para a incidência da despesa de pessoal próprio nas despesas de exploração de água (IFA2008) são apresentados na Tabela 4.6. O efeito médio é negativo e altamente significativo (ATT = -13,81 p.p.; $p < 0,001$). Tendo em vista que a incidência média do gasto com pessoal próprio nos municípios tratados no período pré-concessão era de 54,2%, esse efeito representa uma redução de aproximadamente 25,5% sobre o nível de referência, o que indica uma reconfiguração expressiva da estrutura de custos operacionais após a concessão. Esse resultado é consistente com a lógica da privatização: operadores privados tendem a substituir quadros próprios por terceirização e a otimizar o dimensionamento de equipes. As coortes de 2020, 2021, 2022 e 2023 apresentam reduções significativas que variam entre -9,22 e -16,38 p.p., com o efeito mais intenso observado na coorte de 2023. O coeficiente da coorte de 2024 não é significativo, refletindo a limitação de apenas um período pós-tratamento disponível para essa coorte.

Tabela 4.6 – Efeito médio e por coortes da concessão plena sobre a incidência da despesa de pessoal próprio nas despesas de exploração de água (IFA2008). Brasil, 2026.

Coorte	ATT	Erro Padrão	Valor-p	IC 95%
2020	-11,4114*	5,1619	0,0271	[-21,5288; -1,2940]
2021	-11,1013**	3,4019	0,0011	[-17,7689; -4,4336]
2022	-9,2242***	2,2798	0,0001	[-13,6926; -4,7558]
2023	-16,3826***	0,6734	0,0000	[-17,7025; -15,0627]
2024	22,1879	26,2320	0,3976	[-29,2267; 73,6026]
Efeito médio	-13,8111***	0,9065	0,0000	[-15,5879; -12,0344]

Notas: Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Significância: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$; . $p < 0,10$. Fonte: Resultados da pesquisa.

O estudo de eventos da Figura 4.6 apresenta um padrão distinto dos demais. Entre os períodos -5 e -1, os coeficientes convergem para zero com precisão crescente, o que sustenta a validade da hipótese de tendências paralelas no período relevante. O coeficiente negativo e de grande magnitude no período -6 pode ser atribuído ao número reduzido de municípios com seis anos de antecedência observados. A partir do período 0, instala-se uma redução sistemática e significativa da incidência do gasto com pessoal, que atinge seu pico negativo no período 1 e se mantém em patamares negativos até o período 3, antes de uma reversão parcial no período 4, novamente com intervalo de confiança amplo dado o horizonte mais distante.

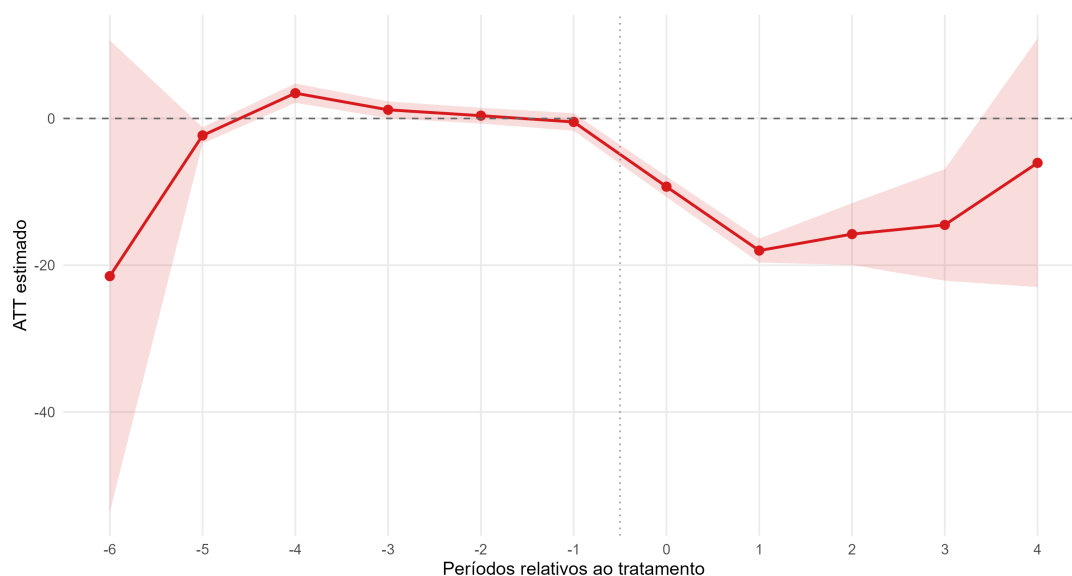


Figura 4.6 – Estudo de eventos da incidência da despesa de pessoal próprio nas despesas de exploração de água (IFA2008). Brasil, 2026. Fonte: Resultados da pesquisa.

4.4.5 Demais efeitos

Os resultados agregados (efeitos médios) para os demais indicadores dos grupos IAG e IFA são apresentados nas Tabelas 4.7 e 4.8 e revelam um conjunto rico de efeitos que permitem caracterizar com maior precisão os mecanismos pelos quais a concessão plena afeta a prestação do serviço de abastecimento de água.

Tabela 4.7 – ATT Agregado — Grupo IAG. Estimador Callaway & Sant’Anna (2021). Brasil, 2017–2024.

Indicador	ATT	Erro Padrão	Valor-p	IC 95%
IAG0002	-0.8118	0.7346	0.2691	[-2.2515; 0.6280]
IAG1001	0.1572	0.2673	0.5565	[-0.3667; 0.6810]
IAG1003	-0.3488	0.5573	0.5313	[-1.4411; 0.7434]
IAG2001	-4.5689***	0.8396	0.0000	[-6.2145; -2.9232]
IAG2002	-1.3560.	0.8056	0.0923	[-2.9350; 0.2230]
IAG2003	26.5968***	2.3348	0.0000	[22.0206; 31.1730]
IAG2004	-0.3115	0.2323	0.1799	[-0.7668; 0.1437]
IAG2005	2.8654***	0.3732	0.0000	[2.1338; 3.5969]
IAG2006	-3.2288	2.7534	0.2409	[-8.6255; 2.1679]
IAG2008	3.1556	2.6981	0.2422	[-2.1327; 8.4440]
IAG200G	-1.5853**	0.6137	0.0098	[-2.7881; -0.3824]
IAG2010	-4.1873***	0.9154	0.0000	[-5.9814; -2.3932]
IAG2011	0.6347	1.0989	0.5635	[-1.5191; 2.7886]
IAG2012	-0.6347	1.1087	0.5670	[-2.8079; 1.5384]
IAG2013	3.5355***	0.8544	0.0000	[1.8608; 5.2102]
IAG2015	143.7926	117.2206	0.2199	[-85.9598; 373.5450]
IAG2023	-7.3339*	3.3375	0.0280	[-13.8755; -0.7923]
IAG3001	-1572.6922	1231.5596	0.2016	[-3986.5490; 841.1646]
IAG3003	22.1805***	2.8575	0.0000	[16.5798; 27.7813]
IAG3004	7.2489	5.2589	0.1681	[-3.0586; 17.5564]

Notas: Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Significância: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$. Descrição e unidade dos indicadores — IAG0001: Atendimento da população total com rede de abastecimento de água (Percentual); IAG0002: Atendimento da população urbana com rede de abastecimento de água (Percentual); IAG1001: Extensão de rede de distribuição de água por ligação (m/ligação); IAG1003: Incidência da hidrometração de água (Percentual); IAG2001: Micromedição de água em relação ao volume disponibilizado para distribuição (Percentual); IAG2002: Micromedição do volume de água consumido (Percentual); IAG2003: Macromedição do volume de água de entrada no sistema de distribuição (Percentual); IAG2004: Consumo micromedido de água por economia (m^3 /economia/mês); IAG2005: Consumo de água faturado por economia (m^3 /economia/mês); IAG2006: Consumo total médio per capita de água (l/hab/dia); IAG2008: Volume de água disponibilizado para distribuição por economia (m^3 /economia/mês); IAG200G: Consumo total médio de água por economia (m^3 /economia/mês); IAG2010: Consumo de água em relação ao volume disponibilizado para distribuição (Percentual); IAG2011: Volume faturado de água em relação ao volume de entrada no sistema de distribuição (Percentual); IAG2012: Perdas de faturamento de água (Percentual); IAG2013: Perdas totais de água na distribuição (Percentual); IAG2015: Perdas totais de água por ligação (l/lig/dia); IAG2023: Consumo médio de energia elétrica no serviço de abastecimento de água (KWh/ m^3); IAG3001: Economias ativas de água atingidas por paralisações (economias/paralisação); IAG3003: Tempo médio de reparo de vazamentos de água (horas/reparo); IAG3004: Duração média das paralisações (horas/paralisação). Fonte: Resultados da pesquisa.

Tabela 4.8 – ATT Agregado — Grupo IFA. Estimador Callaway & Sant’Anna (2021). Brasil, 2017–2024.

Indicador	ATT	Erro Padrão	Valor-p	IC 95%
IFA1003	0.5607.	0.2936	0.0562	[-0.0148; 1.1361]
IFA1004	9.0315***	0.9670	0.0000	[7.1361; 10.9269]
IFA1005	-25.4181***	5.3486	0.0000	[-35.9012; -14.9349]
IFA2001	-17.9977***	3.8467	0.0000	[-25.5373; -10.4581]
IFA2002	0.5233	2.0981	0.8030	[-3.5890; 4.6356]
IFA2004	-1.8028**	0.5724	0.0016	[-2.9247; -0.6809]
IFA2005	27.9005	59.6242	0.6398	[-88.9629; 144.7639]
IFA2006	-62756.8527***	9648.6136	0.0000	[-81668.1354; -43845.5699]
IFA2007	1.2156.	0.6872	0.0769	[-0.1313; 2.5624]
IFA200G	-13.4498***	1.6429	0.0000	[-16.6699; -10.2296]
IFA2010	-4.2520***	0.7082	0.0000	[-5.6401; -2.8639]
IFA2011	-5.1616***	1.4612	0.0004	[-8.0256; -2.2976]
IFA2012	-15.2796***	0.9711	0.0000	[-17.1829; -13.3763]
IFA2013	-13.2161***	1.1697	0.0000	[-15.5088; -10.9234]
IFA2014	8.6145***	1.0473	0.0000	[6.5618; 10.6673]

Notas: Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Significância: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$. Descrição dos indicadores — IFA1001: Receita operacional direta média de usuários de água (R\$/m³); IFA1003: Participação da receita operacional indireta na receita operacional total de água (percentual); IFA1004: Evasão de receitas do serviço de água (percentual); IFA1005: Dias de faturamento comprometidos com contas a receber de água (dias); IFA2001: Desempenho financeiro do serviço de abastecimento de água (Percentual); IFA2002: Despesa total média de água incluindo tributos (R\$/m³); IFA2004: Despesa de exploração média de água (R\$/m³); IFA2005: Despesa de exploração média de água por economia ativa (R\$/economia.ano); IFA2006: Despesa média com pessoal próprio do serviço de abastecimento de água (R\$/empregado.ano); IFA2007: Despesa média de energia elétrica do serviço de abastecimento de água (R\$/KWh); IFA2008: Incidência da despesa de pessoal próprio nas despesas de exploração de água (Percentual); IFA200G: Incidência da despesa de pessoal total nas despesas de exploração de água (Percentual); IFA2010: Incidência da despesa de energia elétrica nas despesas de exploração de água (Percentual); IFA2011: Margem da despesa de exploração de água (Percentual); IFA2012: Margem da despesa com pessoal próprio do serviço de abastecimento de água (Percentual); IFA2013: Margem da despesa com pessoal total do serviço de abastecimento de água (Percentual); IFA2014: Margem da despesa com serviço da dívida de água (Percentual). Fonte: Resultados da pesquisa.

4.4.5.1 Micromedição e consumo

Os indicadores de micromedição apresentam deterioração significativa após a concessão. A micromedição em relação ao volume disponibilizado para distribuição (IAG2001) recua -4,57 p.p. ($p < 0,001$) enquanto que a micromedição do volume consumido (IAG2002) apresenta queda marginal de -1,36 p.p. ($p = 0,092$). Esse resultado é contraintuitivo à luz da expectativa de que operadores privados teriam maior incentivo à hidrometração, uma vez que a micromedição é instrumento central de cobrança dos usuários e do controle de perdas, outra meta do NMLSB. Uma hipótese é que a transição de operadores gere descontinuidade temporária nos sistemas de leitura e cadastro, o que reduziria esses índices no curto prazo. Alternativamente, e em linha

com a ressalva levantada para IAG0001, o resultado pode refletir mudança na metodologia de reporte entre o operador público e o privado. Já a incidência da hidrometração (IAG1003) não apresenta efeito significativo (ATT = -0,35 p.p.; $p = 0,531$), o que sugere que a base instalada de hidrômetros não se alterou de forma sensível no horizonte observado.

No que tange ao consumo, o volume faturado por economia (IAG2005) registra aumento significativo de 2,87 m³/economia/mês ($p < 0,001$), ao passo que o consumo micromedido por economia (IAG2004) e o consumo total médio por economia (IAG200G) não apresentam variações robustas, com o último exibindo leve queda de -1,59 m³/economia/mês ($p = 0,010$). A divergência entre o volume faturado e o consumido é consistente com o aumento simultâneo da evasão de receitas (IFA1004, discutido adiante): o operador privado pode estar faturando volumes maiores sem que o consumo efetivo aumente, possivelmente pela incorporação de economias anteriormente não faturadas ou pela revisão de categorias tarifárias.

4.4.5.2 Macromedição e perdas

A macromedição do volume de entrada no sistema de distribuição (IAG2003) apresenta o maior efeito positivo do grupo IAG (ATT = +26,60 p.p.; $p < 0,001$), sugerindo expansão expressiva da capacidade de monitoramento do sistema de abastecimento. Esse resultado é coerente com a expectativa de que operadores privados invistam em instrumentação da rede como pré-requisito para a gestão de perdas e o cumprimento de metas contratuais.

De modo paradoxal, os indicadores de perdas não apresentam melhora. O consumo em relação ao volume disponibilizado (IAG2010) recua -4,19 p.p. ($p < 0,001$), indicando que a parcela do volume que efetivamente chega às economias diminuiu. As perdas totais na distribuição (IAG2013) aumentam +3,54 p.p. ($p < 0,001$), e o indicador de perdas por ligação (IAG2015) apresenta coeficiente positivo de 143,79 l/lig/dia, embora sem significância estatística ($p = 0,220$). Ainda, o volume faturado em relação ao volume de entrada e perdas de faturamento (respectivamente, IAG2011 e IAG2012) não apresentam efeitos significativos. O conjunto desses resultados sugere que, apesar do investimento em macromedição, as concessões plenas ainda não se traduziram em redução efetiva de perdas no período analisado, o que pode refletir o caráter de longo prazo dos investimentos necessários em reabilitação de redes, ou ainda o fato de que a macromedição é condição necessária mas não suficiente para a redução de perdas.

4.4.5.3 Qualidade operacional

Os indicadores de qualidade operacional revelam piora significativa após a concessão. O tempo médio de reparo de vazamentos (IAG3003) aumentou 22,18 horas/reparo ($p < 0,001$), e a duração média das paralisações (IAG3004) apresentou aumento de 7,25 horas/paralisação, embora sem significância estatística ($p = 0,168$). As economias atingidas por paralisações (IAG3001) exibem coeficiente negativo e de grande magnitude (-1.572,69 economias/paralisação), mas também sem significância ($p = 0,202$). O aumento no tempo de reparo de vazamentos é

preocupante e pode refletir a reorganização das equipes de manutenção no período de transição, com eventual redução do quadro próprio, consistente com a queda de IFA2008, ainda não compensada por uma rede de terceiros igualmente responsiva. O consumo médio de energia elétrica (IAG2023) apresenta redução significativa de $-7,33 \text{ KWh/m}^3$ ($p = 0,028$), sugerindo ganhos de eficiência energética que podem estar associados à modernização de equipamentos de bombeamento, por exemplo.

4.4.5.4 Dimensão financeira

A tabela do grupo IFA revela uma reconfiguração financeira ampla e estatisticamente robusta. Do lado das receitas, a evasão de receitas (IFA1004) aumenta $+9,03 \text{ p.p.}$ ($p < 0,001$), indicando que a inadimplência ou o volume não cobrado cresceu após a concessão, possivelmente em decorrência de reajustes tarifários ou da incorporação de áreas com menor capacidade de pagamento. Em contrapartida, o número de dias de faturamento comprometidos com contas a receber (IFA1005) reduz $-25,42 \text{ dias}$ ($p < 0,001$), sugerindo que, embora a evasão aumente, o operador privado é mais eficiente na cobrança dos valores faturados, reduzindo o prazo médio de recebimento.

O desempenho financeiro (IFA2001), razão entre receitas e despesas, recua $-17,99 \text{ p.p.}$ ($p < 0,001$), o que indica deterioração do equilíbrio financeiro operacional. Esse resultado, combinado com a ausência de efeito sobre IFA1001 (tarifa) e IFA2002 (despesa total por m^3), sugere que a pressão sobre o resultado financeiro não decorre de aumento de custos unitários, mas possivelmente de expansão de atendimento a áreas menos rentáveis, o que pode ser considerado coerente com os ganhos de cobertura reportados em IAG0001 ou de aumento da evasão de receitas identificado em IFA1004.

Do lado dos custos, a despesa de exploração média por m^3 (IFA2004) reduziu $-1,80 \text{ R\$/m}^3$ ($p = 0,002$), sinalizando ganhos de eficiência operacional nos custos diretos de exploração. Esse resultado é reforçado pelos indicadores de composição de despesa: IFA200G ($-13,45 \text{ p.p.}$) indica forte redução da incidência de pessoal total nas despesas de exploração; IFA2010 ($-4,25 \text{ p.p.}$) aponta redução da incidência de energia elétrica; e IFA2011 ($-5,16 \text{ p.p.}$), IFA2012 ($-15,28 \text{ p.p.}$) e IFA2013 ($-13,22 \text{ p.p.}$) confirmam a queda nas margens de despesa com pessoal próprio e total. Em conjunto, esses resultados são consistentes com um processo de reestruturação de custos operacionais centrado na substituição de mão de obra própria, capturado também por pela queda da despesa média com pessoal próprio por empregado (IFA2006) de $-\text{R\$ } 62.757/\text{empregado.ano}$ ($p < 0,001$), magnitude que reflete tanto redução do quadro quanto possível alteração na composição salarial da força de trabalho remanescente.

Por fim, a margem da despesa com serviço da dívida (IFA2014) aumentou $+8,61 \text{ p.p.}$ ($p < 0,001$), o que é esperado: operadores privados financiam investimentos via endividamento, e o serviço da dívida passa a representar parcela crescente das despesas de exploração. Esse aumento é o contraponto natural da redução de pessoal e pode indicar substituição de custo fixo de mão de

obra por custo financeiro de capital, uma reconfiguração estrutural da função de custos que, no longo prazo, depende da capacidade de geração de receita para ser sustentável.

4.4.6 Efeitos sobre morbidade hospitalar

Os efeitos da concessão plena sobre a morbidade hospitalar por doenças relacionadas ao saneamento são apresentados na Tabela 4.9. Os resultados não revelam impacto estatisticamente significativo em nenhum dos grupos etários analisados: o ATT é de 1,91 internações para menores de 1 ano ($p = 0,503$), 4,21 para crianças de 1 a 4 anos ($p = 0,500$) e 6,98 para o total da população ($p = 0,336$), com intervalos de confiança amplos que incluem zero em todos os casos.

Tabela 4.9 – Efeito da concessão plena sobre morbidade hospitalar por doenças relacionadas ao saneamento. Estimador Callaway & Sant’Anna (2021). Brasil, 2017–2024.

Grupo etário	ATT	Erro Padrão	IC 95%	Valor-p
< 1 ano	1.9089	2.8489	[-3.6749; 7.4926]	0.5028
1–4 anos	4.2075	6.2410	[-8.0249; 16.4399]	0.5002
Total	6.9791	7.2532	[-7.2371; 21.1954]	0.3359

Notas: ATT overall (agregado simples). Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Anos 2020–2022 excluídos (pandemia COVID-19). Significância: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$. Fonte: Resultados da pesquisa.

A ausência de efeito detectável é interpretável à luz de pelo menos três considerações. Primeira, o horizonte temporal disponível pode ser insuficiente para captar impactos sobre saúde: melhorias na qualidade e cobertura do abastecimento de água tendem a se traduzir em reduções de morbidade com defasagem de anos, na medida em que dependem não apenas da expansão da rede mas também de mudanças comportamentais e da redução de exposição acumulada a fatores de risco. Segunda, os próprios resultados operacionais discutidos anteriormente sugerem que as concessões ainda não produziram melhoras inequívocas em qualidade e perdas no horizonte observado, o que torna esperada a ausência de efeitos sobre saúde nesse período. Terceira, a exclusão dos anos de 2020 a 2022 em razão da pandemia de COVID-19 reduz substancialmente o número de períodos pós-tratamento disponíveis para as coortes de 2020 e 2021, comprometendo a precisão das estimativas para esses grupos e contribuindo para os erros-padrão elevados observados na tabela.

Cabe notar que os coeficientes positivos, embora não significativos, são consistentes com a deterioração de indicadores operacionais como o aumento do tempo de reparo de vazamentos (IAG3003) e o aumento de perdas totais (IAG2013) identificados na seção anterior, que poderiam, em tese, elevar a exposição a fontes contaminadas. Contudo, a imprecisão das estimativas impede qualquer inferência causal nessa direção. Em conjunto, os resultados de morbidade reforçam a necessidade de um horizonte de avaliação mais longo para que os efeitos das concessões plenas sobre saúde pública possam ser adequadamente mensurados.

4.4.7 Teste de Robustez

Os resultados dos testes de robustez são apresentados nas Tabelas 4.10 e 4.11, que reestimam todos os modelos variando a proporção de pareamentos entre 1:1 (da seção de resultados), 1:2 e 1:5. A lógica do exercício é verificar se os resultados principais são sensíveis à escolha do número de controles por tratado: uma proporção maior de controles aumenta o tamanho da amostra e potencialmente a precisão das estimativas, ao custo de incorporar controles mais distantes no espaço de covariadas.

Tabela 4.10 – Análise de robustez — ATT médio por tamanho de amostra PSM. Estimador Callaway & Sant'Anna (2021). Amostra PSM-FLEXÍVEL com proporções de pareamento 1:1, 1:2 e 1:5. Brasil, 2017–2024.

Indicador	ATT (1:1)	ATT (1:2)	ATT (1:5)
IAG			
IAG0001	4.7813***	4.5195***	3.8161***
IAG0002	-0.8118	-0.3576	0.2064
IAG1001	0.1572	0.0377	-0.0043
IAG1003	-0.3488	-0.3194	-0.6265
IAG2001	-4.5689***	-4.5094***	-4.3832***
IAG2002	-1.3560.	-1.4010.	-1.3950.
IAG2003	26.5968***	26.4079***	26.0204***
IAG2004	-0.3115	-0.3966*	-0.4620**
IAG2005	2.8654***	2.6509***	-1.9293
IAG2006	-3.2288	-6.0179*	-13.4707
IAG2008	3.1556	2.8095	0.8910
IAG200G	-1.5853**	-1.5321*	-1.5502*
IAG2010	-4.1873***	-4.2808***	-4.3828***
IAG2011	0.6347	0.3303	0.0547
IAG2012	-0.6347	-0.3303	-0.0547
IAG2013	3.5355***	3.5786***	3.5271***
IAG2015	143.7926	137.6514	111.5468
IAG2023	-7.3339*	-5.4384**	-4.9323**
IAG3001	-1572.6922	-1591.5624*	-2277.7422**
IAG3003	22.1805***	24.4077***	19.2519***
IAG3004	7.2489	10.1484**	13.4977***
IFA			
IFA1001	0.8150	0.7013	14.5509
IFA1003	0.5607*	0.5692*	0.5468.
IFA1004	9.0315***	9.0871***	9.6420***
IFA1005	-25.4181***	-24.4755***	-26.3663***
IFA2001	-17.9977***	-16.9410***	-18.9766***
IFA2002	0.5233	0.4838	0.9927
IFA2004	-1.8028**	-1.8863***	-1.4673*
IFA2005	27.9005	0.3491	-61.9684
IFA2006	-62756.8527***	-76003.3799***	-77919.3476***
IFA2007	1.2156.	1.1793	0.9584
IFA2008	-13.8111***	-13.6304***	-12.6885***
IFA200G	-13.4498***	-12.3046***	-11.5286***
IFA2010	-4.2520***	-4.2942***	-4.7013***
IFA2011	-5.1616***	-4.7311**	-4.6052***
IFA2012	-15.2796***	-15.3431***	-14.2038***
IFA2013	-13.2161***	-12.9039***	-11.6534***
IFA2014	8.6145***	8.5338***	8.6895***

Notas: Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Significância: *** p<0.001; ** p<0.01; * p<0.05; . p<0.10. Fonte: Resultados da pesquisa.

Tabela 4.11 – Análise de robustez — Efeito sobre morbidade hospitalar por tamanho de amostra PSM. Callaway & Sant’Anna (2021). Brasil, 2017–2024.

Grupo etário	ATT (1:1)	ATT (1:2)	ATT (1:5)
< 1 ano	1.9089	0.9825	0.1492
1–4 anos	4.2075	1.8612	0.1447
Total	6.9791	2.8174	1.2432

Notas: ATT overall (agregado simples). Erros-padrão clusterizados por município. Método doubly-robust (DR). Controles: never-treated. Anos 2020–2022 excluídos (pandemia COVID-19). Significância: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$. Fonte: Resultados da pesquisa.

De modo geral, os resultados são notavelmente estáveis. Os indicadores que apresentaram significância estatística na especificação principal (1:1) mantêm sinal e significância em todas as três amostras, e os que não eram significativos permanecem sem resultado robusto. Destacam-se como particularmente estáveis os efeitos sobre IAG0001 (+4,78, +4,52 e +3,82 p.p.), IAG2001 (-4,57, -4,51 e -4,38 p.p.), IAG2003 (+26,60, +26,41 e +26,02 p.p.), IAG2010 (-4,19, -4,28 e -4,38 p.p.) e IAG2013 (+3,54, +3,58 e +3,53 p.p.) no grupo IAG, e IFA1004 (+9,03, +9,09 e +9,64 p.p.), IFA1005 (-25,42, -24,48 e -26,37 dias), IFA2001 (-18,00, -16,94 e -18,98 p.p.), IFA2004 (-1,80, -1,89 e -1,47 R\$/m³), IFA2008 (-13,81, -13,63 e -12,69 p.p.) e IFA2014 (+8,61, +8,53 e +8,69 p.p.) no grupo IFA. A estabilidade desses coeficientes ao longo de três tamanhos amostrais distintos reforça a confiabilidade das estimativas principais.

Algumas exceções merecem atenção. O volume faturado por economia (IAG2005) apresenta efeito positivo e significativo nos proporções de pareamento 1:1 e 1:2, mas perde sinal e significância no ratio 1:5, o que sugere que o efeito é sensível à composição da amostra de controles. A receita operacional direta por m³ (IFA1001) que já não era significativa na especificação principal, apresenta coeficiente de 14,55 no ratio 1:5, muito distante dos valores obtidos nos proporções de pareamento 1:1 (0,82) e 1:2 (0,70), sinalizando instabilidade que pode refletir a incorporação de controles muito distantes no espaço de covariadas nessa especificação mais ampla. De modo semelhante, IFA2005 e IAG2006 apresentam variações de magnitude e sinal entre os proporções de pareamento. IAG3001 e IAG3004, por sua vez, tornam-se significativos nos proporções de pareamento 1:2 e 1:5 mas não no 1:1, o que indica precisão insuficiente na amostra mais restrita e não necessariamente ausência de efeito real, resultado esse que merece acompanhamento com dados futuros.

Para a morbidade hospitalar, os coeficientes declinam monotonicamente com o aumento da razão de pareamento (de 1,91, 4,21 e 6,98 na proporção 1:1 para 0,15, 0,14 e 1,24 no ratio 1:5) e permanecem não significativos em todas as especificações e grupos etários, confirmando a robustez da ausência de efeito detectável sobre saúde no horizonte analisado.

4.5 Discussão

Os resultados deste estudo permitem uma avaliação dos efeitos das concessões plenas de abastecimento de água no Brasil após o Novo Marco Legal do Saneamento Básico (Lei n.º 14.026/2020). De forma conjunta, os achados mostram uma trajetória de reestruturação operacional e financeira nos municípios concedidos que é, ao mesmo tempo, consistente com ganhos gerenciais de curto prazo e inconclusiva quanto à melhoria efetiva da qualidade do serviço, tensão que remete diretamente ao debate teórico sobre o dilema custo-qualidade na provisão privada de serviços públicos.

O dilema custo-qualidade no contexto do NMLSB

A hipótese do dilema custo-qualidade, formulada por Hart, Shleifer e Vishny (1997), é a de que, em setores nos quais a qualidade é não-contratável e inexistem mecanismos de mercado que remunerem sua manutenção, os incentivos à redução de custos de um operador privado podem conduzir à deterioração do nível de serviço ofertado. O setor de saneamento reúne as condições que tornam esse dilema potencialmente presente: os consumidores não têm escolha entre fornecedores, os indicadores de qualidade são de difícil verificação e, historicamente, os contratos de concessão brasileiros não previam metas explícitas de qualidade (Saiani; Azevedo, 2018). Como sintetiza Lazzarini (2023, p. 24), a privatização pode ser especialmente problemática nos casos em que os operadores privados tenham grandes lucros potenciais ao ignorar dimensões sociais ou públicas difíceis de especificar em contrato, e o saneamento, por sua natureza de monopólio natural e bem essencial, se encaixa nessa descrição. Ainda, como os governos frequentemente têm dificuldade para elaborar contratos que especifiquem os atributos de serviço exigidos (Lazzarini, 2023, p. 24), a completude contratual é o ponto institucional central que determina se o dilema se manifesta ou não na prática.

Os resultados aqui obtidos são evidências empíricas que permitem avaliar em que medida esse dilema se manifesta nas concessões recentes. Do lado dos custos, os achados são inequívocos: houve redução da despesa de exploração por m³ (IFA2004: –R\$ 1,80/m³), queda expressiva da incidência de pessoal próprio nas despesas de exploração (IFA2008: –13,81 p.p.) e da despesa por empregado (IFA2006: –R\$ 62.757/empregado.ano), além de redução da incidência de energia elétrica (IFA2010: –4,25 p.p.) e das margens de pessoal (IFA2012 e IFA2013). Esse conjunto de resultados configura uma reestruturação de custos operacionais baseada na substituição de mão de obra própria, consistente com o padrão documentado por Dassan *et al.* (2024), que identificam reduções tarifárias nos primeiros anos após a privatização, e com o argumento de Saiani e Azevedo (2018) de que ganhos de eficiência gerencial são o principal vetor de resultado de curto prazo nas concessões locais.

A dimensão da qualidade, porém, apresenta resultados mais ambíguos e preocupantes em alguns aspectos. O aumento do tempo médio de reparo de vazamentos (IAG3003: +22,18 horas/-

reparo) e o crescimento das perdas totais na distribuição (IAG2013: +3,54 p.p.) sugerem que a redução do quadro de pessoal próprio pode ter gerado descontinuidades na capacidade operacional de manutenção da rede no horizonte observado. Embora o investimento em macromedição tenha sido expressivo (IAG2003: +26,60 p.p.), o que é condição necessária para uma gestão eficiente de perdas, os indicadores de perdas não melhoraram em contrapartida, o que é consistente com o argumento de que essas transformações requerem horizontes temporais mais longos e investimentos complementares de reabilitação de redes (Dassan *et al.*, 2024).

É precisamente nessa tensão entre a redução de custos operacionais e a ausência de melhoria nos indicadores de qualidade que a hipótese do dilema encontra espaço para ser avaliada. Cabe aqui afirmar que não se deve confundir a confirmação parcial do dilema com uma condenação genérica da privatização. Como adverte Lazzarini (2023, p. 51), em vez de escolher a dedo um argumento teórico e defender de forma genérica a atuação de empresas públicas ou privadas, o exercício adequado é analisar a fundo os prós e contras de cada alternativa e propor maneiras de lidar com seus dilemas inerentes. Nesse sentido, os resultados aqui obtidos apontam não para a inviabilidade da concessão privada, mas para a centralidade do desenho contratual e da capacidade regulatória do Estado como determinantes do seu sucesso.

Uma interpretação possível é a de que o dilema está presente, ainda que de forma incipiente: os operadores privados reestruturam custos com relativa rapidez, mas o repasse desses ganhos para a qualidade do serviço ainda não se materializou dentro do horizonte aqui analisado. Alternativamente, e em linha com Saiani e Azevedo (2018), os resultados operacionais negativos podem refletir efeitos transitórios do período de transição entre operadores, e não uma trajetória estrutural de deterioração de qualidade. A distinção entre essas interpretações exige dados de mais longo prazo.

Tarifas e equilíbrio financeiro

A questão tarifária é um dos temas mais debatidos na literatura sobre privatizações de saneamento. A evidência internacional é diversa: aumentos expressivos foram documentados em países como Hungria e República Checa após a abertura ao setor privado (Lobina; Hall, 2000), enquanto estudos para a Argentina apontam que reduções tarifárias iniciais foram rapidamente revertidas por reajustes subsequentes (Galiani; Gertler; Schargrotsky, 2005). Para o Brasil, Dassan *et al.* (2024) documentam que as tarifas reais caem nos primeiros anos após a privatização, mas crescem a taxas mais elevadas ao longo do tempo entre os operadores privados vis-à-vis os públicos.

Os resultados do presente estudo são consistentes com esse padrão heterogêneo: o efeito médio sobre IFA1001 não é estatisticamente significativo, mas as coortes individuais revelam variações que alcançam entre -25% e +35% do nível de referência pré-concessão. A coorte de 2021 registra o maior aumento, de R\$ 3,96/m³, enquanto as coortes de 2020, 2022 e 2024 exibem reduções significativas. Essa alternância de sinais não é aleatória, ela reflete diferenças no

modelo de leilão adotado em cada rodada. As primeiras concessões realizadas após o NMLSB, como as dos blocos de Alagoas e Rio de Janeiro, adotaram o critério de maior outorga como único parâmetro de julgamento. Fagundes, Cunha Marques e Malheiros (2024) analisam essas concessões e documentam impactos tarifários de 13,2% a 46,3% sobre as tarifas médias, com comprometimento expressivo da renda das famílias mais vulneráveis — chegando a 10% da renda dos beneficiários do *Bolsa Família* em Alagoas, mesmo com desconto de tarifa social. Os autores concluem que o modelo baseado exclusivamente na maior outorga é perverso do ponto de vista da modicidade tarifária, uma vez que transfere para os usuários o custo de um prêmio bilionário sem qualquer garantia de acessibilidade. Essa evidência é consistente com o aumento tarifário observado na coorte de 2021 no presente estudo, que corresponde justamente ao período em que as primeiras concessões baseadas nesse modelo entraram em operação.

De modo diferente, os leilões posteriores passaram a adotar o critério de menor tarifa, o que altera estruturalmente os incentivos do concessionário (Fagundes; Cunha Marques; Malheiros, 2024; ANA, 2024). As coortes de 2022 e 2024, nas quais são observadas reduções tarifárias significativas, são consistentes com esse redesenho regulatório. Esse padrão ilustra um ponto fundamental destacado por Lazzarini (2023, p. 280): nos monopólios naturais, uma regulamentação tarifária defeituosa pode fazer com que os preços estabelecidos sejam muito altos, provocando divergências públicas; por outro lado, quando o governo tenta manter os preços artificialmente baixos, o investimento privado pode ser insuficiente. O equilíbrio entre modicidade tarifária e sustentabilidade do investimento não se resolve automaticamente com a privatização, ele depende do desenho do processo licitatório e de regras regulatórias transparentes e estáveis.

O quadro financeiro mais amplo merece análise adicional. A deterioração do desempenho financeiro operacional (IFA2001: -17,99 p.p.) combinada com o aumento da evasão de receitas (IFA1004: +9,03 p.p.) e a elevação da margem de serviço da dívida (IFA2014: +8,61 p.p.) configuram uma reconfiguração estrutural da função de custos: a redução de pessoal próprio é parcialmente substituída por custo de capital financeiro, o que é esperado em modelos de concessão que dependem de endividamento para financiar investimentos. Nesse sentido, Fagundes, Cunha Marques e Malheiros (2024) alertam que os prêmios de outorga bilionários, custeados integralmente pelas tarifas, podem comprometer a viabilidade financeira de longo prazo dos contratos e aumentar a probabilidade de renegociações precoces, historicamente desfavoráveis ao interesse público. A sustentabilidade de longo prazo do modelo de concessão depende, portanto, não apenas da capacidade operacional do concessionário, mas da adequação do processo de licitação que deu origem ao contrato.

Cobertura, inclusão e os limites do mercado

O efeito positivo e significativo sobre o atendimento total (IAG0001: +4,78 p.p.) é um achado que por um lado, alinha-se à literatura que documenta expansão de cobertura após concessões plenas no Brasil (Dassan *et al.*, 2024) e em outros contextos, como a Argentina (Galiani; Gertler;

Schargrodsky, 2005). Por outro lado, a ausência de efeito significativo justamente na coorte de 2020, a mais antiga e, portanto, com maior horizonte pós-tratamento, levanta dúvidas sobre se os ganhos observados nas coortes posteriores refletem expansão física de rede ou são puramente vieses de declaração motivados pelos incentivos de metas contratuais do NMLSB.

A questão da cobertura toca num ponto estrutural do debate sobre privatização e inclusão. Lazzarini (2023, p. 58) observa que os provedores privados podem se concentrar nas populações de alta renda, que conseguem pagar não apenas tarifas variáveis pela água que consomem, mas também tarifas fixas de conexão para cobrir os investimentos iniciais em infraestrutura, enquanto que a população de baixa renda, mesmo quando o serviço está presente em sua área, pode não conseguir pagar essas tarifas e buscar fontes alternativas e menos seguras. Esse risco é particularmente relevante no contexto brasileiro, em que os municípios concedidos apresentavam, antes do tratamento, cobertura média de apenas 63,0%, nível bastante inferior ao grupo de controle (74,6%), sugerindo que a expansão para os estratos não atendidos (tipicamente mais pobres e mais distantes) é o principal desafio de universalização que cabe às novas concessões enfrentar. Os resultados operacionais e financeiros aqui documentados não permitem concluir se esse desafio está sendo cumprido, mas indicam que a pressão de curto prazo sobre a rentabilidade operacional pode atuar em sentido contrário ao da inclusão.

Como os dados do SINISA são autodeclarados pelo próprio prestador, uma revisão cadastral de ligações e economias na assunção do contrato pode inflar o indicador de cobertura sem que haja conexão efetiva de novas residências. A validação desses ganhos por fontes independentes, como pesquisas domiciliares, é condição necessária para interpretar o efeito com maior confiança causal.

Saúde pública

A ausência de efeito detectável sobre a morbidade hospitalar por doenças relacionadas ao saneamento é o resultado que mais se distingue dos achados de estudos anteriores. Galiani, Gertler e Schargrodsky (2005) documentaram redução de 8% na mortalidade infantil nos municípios argentinos privatizados, com efeitos que chegaram a 26% nas áreas mais pobres, após horizontes de observação de até cinco anos. Saiani e Azevedo (2018), com um painel de treze anos sobre as privatizações municipais brasileiras de 1995 a 2008, encontraram que a concessão plena a operadores privados locais se associa a reduções significativas de morbidade por doenças relacionadas ao saneamento. Chaves e Costa (2025), ao atualizar essa análise para o período 1998–2021, obtiveram resultados mistos, mas ainda com evidência de melhora em parcela dos municípios tratados.

A principal razão para a divergência em relação a esses estudos é temporal: as concessões analisadas no presente trabalho são recentes, com a coorte mais antiga tendo no máximo quatro anos de operação, e os anos de 2020 a 2022 foram excluídos em função da pandemia de COVID-19, o que reduz substancialmente os períodos pós-tratamento disponíveis para as coortes de 2020 e

2021. Conforme argumentado por Dassan *et al.* (2024), os efeitos sobre variáveis de resultado de mais longo prazo, como saúde pública, tendem a emergir apenas após um ciclo completo de investimentos, estimado em pelo menos dois a quatro anos de operação. Adicionalmente, os próprios indicadores operacionais discutidos anteriormente (ausência de redução de perdas, aumento no tempo de reparo de vazamentos e possíveis artefatos de reporte nos indicadores de cobertura) sugerem que a melhoria efetiva da qualidade do serviço ainda não se consolidou a ponto de gerar externalidades de saúde detectáveis. Nesse sentido, a ausência de efeito é esperada e não invalida a hipótese de que a concessão plena possa produzir benefícios de saúde em horizontes mais longos, uma vez que as melhorias operacionais se materializem.

Implicações para a política regulatória

Os resultados oferecem lições relevantes para a regulação do setor no contexto do NMLSB, e seu conjunto pode ser lido à luz de uma tese central: a privatização, por si só, não determina o resultado. O que determina é a qualidade institucional do ambiente em que ela ocorre. Essa é, em essência, a tese de Lazzarini (2023, p. 108): a privatização certa exige governos capazes: capazes de elaborar contratos que especifiquem os atributos de serviço exigidos, de monitorar o cumprimento de metas, de aplicar penalidades em caso de desvio e de defender o interesse público frente ao poder de barganha dos concessionários.

A reestruturação de custos observada é real e rápida, mas sua translação em benefícios para os usuários, seja em qualidade do serviço, seja em universalização, é mais lenta e depende de incentivos regulatórios adequados. O aumento da evasão de receitas e a deterioração do desempenho financeiro operacional sinalizam que a sustentabilidade econômico-financeira das concessões não é automática e requer monitoramento ativo. A coexistência de ganhos de eficiência com piora em indicadores de qualidade operacional é consistente com a preocupação teórica de Hart, Shleifer e Vishny (1997) sobre o dilema custo-qualidade, e reforça a necessidade de que os contratos de concessão incorporem metas mensuráveis e progressivas de qualidade, com mecanismos efetivos de incentivo e penalidade. Lazzarini (2023, p. 172) argumenta que, com boas métricas, a remuneração por resultados deveria ser mais utilizada nos contratos entre atores públicos e privados, recomendação essa que corrobora a Resolução Normativa ANA n.º 183/2024, que prevê indicadores de desempenho atrelados ao ciclo tarifário (ANA, 2024).

O problema do refém também é relevante nesse contexto. Uma vez que um operador privado assume a concessão e realiza investimentos específicos em infraestrutura, o poder concedente pode ter dificuldade de impor renegociações ou de encerrar o contrato sem custos elevados enquanto que o operador pode tentar melhorar sua posição de barganha ao longo do tempo (Lazzarini, 2023, p. 100). Esse risco é tanto maior quanto mais ambíguas forem as cláusulas contratuais originais, o que reforça o argumento em favor de contratos detalhados e reguladores tecnicamente capacitados desde o início da concessão.

Por fim, a heterogeneidade dos efeitos entre coortes reforça que o desempenho das

concessões não é determinado apenas pela mudança de titularidade do operador, mas pelo conjunto de condições institucionais, contratuais e de mercado sob as quais cada concessão é celebrada. Como previsões simples podem não se concretizar quando ocorrem mudanças de contexto (Lazzarini, 2023, p. 25), avaliar essa heterogeneidade de forma sistemática, à medida que novos dados forem disponibilizados, constitui a agenda de pesquisa mais relevante para a compreensão dos efeitos de longo prazo do NMLSB sobre o saneamento básico no Brasil.

4.6 Conclusões

Este estudo investigou os efeitos causais das concessões plenas de abastecimento de água celebradas a partir do Novo Marco Legal do Saneamento Básico sobre um conjunto amplo de indicadores operacionais, financeiros e de saúde pública, utilizando o estimador de diferenças-em-diferenças escalonado de Callaway e Sant'Anna (2021) com pareamento por escore de propensão em um painel de municípios brasileiros para o período 2017–2024.

Os achados revelam um padrão coerente de reestruturação operacional de curto prazo: os municípios concedidos experimentaram reduções expressivas na incidência de despesas com pessoal próprio (–13,81 p.p.), na despesa de exploração por m³ (–R\$ 1,80/m³) e na incidência de energia elétrica (–4,25 p.p.), além de expansão significativa da macromedicação (+26,60 p.p.) e aumento reportado na cobertura de abastecimento (+4,78 p.p.). Esses resultados são consistentes com a expectativa teórica de que a privatização produz ganhos gerenciais mensuráveis no curto prazo (Saiani; Azevedo, 2018; Dassan *et al.*, 2024).

Contudo, esse quadro positivo é contrabalançado por evidências de deterioração em dimensões relevantes de qualidade: o tempo médio de reparo de vazamentos aumentou 22,18 horas/reparo, as perdas totais na distribuição cresceram 3,54 p.p. e o desempenho financeiro operacional recuou 17,99 p.p. Esses resultados são interpretados como manifestação incipiente do dilema custo-qualidade (Hart; Shleifer; Vishny, 1997): a reestruturação de custos ainda não se traduziu em melhoria efetiva do serviço, e a piora em indicadores de manutenção é consistente com a redução do quadro próprio de pessoal ainda não compensada por uma rede equivalente de prestadores terceirizados. Nenhum efeito significativo foi detectado sobre a morbidade hospitalar por doenças relacionadas ao saneamento, resultado esperado dado o horizonte temporal ainda reduzido das concessões e o impacto da pandemia de COVID-19 sobre a disponibilidade de períodos pós-tratamento.

Do ponto de vista da política tarifária, os achados revelam que o efeito médio sobre a receita operacional por m³ é nulo, mas esconde dinâmicas opostas entre coortes: enquanto as concessões celebradas sob o critério de maior outorga, prevalente nas primeiras rodadas do NMLSB, estão associadas a aumentos tarifários, as coortes posteriores, sob critério de menor tarifa, registram reduções. Esse padrão corrobora a análise de Fagundes, Cunha Marques e Malheiros (2024), que demonstra que o modelo de maior outorga transfere integralmente para os usuários o custo de

prêmios bilionários, com impactos desproporcionais sobre famílias em situação de pobreza. A mudança de critério licitatório representa, portanto, um aprendizado institucional relevante com implicações diretas sobre a modicidade tarifária.

Implicações para a política pública

Os resultados não autorizam uma resposta simples à pergunta sobre se vale a pena privatizar os serviços de saneamento. Tal como argumenta Lazzarini (2023, p. 51), o exercício adequado não é defender de forma genérica a atuação de empresas públicas ou privadas, mas analisar os prós e contras de cada alternativa e propor maneiras de lidar com seus dilemas inerentes. A evidência aqui produzida sugere que as concessões plenas têm potencial real de gerar ganhos operacionais e de cobertura, mas que esses ganhos dependem criticamente de três condições institucionais.

A primeira é o desenho do processo licitatório. O critério de julgamento dos leilões determina estruturalmente os incentivos do concessionário: modelos baseados em maior outorga tendem a elevar tarifas e comprometer a acessibilidade, especialmente para populações vulneráveis, enquanto modelos baseados em menor tarifa alinham melhor os incentivos privados ao interesse público (Fagundes; Cunha Marques; Malheiros, 2024; ANA, 2024).

A segunda é a completude e monitorabilidade dos contratos. A ausência histórica de metas explícitas de qualidade nos contratos de concessão brasileiros (Saiani; Azevedo, 2018) e a dificuldade dos governos em especificar atributos de serviço não-contratáveis (Lazzarini, 2023, p. 108) criam espaço para que a redução de custos se dê com base na piora da qualidade. A Resolução Normativa ANA n.º 183/2024, ao estabelecer indicadores de desempenho vinculados ao ciclo tarifário, é um avanço institucional na direção correta (ANA, 2024), mas sua efetividade depende da capacidade técnica das entidades reguladoras infranacionais de monitorá-los e aplicá-los.

A terceira é o horizonte de avaliação. Os estudos que encontraram efeitos positivos da privatização sobre saúde pública no Brasil cobriram horizontes de até treze anos (Saiani; Azevedo, 2018), substancialmente superiores ao máximo de quatro anos disponível no presente estudo. A ausência de efeitos sobre morbidade não é evidência de que eles não existirão, mas sim de que ainda é cedo para observá-los. Isso impõe aos formuladores de política a necessidade de mecanismos de avaliação contínua das concessões em curso.

Contribuições e limitações

As contribuições deste estudo para a literatura são de duas ordens. Do ponto de vista metodológico, adota-se o estimador de Callaway e Sant'Anna (2021) combinado com pareamento por escore de propensão, abordagem que acomoda a heterogeneidade de efeitos entre coortes e supera as limitações bem documentadas do estimador de efeitos fixos convencional em contextos

de tratamento escalonado (Baker *et al.*, 2025). Do ponto de vista substantivo, o estudo é o primeiro a avaliar os efeitos do NMLSB com dados até 2024, incorporando os indicadores consolidados do SINISA em nível municipal e permitindo capturar os efeitos da regionalização induzida pela reforma, em linha com a recomendação metodológica de Oliveira e Saiani (2022) e Chaves e Costa (2025).

As principais limitações dizem respeito à natureza dos dados e ao horizonte temporal. Os indicadores do SINISA são autodeclarados pelo próprio prestador, o que cria incentivos para viés de declaração, especialmente em variáveis de cobertura sujeitas a metas contratuais. Isso implica que parte dos ganhos reportados em IAG0001 pode refletir revisão cadastral e não expansão física de rede, limitação que somente pode ser superada com triangulação por fontes independentes, como pesquisas domiciliares. O horizonte de análise, condicionado pela disponibilidade de dados e pela exclusão dos anos pandêmicos, restringe a capacidade de detectar efeitos que se manifestam com maior defasagem, como os impactos sobre saúde e os resultados de investimentos em infraestrutura de maior maturação.

Agenda de pesquisa

A incorporação de dados dos anos subsequentes, à medida que as coortes mais recentes acumulem períodos pós-tratamento, é a extensão mais imediata e necessária deste trabalho. Em paralelo, a validação dos ganhos de cobertura por fontes censitárias ou de pesquisa domiciliar é condição para distinguir expansão real de declaração enviesada. Do ponto de vista metodológico, a comparação sistemática entre os resultados do estimador de Callaway e Sant’Anna (2021) e alternativas como o controle sintético e o estimador de Callaway e Sant’Anna (2021) com variáveis instrumentais permitiria avaliar a sensibilidade dos achados a diferentes hipóteses de identificação. Por fim, análises desagregadas por porte municipal, região geográfica e modelo contratual específico (critério de julgamento, estrutura de metas, entidade reguladora) seriam valiosas para identificar as condições sob as quais as concessões plenas produzem resultados mais favoráveis, contribuindo para o desenho mais eficaz das rodadas futuras de licitação no âmbito do NMLSB.

Considerações Finais

Esta tese investigou os efeitos da participação privada no saneamento brasileiro a partir de quatro perspectivas complementares: uma conceitual, uma bibliométrica e duas empíricas. Reunidos, os resultados oferecem uma resposta articulada à hipótese central de que a propriedade, por si só, não determina a eficiência dos prestadores nem garante a modicidade tarifária, o que prevalece é o ambiente regulatório e o desenho dos contratos de concessão.

A resposta à pergunta sobre se vale a pena privatizar os serviços de saneamento não é simples. Como observa Lazzarini (2023), o exercício adequado não é defender genericamente a atuação de empresas públicas ou privadas, mas analisar os prós e contras de cada alternativa e propor maneiras de lidar com seus dilemas inerentes. A evidência aqui produzida sugere que as concessões plenas têm potencial real de gerar ganhos operacionais e de cobertura, mas que esses ganhos dependem criticamente das condições institucionais em que se inserem: o desenho do leilão, a completude dos contratos e a capacidade regulatória. A universalização do saneamento básico no Brasil é possível, mas requer um modelo que combine eficiência econômica, sustentabilidade financeira e equidade social bem como uma regulação à altura desse desafio.

Síntese dos artigos

O primeiro artigo caracterizou o mercado de água e saneamento como um monopólio natural marcado por elevados custos fixos, economias de escala, longa maturação do investimento e demanda inelástica. A revisão das experiências internacionais (Reino Unido, Chile, Argentina, Bolívia, França, Alemanha e Estados Unidos) evidenciou que os resultados da participação privada são heterogêneos: há casos de melhora em eficiência e cobertura, mas também de aumento de tarifas, deterioração da qualidade e remunicipalização. O denominador comum das experiências bem-sucedidas é uma regulação robusta e transparente, capaz de alinhar os incentivos privados ao interesse público. No Brasil, o NMLSB representa um avanço institucional ao promover a regionalização e ao estabelecer uma autoridade regulatória nacional, mas sua implementação ainda carece de mecanismos de salvaguarda de acessibilidade, especialmente para populações de baixa renda.

O segundo artigo mapeou a produção científica sobre participação privada no saneamento a partir de uma revisão bibliométrica de 392 documentos publicados entre 1972 e 2025. A taxa de crescimento anual de 6,34% do campo reflete seu dinamismo, com aceleração especialmente após 1997, o que coincide com as grandes ondas de privatização global, e novo impulso após 2020, propiciado pelo debate em torno do NMLSB. A análise de redes revelou dois clusters intelectuais bem definidos: um central, ancorado em teoria da regulação e estudos de impacto

de privatização, e outro especializado em análise de eficiência por métodos de fronteira. A emergência do Brasil como tópico de tendência na produção recente reforça a relevância das contribuições empíricas desta tese para a literatura internacional. Como agenda de pesquisa prioritária para o país, o estudo identificou sete eixos: (i) estudos econométricos de impacto com desenhos causais robustos sobre os efeitos do NMLSB; (ii) análises de eficiência comparativa entre prestadores públicos e privados, incluindo o papel do capital humano nas PPPs; (iii) investigação sobre escala e regionalização e o trade-off custo-qualidade nos prestadores locais; (iv) regulação e captura regulatória; (v) equidade tarifária e impactos distributivos das reformas sobre famílias de baixa renda; (vi) comportamento do consumidor em relação à titularidade do serviço; e (vii) determinantes e design da privatização. Esta tese responde diretamente ao primeiro e ao segundo desses eixos.

O terceiro artigo estimou a eficiência técnica de custo de prestadores brasileiros de água e esgoto no período 2015–2022 por meio de uma função de custo variável translogarítmica estimada por SFA com ineficiência tempo-variante. Os resultados revelam um setor com eficiência média elevada, mas com dispersão expressiva entre prestadores e margem de ineficiência residual significativa, que representa um potencial considerável de reduções de custos operacionais e constitui o principal argumento econômico para políticas de benchmarking regulatório. O achado mais relevante para o debate sobre privatização é a ausência de efeito estatisticamente significativo da propriedade privada sobre os custos, resultado robusto a especificações alternativas. Esse resultado consolida a conclusão de que, no período analisado, o que determina a eficiência não é a natureza jurídica do prestador, mas o ambiente institucional em que opera, ou seja, extensão da rede, condições de micromedição, densidade de atendimento e estrutura de gestão.

O quarto artigo avaliou os efeitos causais das concessões plenas de abastecimento de água celebradas a partir do NMLSB sobre um conjunto amplo de resultados municipais, usando o estimador de diferenças-em-diferenças escalonado com pareamento por escore de propensão em um painel de municípios para o período 2017–2024. Os resultados apontam uma reestruturação operacional de curto prazo inequívoca: reduções expressivas nas despesas com pessoal e energia, expansão da macromedição e aumento reportado na cobertura de abastecimento. Esse quadro de ganhos gerenciais iniciais coexiste, porém, com piora em indicadores de qualidade operacional como maior tempo de reparo de vazamentos, aumento de perdas na distribuição e deterioração do desempenho financeiro, o que é interpretado como manifestação incipiente do dilema custo-qualidade. Os efeitos tarifários são heterogêneos entre coortes e dependem criticamente do modelo de leilão: concessões sob critério de maior outorga estão associadas a aumentos, enquanto as realizadas sob critério de menor tarifa registram reduções. Nenhum efeito significativo foi detectado sobre a morbidade hospitalar, resultado esperado dado o horizonte temporal ainda reduzido das concessões e o impacto da pandemia sobre os períodos pós-tratamento disponíveis.

Implicações para a política regulatória

De forma conjunta, os resultados desta tese têm três implicações diretas para os reguladores e formuladores de política. A primeira diz respeito ao desenho dos leilões. O critério de julgamento determina estruturalmente os incentivos do concessionário. Modelos baseados em maior outorga tendem a elevar tarifas e comprometer a acessibilidade, especialmente para populações vulneráveis. Modelos baseados em menor tarifa alinham melhor os incentivos privados ao interesse público. A mudança de critério observada nas rodadas mais recentes do NMLSB representa um aprendizado institucional relevante que deve ser preservado e aprofundado.

A segunda refere-se à qualidade e monitorabilidade dos contratos. A ausência histórica de metas explícitas de qualidade nos contratos de concessão brasileiros e a dificuldade de especificar atributos de serviço não-contratáveis criam espaço para que a redução de custos se dê à custa da qualidade tal como o dilema custo-qualidade previsto por Hart, Shleifer e Vishny (1997). A Resolução Normativa ANA n.º 183/2024, ao estabelecer indicadores de desempenho vinculados ao ciclo tarifário, é um avanço institucional na direção correta, mas sua efetividade depende da capacidade técnica das entidades reguladoras infranacionais de monitorá-los e aplicá-los.

A terceira trata do foco regulatório. Os resultados sugerem que as atenções no período pós-NMLSB devem se voltar não para a forma de propriedade em si, mas para os determinantes estruturais da eficiência identificados como relevantes: redução de perdas, expansão da micromedição, densificação da rede e fortalecimento da capacidade de gestão dos prestadores locais. Instrumentos que promovam competição por comparação entre prestadores, como os mecanismos de *benchmarking* previstos no NMLSB, são mais promissores para elevar a eficiência do que mudanças de propriedade desacompanhadas de reformas contratuais e regulatórias.

Limitações e agenda de pesquisa

Esta tese apresenta limitações que devem orientar estudos futuros. O horizonte temporal do quarto artigo restringe a capacidade de detectar efeitos que se manifestam com maior defasagem, como os impactos sobre saúde pública e os resultados de investimentos de maior maturação. Os indicadores do SINISA, por serem autodeclarados pelos prestadores, criam incentivos para viés de declaração, especialmente em variáveis de cobertura sujeitas a metas contratuais. A fronteira estocástica cobre o período pré-NMLSB, de modo que a avaliação dos efeitos da reforma sobre a eficiência demanda a incorporação dos dados do período subsequente, à medida que se tornem disponíveis. A revisão bibliométrica, por sua vez, está circunscrita a duas bases e a idiomas anglófonos, podendo subrepresentar contribuições interdisciplinares.

Como extensões imediatas desta pesquisa, destacam-se: (i) a incorporação de dados dos anos subsequentes para ampliar o horizonte de avaliação das concessões; (ii) a validação dos ganhos de cobertura reportados por fontes censitárias ou de pesquisa domiciliar independente;

(iii) a atualização da fronteira estocástica com dados pós-NMLSB para mensurar os efeitos da reforma sobre a eficiência; e (iv) análises desagregadas por porte municipal, região geográfica e modelo contratual específico, capazes de identificar as condições sob as quais as concessões plenas produzem resultados mais favoráveis.

Referências

- ABBOTT, M.; COHEN, B. Productivity and efficiency in the water industry. **Utilities Policy**, Elsevier BV, v. 17, p. 233–244, 2009. ISSN 09571787. DOI 10.1016/j.jup.2009.05.001. Citado na p. 75.
- ABCON SINDCON. **Panorama da Participação Privada no Saneamento 2024: Marco Legal – 1500 Dias de Avanços e Desafios**. 2024. Disponível em PDF. Disponível em: https://abconsindcon.com.br/wp-content/uploads/2024/08/Panorama_2024_MIOLO_PT_v4.1.pdf#page=6.21. Citado nas pp. 41 e 113.
- ALLEN, J.; PRYKE, M. Financialising household water: Thames Water, MEIF, and ‘ring-fenced’ politics. **Cambridge Journal of Regions, Economy and Society**, v. 6, n. 3, p. 419–439, 05 2013. ISSN 1752-1378. DOI 10.1093/cjres/rst010. Disponível em: <https://doi.org/10.1093/cjres/rst010>. Citado na p. 24.
- ANA. **Resolução ANA n.º 183, de 5 de fevereiro de 2024**: Aprova a Norma de Referência da Agência Nacional de Águas e Saneamento Básico n.º 6/2024, que dispõe sobre os modelos de regulação tarifária dos serviços públicos de abastecimento de água e esgotamento sanitário. Brasília, 2024. Diário Oficial da União, ed. 27, seq. 1, p. 35, 7 fev. 2024. Disponível em: <https://www.gov.br/ana/pt-br/legislacao/resolucoes/resolucoes-regulatorias/2024/183>. Acesso em: 5 abr. 2026. Citado nas pp. 137, 139 e 141.
- ANDRES, L. A.; THIBERT, M.; CORDOBA, C. L.; DANILENKO, A. V.; JOSEPH, G.; BORJA-VEGA, C. **Doing More with Less: Smarter Subsidies for Water Supply and Sanitation**. Washington, DC, 2019. Disponível em: <https://documents1.worldbank.org/curated/en/330841560517317845/Doing-More-with-Less-Smarter-Subsidies-for-Water-Supply-and-Sanitation.pdf>. Citado nas pp. 28 e 29.
- ANWANDTER, L.; OZUNA, T. Can public sector reforms improve the efficiency of public water utilities? **Environment and Development Economics**, v. 7, p. 687–700, 10 2002. ISSN 1355770X. DOI 10.1017/S1355770X02000414. Citado na p. 77.
- AQUINO, M.; LEDO, B. Sede de quê? Efeitos da interferência política no setor de saneamento sobre o bem-estar e a expansão do sistema. **Revista de Defesa da Concorrência**, v. 11, n. 2, p. 124–142, dec 2023. ISSN 2318-2253. DOI 10.52896/rdc.v11i2.1066. Disponível em: <https://revista.cade.gov.br/index.php/revistadedefesadaconcorrencia/article/view/1066>. Citado nas pp. 67, 69 e 70.
- ARIA, M.; CUCCURULLO, C. bibliometrix: An r-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959–975, 2017. ISSN 1751-1577. DOI

<https://doi.org/10.1016/j.joi.2017.08.007>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157717300500>. Citado nas pp. 49 e 51.

AUBERT, C.; REYNAUD, A. The impact of regulation on cost efficiency: An empirical analysis of wisconsin water utilities. **Journal of Productivity Analysis**, v. 23, p. 383–409, 7 2005. ISSN 0895-562X. DOI 10.1007/s11123-005-2216-8. Disponível em: <http://link.springer.com/10.1007/s11123-005-2216-8>. Citado nas pp. 77 e 78.

AVERCH, H.; JOHNSON, L. L. Behavior of the Firm Under Regulatory Constraint. **American Economic Review**, v. 52, n. 5, p. 1052–1069, 1962. ISSN 0036-8075. Citado na p. 69.

BAGNOLI, L.; BERTOMEU-SANCHEZ, S.; ESTACHE, A.; VAGLIASINDI, M. Does the ownership of utilities matter for social outcomes? A survey of the evidence for developing countries. **Journal of Economic Policy Reform**, Routledge, v. 26, n. 1, p. 24–43, 2023. ISSN 17487889. DOI 10.1080/17487870.2021.1997747. Disponível em: <https://doi.org/10.1080/17487870.2021.1997747>. Citado nas pp. 66 e 69.

BAKER, A.; CALLAWAY, B.; CUNNINGHAM, S.; GOODMAN-BACON, A.; SANT'ANNA, P. H. C. Difference-in-Differences Designs: A Practitioner's Guide. n. 2020, p. 1–75, 2025. Disponível em: <http://arxiv.org/abs/2503.13323>. Citado nas pp. 114, 116 e 142.

BAKKER, K. Privatizing water, producing scarcity: The Yorkshire drought of 1995. **Economic Geography**, v. 76, n. 1, p. 4–27, 2000. ISSN 00130095 (ISSN). Publisher: Clark University. DOI 10.1111/j.1944-8287.2000.tb00131.x. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034053202&doi=10.1111%2fj.1944-8287.2000.tb00131.x&partnerID=40&md5=8825c237fe07f29f349a4c54e4152b92>. Citado na p. 31.

BAKKER, K. Paying for water: Water pricing and equity in England and Wales. **Transactions of the Institute of British Geographers**, v. 26, n. 2, p. 143–164, 2001. ISSN 00202754. DOI 10.1111/1475-5661.00012. Citado nas pp. 30 e 31.

BARBOSA, A.; LIMA, S. C. de; BRUSCA, I. Governance and efficiency in the brazilian water utilities: A dynamic analysis in the process of universal access. **Utilities Policy**, Elsevier Ltd, v. 43, p. 82–96, 12 2016. ISSN 09571787. DOI 10.1016/j.jup.2016.06.013. Citado na p. 79.

BARRAQUÉ, B. Return of drinking water supply in Paris to public control. **Water Policy**, v. 14, n. 6, p. 903–914, 2012. ISSN 13667017. DOI 10.2166/wp.2012.085. Citado nas pp. 32 e 33.

BARRERA-OSORIO, F.; OLIVERA, M.; OSPINO, C. Does society win or lose as a result of privatization? The case of water sector privatization in Colombia. **Economica**, v. 76, n. 304, p. 649–674, 2009. ISSN 00130427. DOI 10.1111/j.1468-0335.2008.00731.x. Citado nas pp. 110 e 111.

- BASU, M.; DASGUPTA, R. Where do we stand now? A bibliometric analysis of water research in support of the sustainable development goal 6. **Water (Switzerland)**, v. 13, n. 24, p. 1–18, 2021. ISSN 20734441. DOI 10.3390/w13243591. Citado na p. 18.
- BATTESE, G.; COELLI, T. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in india. **Journal of Productivity Analysis**, Springer, v. 3, n. 1/2, p. 153–169, 1992. ISSN 0895562X, 15730441. Disponível em: <http://www.jstor.org/stable/41770578>. Citado nas pp. 74, 81, 85, 95, 100, 101 e 104.
- BAUER, C. J. Bringing Water Markets Down to Earth: The Political Economy of Water Rights in Chile, 1976-95. **World Development**, v. 25, n. 5, p. 639–656, 1997. ISSN 0305750X. DOI 10.1016/S0305-750X(96)00128-3. Citado na p. 31.
- BAUMOL, W. J. On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry. **The American Economic Review**, v. 67, n. 5, p. 809–822, 1977. Disponível em: <http://www.jstor.org/stable/1828065>. Citado na p. 24.
- BEL, G. Public versus private water delivery, remunicipalization and water tariffs. **Utilities Policy**, v. 62, n. November 2019, 2020. ISSN 09571787. DOI 10.1016/j.jup.2019.100982. Citado na p. 117.
- BEL, G.; WARNER, M. Does privatization of solid waste and water services reduce costs? A review of empirical studies. 2008. 1337-1348 p. DOI 10.1016/j.resconrec.2008.07.014. Citado nas pp. 73 e 75.
- BEVERIDGE, R.; HÜESKER, F.; NAUMANN, M. From post-politics to a politics of possibility? Unravelling the privatization of the Berlin Water Company. **Geoforum**, Elsevier Ltd, v. 51, n. September 2011, p. 66–74, 2014. ISSN 00167185. DOI 10.1016/j.geoforum.2013.09.021. Disponível em: <http://dx.doi.org/10.1016/j.geoforum.2013.09.021>. Citado nas pp. 33 e 34.
- BHATTACHARYYA, A.; HARRIS, T. R.; NARAYANAN, R.; RAFFIEE, K. Specification and estimation of the effect of ownership on the economic efficiency of the water utilities. **Regional Science and Urban Economics**, v. 25, p. 759–784, 12 1995. ISSN 01660462. DOI 10.1016/0166-0462(95)02107-8. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/0166046295021078>. Citado nas pp. 75 e 78.
- BOYCKO, M.; SHLEIFER, A.; VISHNY, R. W. A Theory of Privatisation. **The Economic Journal**, v. 106, n. 435, p. 309, mar 1996. ISSN 00130133. DOI 10.2307/2235248. Disponível em: <https://academic.oup.com/ej/article/106/435/309-319/5159022>. Citado na p. 69.
- BRASIL. Novo Marco Legal do Saneamento Básico. Lei Federal nº 14.026, de 15 de julho de 2020. Brasília: [s.n.], 2020. 31 p. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/114026.htm. Citado nas pp. 39, 40, 41, 48, 70 e 107.

- BRASIL. **Diagnóstico Temático: Serviços de Água e Esgoto. Visão Geral ano de referência 2022**. Brasília, 2023. –107 p. Disponível em: https://www.gov.br/cidades/pt-br/aceso-a-informacao/acoes-e-programas/saneamento/snis/produtos-do-snis/diagnosticos/DIAGNOSTICO_TEMATICO_VISAO_GERAL_AE_SNIS_2023.pdf. Citado nas pp. 18 e 29.
- BRASIL. Ministério das Cidades. **Glossário de Indicadores - Água e Esgotos 2022**. [S.l.], 2023. Disponível em: https://www.gov.br/cidades/pt-br/aceso-a-informacao/acoes-e-programas/saneamento/snis/produtos-do-snis/diagnosticos/Glossario_Indicadores_AE2022.pdf. Citado na p. 113.
- BRASIL. Ministério das Cidades. **Portaria MCID nº 788, de 1º de agosto de 2024**. 2024. Diário Oficial da União. Estabelece os procedimentos gerais para o cumprimento do disposto no inciso IV do caput do art. 50 da Lei nº 11.445, de 5 de janeiro de 2007, e no inciso IV do caput do art. 7º do Decreto nº 11.599, de 12 de julho de 2023. Disponível em: <https://www.in.gov.br/web/dou/-/portaria-mcid-n-788-de-1-de-agosto-de-2024-578754081>. Acesso em: 15 fev. 2025. Citado na p. 39.
- BRASIL. Ministério das Cidades. **Relatório SINISA Abastecimento de Água 2024**. 2024. Disponível em: https://www.gov.br/cidades/pt-br/aceso-a-informacao/acoes-e-programas/saneamento/sinisa/resultados-sinisa/copy_of_RELATORIO_SINISA_ABASTECIMENTO_DE_AGUA_2024_v2.pdf. Citado nas pp. 42 e 43.
- Brasil. Ministério das Cidades. **Relatório dos Serviços de Abastecimento de Água – SINISA 2025**. Brasília, 2025. Ano de referência 2024. Disponível em: <https://www.gov.br/cidades/sinisa>. Citado nas pp. 43 e 73.
- Brasil. Ministério das Cidades. **Relatório dos Serviços de Esgotamento Sanitário – SINISA 2025**. Brasília, 2025. Ano de referência 2024. Disponível em: <https://www.gov.br/cidades/sinisa>. Citado nas pp. 43 e 73.
- BRUNO, E. M.; JESSOE, K. Using Price Elasticities of Water Demand to Inform Policy. **Annual Review of Resource Economics**, v. 13, n. 1, p. 427–441, oct 2021. ISSN 1941-1340. DOI 10.1146/annurev-resource-110220-104549. Disponível em: <https://www.annualreviews.org/doi/10.1146/annurev-resource-110220-104549>. Citado na p. 27.
- BUNTAINE, M.; GREENSTONE, M.; HE, G.; LIU, M.; WANG, S.; ZHANG, B. Does the Squeaky Wheel Get More Grease? The Direct and Indirect Effects of Citizen Participation on Environmental Governance in Chinat. **AMERICAN ECONOMIC REVIEW**, v. 114, n. 3, p. 815–850, mar. 2024. ISSN 0002-8282. DOI 10.1257/aer.20221215. Citado na p. 64.
- BYATT, I. The regulation of water services in the UK. **Utilities Policy**, Elsevier Ltd, v. 24, p. 3–10, 2013. ISSN 09571787. DOI 10.1016/j.jup.2012.07.003. Disponível em: <http://dx.doi.org/10.1016/j.jup.2012.07.003>. Citado na p. 31.

- BYRNES, P.; GROSSKOPF, S.; HAYES, K. Efficiency and Ownership: Further Evidence. **The Review of Economics and Statistics**, v. 68, n. 2, p. 337, may 1986. ISSN 00346535. DOI 10.2307/1925517. Disponível em: <http://www.jstor.org/stable/1925517>. Accessed: <https://www.jstor.org/stable/1925517?origin=crossref>. Citado nas pp. 53, 69, 75 e 78.
- CAI, H.; CHEN, Y.; GONG, Q. Polluting thy neighbor: Unintended consequences of China's pollution reduction mandates. **JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT**, v. 76, p. 86–104, mar. 2016. ISSN 0095-0696. DOI 10.1016/j.jeem.2015.01.002. Citado na p. 53.
- CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. **Journal of Economic Surveys**, v. 22, n. 1, p. 31–72, 2008. ISSN 09500804. DOI 10.1111/j.1467-6419.2007.00527.x. Citado nas pp. 116 e 117.
- CALLAWAY, B.; SANT'ANNA, P. H. Difference-in-Differences with multiple time periods. **Journal of Econometrics**, Elsevier B.V., v. 225, n. 2, p. 200–230, 2021. ISSN 18726895. DOI 10.1016/j.jeconom.2020.12.001. Disponível em: <https://doi.org/10.1016/j.jeconom.2020.12.001>. Citado nas pp. 20, 109, 112, 114, 115, 116, 117, 120, 140, 141 e 142.
- CARDOSO, J. de M.; ZOGHBI, A. C. P.; RESENDE, J. G. de L. Privatization and sanitation outcomes: evidence from brazilian municipalities. **Applied Economics Letters**, Routledge, v. 0, n. 0, p. 1–7, 2026. DOI 10.1080/13504851.2026.2622550. Disponível em: <https://doi.org/10.1080/13504851.2026.2622550>. Citado nas pp. 19 e 21.
- CARVALHO, A. A. de. Regulação econômica e contratos sob o novo marco legal do saneamento básico : estudo de caso da concessão da região metropolitana de Maceió. In: Mauro Santos Silva (Ed.). **Concessões e parcerias público-privadas : políticas públicas para provisão de infraestrutura**. Brasília: Ipea, 2022. cap. 15, p. 468. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/11495>. Citado na p. 112.
- CARVALHO, P.; MARQUES, R. C. Estimating size and scope economies in the portuguese water sector using the bayesian stochastic frontier analysis. **Science of the Total Environment**, Elsevier B.V., v. 544, p. 574–586, 2 2016. ISSN 18791026. DOI 10.1016/j.scitotenv.2015.11.169. Citado nas pp. 76 e 78.
- CASARIN, A. A.; DELFINO, J. A.; DELFINO, M. E. Failures in water reform: Lessons from the Buenos Aires's concession. **Utilities Policy**, v. 15, n. 4, p. 234–247, 2007. ISSN 09571787. DOI 10.1016/j.jup.2007.02.004. Citado nas pp. 34 e 35.
- CASHMAN, A.; LEWIS, L. Topping up or watering down? Sustainable development in the privatized UK water industry. **Business Strategy and the Environment**, v. 16, n. 2, p. 93–105, 2007. ISSN 09644733. DOI 10.1002/bse.472. Citado na p. 30.
- CETRULO, T. B.; MARQUES, R. C.; MALHEIROS, T. F. An analytical review of the efficiency of water and sanitation utilities in developing countries. **Water Research**, Elsevier Ltd,

- v. 161, p. 372–380, 2019. ISSN 18792448. DOI 10.1016/j.watres.2019.05.044. Disponível em: <https://doi.org/10.1016/j.watres.2019.05.044>. Citado nas pp. 107, 112 e 113.
- CHAVES, H. P.; FILHO, T. de C. T. B.; MARQUES, R. C.; MALHEIROS, T. F. **Performance evaluation of water supply service providers: a review with a focus on Brazil**. [S.l.]: IWA Publishing, 2024. 3118-3131 p. DOI 10.2166/ws.2024.198. Citado na p. 80.
- CHAVES, R. F.; COSTA, A. B. Private ownership of water and wastewater systems: Assessing health impacts. **Utilities Policy**, Elsevier Ltd, v. 95, n. August 2024, p. 101903, 2025. ISSN 09571787. DOI 10.1016/j.jup.2025.101903. Disponível em: <https://doi.org/10.1016/j.jup.2025.101903>. Citado nas pp. 68, 69, 111, 112, 113, 114, 138 e 142.
- CHISARI, O.; ESTACHE, A.; ROMERO, C. Winners and losers from the privatization and regulation of utilities: Lessons from a general equilibrium model of Argentina. **World Bank Economic Review**, v. 13, n. 2, p. 357–378, 1999. ISSN 02586770. DOI 10.1093/wber/13.2.357. Citado na p. 35.
- CHONG, E.; SAUSSIÉ, S.; SILVERMAN, B. S. Water under the bridge: Determinants of franchise renewal in water provision. **Journal of Law, Economics, and Organization**, v. 31, p. i3–i39, 2015. ISSN 14657341. DOI 10.1093/jleo/ewv010. Citado nas pp. 32 e 33.
- CHRISTENSEN, L. R.; GREENE, W. H. Economies of scale in u.s. electric power generation. **Journal of Political Economy**, v. 84, p. 655–676, 8 1976. ISSN 0022-3808. DOI 10.1086/260470. Disponível em: <https://www.journals.uchicago.edu/doi/10.1086/260470>. Citado nas pp. 82, 83, 84 e 85.
- CHRISTENSEN, L. R.; JORGENSEN, D. W.; LAU, L. J. Transcendental logarithmic production frontiers. **The Review of Economics and Statistics**, v. 55, p. 28, 2 1973. ISSN 00346535. DOI 10.2307/1927992. Disponível em: <https://www.jstor.org/stable/1927992?origin=crossref>. Citado nas pp. 82, 83 e 84.
- CIRIACY-WANTRUP, S. V. Projections of Water Requirements in the Economics of Water Policy. **Journal of Farm Economics**, v. 43, n. 2, p. 197, may 1961. ISSN 10711031. DOI 10.2307/1235779. Disponível em: <https://academic.oup.com/ajae/article-lookup/doi/10.2307/1235779>. Citado na p. 27.
- CLARKE, G. R.; KOSEC, K.; WALLSTEN, S. Has private participation in water and sewerage improved coverage? Empirical evidence from Latin America. **Journal of International Development**, v. 21, n. 3, p. 327–361, apr 2009. ISSN 09541748. DOI 10.1002/jid.1458. Disponível em: <http://doi.wiley.com/10.1002/jid.1458>. Citado nas pp. 35, 36, 38, 107, 108, 109 e 110.
- COELLI, T.; HENNINGSSEN, A. **frontier: Stochastic Frontier Analysis**. [S.l.], 2020. R package version 1.1-8. Disponível em: <https://CRAN.R-Project.org/package=frontier>. Citado na p. 85.

- COELLI, T. J.; Prasada Rao, D. S.; O'DONNELL, C. J.; BATTESE, G. E. **An introduction to efficiency and productivity analysis**. [S.l.: s.n.], 2005. 1–349 p. ISBN 0387242651. DOI 10.1007/b136381. Citado nas pp. 85 e 86.
- CORREA-PARRA, J.; VERGARA-PERUCICH, J. F.; AGUIRRE-NUÑEZ, C. Water privatization and inequality: Gini coefficient for water resources in chile. **Water (Switzerland)**, v. 12, n. 12, p. 1–13, 2020. ISSN 20734441. DOI 10.3390/w12123369. Citado nas pp. 31 e 32.
- CRAIN, W. M.; ZARDKOOHI, A. A test of the property-rights theory of the firm: water utilities in the united states. **The Journal of Law and Economics**, The University of Chicago Law School, v. 21, n. 2, p. 395–408, 1978. Citado na p. 53.
- CRAIN, W. M.; ZARDKOOHI, A. A test of the property-rights theory of the firm: Water utilities in the united states. **The Journal of Law and Economics**, v. 21, p. 395–408, 10 1978. ISSN 0022-2186. DOI 10.1086/466927. Disponível em: <https://www.journals.uchicago.edu/doi/10.1086/466927>. Citado nas pp. 75 e 78.
- Da Silva E Souza, G.; De Faria, R. C.; MOREIRA, T. B. S. Estimating the relative efficiency of Brazilian publicly and privately owned water utilities: A stochastic cost frontier approach. **Journal of the American Water Resources Association**, v. 43, n. 5, p. 1237–1244, 2007. ISSN 1093474X. DOI 10.1111/j.1752-1688.2007.00106.x. Citado nas pp. 74, 79, 80, 81, 83, 84, 87, 88, 93, 101, 102 e 104.
- DANELON, A. F. **Eficiência econômica do tratamento de água e esgoto no Brasil: uma abordagem por fronteira estocástica de custos**. Tese (Doutorado em Ciências – Área de Concentração: Economia Aplicada) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2020. Orientador: Prof. Dr. Humberto Francisco Silva Spolador. 121 p. Citado nas pp. 80, 81, 100, 101, 102, 104 e 105.
- DASSAN, A.; SAMPAIO, J. O.; SILVA, V. A. B.; DE-LOSSO, R. Does Private Means Better? A Water and Sanitation Quasi-Experimental Design. **Emerging Markets Finance and Trade**, Routledge, v. 60, n. 6, p. 1087–1105, 2024. ISSN 15580938. DOI 10.1080/1540496X.2023.2267737. Disponível em: <https://doi.org/10.1080/1540496X.2023.2267737>. Citado nas pp. 19, 24, 28, 29, 38, 39, 66, 69, 70, 111, 112, 114, 135, 136, 137, 139 e 140.
- de Souza Pereira, M.; Magalhães Filho, F. J. C.; de Moraes Lima, P.; TABAK, B. M.; CONSTANTINO, M. Sanitation and water services: Who is the most efficient provider public or private? Evidences for Brazil. **Socio-Economic Planning Sciences**, v. 79, n. May 2021, p. 101149, feb 2022. ISSN 00380121. DOI 10.1016/j.seps.2021.101149. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0038012121001415>. Citado nas pp. 66, 69, 70, 71, 80, 81 e 101.

- DEVOTO, F.; DUFLO, E.; DUPAS, P.; PARIENTÉ, W.; PONS, V. Happiness on tap: Piped water adoption in urban Morocco. **American Economic Journal: Economic Policy**, v. 4, n. 4, p. 68–99, 2012. ISSN 19457731. DOI 10.1257/pol.4.4.68. Citado na p. 53.
- Di Tella, R.; GALIANI, S.; SCHARGRODSKY, E. Reality versus propaganda in the formation of beliefs about privatization. **Journal of Public Economics**, Elsevier B.V., v. 96, n. 5-6, p. 553–567, 2012. ISSN 00472727. DOI 10.1016/j.jpubeco.2011.11.006. Disponível em: <http://dx.doi.org/10.1016/j.jpubeco.2011.11.006>. Citado na p. 35.
- DINAR, A.; TSUR, Y. **The Economics of Water Resources**. Cambridge University Press, 2021. ISBN 9781316678640. DOI 10.1017/9781316678640. Disponível em: <https://www.cambridge.org/core/product/identifier/9781316678640/type/book>. Citado nas pp. 22 e 23.
- DISTEFANO, T. **Water Resources and Economic Processes**. 1 Edition. — New York: Routledge, 2020. — Series: Routledge studies in ecological economics: Routledge, 2020. 1–304 p. ISBN 9780429025013. DOI 10.4324/9780429025013. Disponível em: <https://www.taylorfrancis.com/books/9780429658372>. Citado na p. 18.
- DOMENECH, T.; BLEISCHWITZ, R.; DORANOVA, A.; PANAYOTOPOULOS, D.; ROMAN, L. Mapping Industrial Symbiosis Development in Europe_ typologies of networks, characteristics, performance and contribution to the Circular Economy. **Resources, Conservation and Recycling**, v. 141, p. 76–98, 2019. ISSN 09213449 (ISSN). Publisher: Elsevier B.V. DOI 10.1016/j.resconrec.2018.09.016. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055164495&doi=10.1016%2fj.resconrec.2018.09.016&partnerID=40&md5=21dd832aa021a4a6e9732495aced9d87>. Citado na p. 53.
- DONTHU, N.; KUMAR, S.; MUKHERJEE, D.; PANDEY, N.; LIM, W. M. How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, Elsevier Inc., v. 133, n. April, p. 285–296, 2021. ISSN 01482963. DOI 10.1016/j.jbusres.2021.04.070. Disponível em: <https://doi.org/10.1016/j.jbusres.2021.04.070>. Citado nas pp. 48, 49 e 50.
- DORE, M. H.; KUSHNER, J.; ZUMER, K. Privatization of water in the UK and France - What can we learn? **Utilities Policy**, v. 12, n. 1, p. 41–50, 2004. ISSN 09571787. DOI 10.1016/j.jup.2003.11.002. Citado nas pp. 30, 31, 33 e 112.
- DURÁN-SÁNCHEZ, A.; ÁLVAREZ-GARCÍA, J.; GONZÁLEZ-VÁZQUEZ, E.; Del Río-Rama, M. d. I. C. Wastewater Management: Bibliometric Analysis of Scientific Literature. **Water**, v. 12, n. 11, p. 2963, oct 2020. ISSN 2073-4441. DOI 10.3390/w12112963. Disponível em: <https://www.mdpi.com/2073-4441/12/11/2963>. Citado na p. 48.
- ERBETTA, F.; CAVE, M. Regulation and efficiency incentives: Evidence from the england and wales water and sewerage industry. **Review of Network Economics**, v. 6, 1 2007. ISSN 1446-9022. DOI 10.2202/1446-9022.1128. Disponível em: <https://www.degruyter.com/document/doi/10.2202/1446-9022.1128/html>. Citado nas pp. 77 e 78.

- ESTACHE, A.; KOUASSI, E. Sector organization, governance, and the inefficiency of african water utilities. **SSRN Electronic Journal**, 2002. Disponível em: <https://ssrn.com/abstract=636253>. Citado nas pp. 76 e 78.
- ESTACHE, A.; ROSSI, M. A. How different is the efficiency of public and private water companies in Asia? **World Bank Economic Review**, v. 16, n. 1, p. 139–148, 2002. ISSN 02586770. DOI 10.1093/wber/16.1.139. Citado nas pp. 35, 53, 58 e 69.
- ESTACHE, A.; ROSSI, M. A.; ALEXANDER, I.; ALVAREZ, A.; BOURGUIGNON, F.; BURNS, P.; CANAY, I.; COELLI, T.; CRAMPES, C.; DINGHEM, S.; TRUJILLO, L.; OLIVERI, A.; RODRIGUEZ-PARDINA, M. **How Different Is the Efficiency of Public and Private Water Companies in Asia?** [S.l.], 2002. Citado nas pp. 75 e 78.
- ESTACHE, A.; TRUJILLO, L. Efficiency effects of "privatization" in Argentina's water and sanitation services. **Water Policy**, v. 5, n. 4, p. 369–380, 2003. ISSN 13667017. DOI 10.2166/wp.2003.0022. Citado nas pp. 34, 35 e 77.
- FABRE, A.; STRAUB, S. The impact of public–private partnerships (ppps) in infrastructure, health, and education. **Journal of Economic Literature**, v. 61, n. 2, p. 655–715, June 2023. DOI 10.1257/jel.20211607. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jel.20211607>. Citado nas pp. 19, 48, 61, 62, 64, 67, 69, 70, 71 e 110.
- FAGUNDES, T. S.; Cunha Marques, R.; MALHEIROS, T. F. Are Brazilian water auction prices affordable for vulnerable families? **Utilities Policy**, Elsevier Ltd, v. 88, n. March, p. 101749, 2024. ISSN 09571787. DOI 10.1016/j.jup.2024.101749. Disponível em: <https://doi.org/10.1016/j.jup.2024.101749>. Citado nas pp. 40, 41, 42, 68, 69, 70, 109, 137, 140 e 141.
- FEIGENBAUM, S.; TEEPLES, R. Public versus private water delivery: a hedonic cost approach. **The review of Economics and Statistics**, JSTOR, p. 672–678, 1983. Citado na p. 53.
- FEIGENBAUM, S.; TEEPLES, R. Public versus private water delivery: A hedonic cost approach. **The Review of Economics and Statistics**, v. 65, p. 672, 11 1983. ISSN 00346535. DOI 10.2307/1935940. Disponível em: <https://www.jstor.org/stable/1935940?origin=crossref>. Citado nas pp. 75 e 78.
- FERRO, G.; LENTINI, E. J.; MERCADIER, A. C.; ROMERO, C. A. Efficiency in Brazil's water and sanitation sector and its relationship with regional provision, property and the independence of operators. **Utilities Policy**, Elsevier Ltd, v. 28, p. 42–51, 2014. ISSN 09571787. DOI 10.1016/j.jup.2013.12.001. Disponível em: <http://dx.doi.org/10.1016/j.jup.2013.12.001>. Citado nas pp. 73, 74, 79, 80, 81, 82, 83, 84, 87, 88, 93, 100, 101 e 104.
- FILIPPINI, M.; HROVATIN, N.; ZORIĆ, J. Cost efficiency of slovenian water distribution utilities: An application of stochastic frontier methods. **Journal of Productivity Analysis**, v. 29, p.

- 169–182, 4 2008. ISSN 0895562X. DOI 10.1007/s11123-007-0069-z. Citado nas pp. 77, 78, 83 e 84.
- FOX, W. F.; HOFER, R. A. Using homothetic composed error frontiers to measure water utility efficiency. **Southern Economic Journal**, v. 53, p. 461, 10 1986. ISSN 00384038. DOI 10.2307/1059427. Disponível em: <https://www.jstor.org/stable/1059427?origin=crossref>. Citado nas pp. 75 e 78.
- FREIRE, V. G. M. **Ineficiência das provedoras de água e esgoto no Brasil: efeitos do novo Marco Legal do Saneamento**. Dissertação (Mestrado) — Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2025. Dissertação de Mestrado – Área: Economia Aplicada. Orientador: Bruno César Aurichio Ledo. Citado nas pp. 80, 81, 82 e 103.
- GALIANI, S.; GERTLER, P.; SCHARGRODSKY, E. Water for life: The impact of the privatization of water services on child mortality. **JOURNAL OF POLITICAL ECONOMY**, v. 113, n. 1, p. 83–120, fev. 2005. ISSN 0022-3808. DOI 10.1086/426041. Citado nas pp. 24, 29, 34, 48, 53, 58, 59, 60, 68, 69, 70, 107, 111, 112, 114, 136 e 138.
- GARCÍA-LÓPEZ, M.; LÓPEZ-RUIZ, S.; CUADRADO-QUESADA, G. Does economic growth drive equitable water and sanitation access? assessing inequality reduction across 64 nations. **Sustainable Development**, John Wiley and Sons Ltd, 2026. ISSN 10991719. DOI 10.1002/sd.70894. Citado nas pp. 18, 20 e 21.
- GARCÍA-SÁNCHEZ, I. M. Efficiency measurement in spanish local government: The case of municipal water services. **Review of Policy Research**, v. 23, p. 355–372, 3 2006. ISSN 1541-132X. DOI 10.1111/j.1541-1338.2006.00205.x. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.1541-1338.2006.00205.x>. Citado nas pp. 76 e 78.
- GARRICK, D. E.; HANEMANN, M.; HEPBURN, C. Rethinking the economics of water: an assessment. **Oxford Review of Economic Policy**, v. 36, n. 1, p. 1–23, jan 2020. ISSN 0266-903X. DOI 10.1093/oxrep/grz035. Disponível em: <https://academic.oup.com/oxrep/article/36/1/1/5696678>. Citado nas pp. 22, 27, 29 e 61.
- GIBBS, E. The Fundamentals of Public Ownership: Learning from UK Historical Experience and Recent Scottish Policy. **Political Quarterly**, v. 95, n. 1, p. 157–166, 2024. ISSN 1467923X. DOI 10.1111/1467-923X.13348. Citado na p. 31.
- GUASCH, J. L. **Granting and Renegotiating Infrastructure Concessions**. Washington, D.C.: The World Bank, 2004. ISBN 0-8213-5792-1. Disponível em: <https://documents.worldbank.org/pt/publication/documents-reports/documentdetail/678041468765605224/granting-and-renegotiating-infrastructure-concessions-doing-it-right>. Citado na p. 58.
- GUASCH, J. L.; LAFFONT, J. J.; STRAUB, S. Renegotiation of concession contracts in latin america. evidence from the water and transport sectors. **International Journal of Industrial Or-**

- ganization**, v. 26, p. 421–442, 3 2008. ISSN 01677187. DOI 10.1016/j.ijindorg.2007.05.003. Citado na p. 60.
- HAILU, D.; OSORIO, R. G.; TSUKADA, R. Privatization and Renationalization: What Went Wrong in Bolivia's Water Sector? **World Development**, Elsevier Ltd, v. 40, n. 12, p. 2564–2577, 2012. ISSN 0305750X. DOI 10.1016/j.worlddev.2012.05.032. Disponível em: <http://dx.doi.org/10.1016/j.worlddev.2012.05.032>. Citado nas pp. 35, 36 e 110.
- HANNA, T. M.; MCDONALD, D. A. From pragmatic to politicized? The future of water remunicipalization in the United States. **Utilities Policy**, Elsevier Ltd, v. 72, n. July, p. 101276, 2021. ISSN 09571787. DOI 10.1016/j.jup.2021.101276. Disponível em: <https://doi.org/10.1016/j.jup.2021.101276>. Citado nas pp. 19 e 36.
- HANS, V. B. Sanitation – An Economic Perspective. **SSRN Electronic Journal**, p. 6, 2023. ISSN 1556-5068. DOI 10.2139/ssrn.4589984. Disponível em: <https://www.ssrn.com/abstract=4589984>. Citado na p. 19.
- HART, O.; SHLEIFER, A.; VISHNY, R. W. The proper scope of government: Theory and an application to prisons. **Quarterly Journal of Economics**, v. 112, n. 4, p. 1127–1161, 1997. ISSN 00335533. DOI 10.1162/003355300555448. Citado nas pp. 29, 135, 139, 140 e 145.
- HAUGHTON, G. Private profits - public drought: The creation of a crisis in water management for West Yorkshire. **Transactions of the Institute of British Geographers**, v. 23, n. 4, p. 419–433, 1998. ISSN 00202754. DOI 10.1111/j.0020-2754.1998.00419.x. Citado na p. 31.
- HAVSTAD, K.; PETERS, D.; SKAGGS, R.; BROWN, J.; BESTELMEYER, B.; FREDRICKSON, E.; HERRICK, J.; WRIGHT, J. Ecological services to and from rangelands of the United States. **Ecological Economics**, v. 64, n. 2, p. 261–268, 2007. ISSN 09218009 (ISSN). DOI 10.1016/j.ecolecon.2007.08.005. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-36549056975&doi=10.1016%2fj.ecolecon.2007.08.005&partnerID=40&md5=e39748a761c619e979b16cd16600313f>. Citado na p. 53.
- HAZILLA, M.; KOPP, R. Social cost of environmental quality regulations: a general equilibrium analysis. **Journal of Political Economy**, v. 98, n. 4, p. 853–873, 1990. ISSN 00223808 (ISSN). DOI 10.1086/261709. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0025631737&doi=10.1086%2f261709&partnerID=40&md5=a7f84c9b017a6c713ceb1d77bdc242f2>. Citado na p. 53.
- HELLER, L. Basic Sanitation in Brazil: Lessons from the Past, Opportunities from the Present, Challenges for the Future. **Journal of Comparative Social Welfare**, v. 23, n. 2, p. 141–153, 2007. ISSN 1748-6831. DOI 10.1080/17486830701494640. Citado nas pp. 19 e 38.
- HELM, D. Thirty years after water privatization—is the English model the envy of the world? **Oxford Review of Economic Policy**, v. 36, n. 1, p. 69–85, jan 2020. ISSN 0266-903X.

- DOI 10.1093/oxrep/grz031. Disponível em: <https://academic.oup.com/oxrep/article/36/1/69/5696683>. Citado na p. 76.
- HO, D. E.; KING, G.; STUART, E. A.; IMAI, K. MatchIt : Nonparametric Preprocessing for. **Journal Of Statistical Software**, v. 42, n. 8, p. 1–28, 2011. ISSN 1548-7660. Disponível em: <http://www.jstatsoft.org/v42/i08/paper>. Citado na p. 118.
- HON, L. Y.; BOON, T. H.; LEE, C. Technical efficiency in the malaysian water sector: A stochastic frontier approach. **The Journal of Developing Areas**, Johns Hopkins University Press, v. 50, p. 149–165, 2016. DOI 10.1353/jda.2016.0104. Citado na p. 76.
- HOUTVEN, G. L. V.; PATTANAYAK, S. K.; USMANI, F.; YANG, J. C. What are households willing to pay for improved water access? results from a meta-analysis. **ECOLOGICAL ECONOMICS**, v. 136, p. 126–135, 2017. ISSN 0921-8009. DOI 10.1016/j.ecolecon.2017.01.023. Citado na p. 61.
- Instituto Trata Brasil. **Qualidade da Regulação do Saneamento no Brasil e Oportunidades de Melhoria**. [S.l.], 2021. 1–261 p. Disponível em: <http://www.tratabrasil.org.br/pt/estudos/estudos-itb/itbs>. Citado na p. 39.
- ISRAEL, D. K. Impact of increased access and price on household water use in urban Bolivia. **Journal of Environment and Development**, v. 16, n. 1, p. 58–83, 2007. ISSN 10704965. DOI 10.1177/1070496506298190. Citado na p. 35.
- JEPSON, W.; Lee Brown, H. 'If no gasoline, no water': Privatizing drinking water quality in South Texas colonias. **Environment and Planning A**, v. 46, n. 5, p. 1032–1048, 2014. ISSN 14723409. DOI 10.1068/a46170. Citado na p. 36.
- JIANG, K.; ZHANG, J.; ZHANG, L.; WANG, D.; WANG, Y. Sustainable cooperation in the watershed ecological compensation public-private partnership project: Lessons from China's Chishui river basin. **Socio-Economic Planning Sciences**, v. 90, 2023. ISSN 00380121 (ISSN). Publisher: Elsevier Ltd. DOI 10.1016/j.seps.2023.101730. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85171590929&doi=10.1016%2fj.seps.2023.101730&partnerID=40&md5=8e63476e8fcdad0494b0dba2a3324afd>. Citado na p. 64.
- KRESCH, E. P. The buck stops where? Federalism, uncertainty, and investment in the Brazilian water and sanitation sector. **American Economic Journal: Economic Policy**, v. 12, n. 3, p. 374–401, 2020. ISSN 1945774X. DOI 10.1257/POL.20180327. Citado nas pp. 28, 29, 38 e 109.
- KUMBHAKAR, S. C.; LOVELL, C. A. K. **Stochastic Frontier Analysis**. 1. ed. Cambridge: Cambridge University Press, 2000. Citado nas pp. 82 e 84.

- KUMBHAKAR, S. C.; PARMETER, C. F.; ZELENYUK, V. Stochastic frontier analysis: Foundations and advances i. *In*: _____. **Handbook of Production Economics**. [S.l.]: Springer Singapore, 2021. p. 1–40. DOI 10.1007/978-981-10-3450-3₉ – 2. Citado na p. 85.
- KUMBHAKAR, S. C.; WANG, H.-J.; HORNCastle, A. P. **A practitioner's guide to stochastic frontier analysis using Stata**. [S.l.]: Cambridge University Press, 2015. Citado na p. 100.
- LAFFONT, J.; TIROLE, J. **A Theory of Incentives in Procurement and Regulation**. MIT Press, 1993. (Mit Press). ISBN 9780262121743. Disponível em: <https://books.google.com.br/books?id=4iH4Z2wbNqAC>. Citado nas pp. 49, 69 e 70.
- LAFFONT, J.-J.; TIROLE, J. **Competition in telecommunications**. [S.l.]: MIT press, 2001. Citado nas pp. 53 e 58.
- LANNIER, A. L.; PORCHER, S. Efficiency in the public and private french water utilities: prospects for benchmarking. **Applied Economics**, v. 46, p. 556–572, 2 2014. ISSN 00036846. DOI 10.1080/00036846.2013.857002. Citado nas pp. 76 e 78.
- LAZZARINI, S. G. **A privatização certa: Por que as empresas privadas em iniciativas públicas precisam de governos capazes**. São Paulo: Portfolio-Penguin, 2023. ISBN 978-65-5424-019-2. Citado nas pp. 24, 37, 40, 41, 45, 135, 136, 137, 138, 139, 140, 141 e 143.
- LEE, C. Privatization, water access and affordability: Evidence from Malaysian household expenditure data. **Economic Modelling**, Elsevier B.V., v. 28, n. 5, p. 2121–2128, sep 2011. ISSN 02649993. DOI 10.1016/j.econmod.2011.05.003. Disponível em: <http://dx.doi.org/10.1016/j.econmod.2011.05.003><https://linkinghub.elsevier.com/retrieve/pii/S0264999311001143>. Citado nas pp. 111 e 112.
- LEITE, C. H. P.; Moita Neto, J. M.; BEZERRA, A. K. L. Novo marco legal do saneamento básico: alterações e perspectivas. **Engenharia Sanitaria e Ambiental**, v. 27, n. 5, p. 1041–1047, oct 2022. ISSN 1809-4457. DOI 10.1590/s1413-415220210311. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-41522022000501041&tlng=pt. Citado na p. 40.
- LIM, S.-W.; CHOU, W.; CHEN, L. Sankeynetwork: a clear and concise visualization tool for bibliometric data. **MethodsX**, Elsevier, v. 14, p. 103379, 2025. Citado na p. 52.
- LIMA, S.; BROCHADO, A.; MARQUES, R. C. Public-private partnerships in the water sector: A review. **Utilities Policy**, v. 69, n. February, p. 101182, apr 2021. ISSN 09571787. DOI 10.1016/j.jup.2021.101182. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0957178721000163>. Citado nas pp. 19 e 112.
- LIU, M.; SHADBEGIAN, R.; ZHANG, B. Does environmental regulation affect labor demand in China? Evidence from the textile printing and dyeing industry. **Journal of Environmental Economics and Management**, v. 86, p. 277–294, 2017. ISSN 00950696

- (ISSN). Publisher: Academic Press Inc. DOI 10.1016/j.jeem.2017.05.008. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85020273466&doi=10.1016%2fj.jeem.2017.05.008&partnerID=40&md5=0d04f2ee3af006812214a8c272a1fdf6>. Citado na p. 53.
- LOBINA, E.; HALL, D. Public sector alternatives to water supply and sewerage privatization: Case studies. **International Journal of Water Resources Development**, v. 16, n. 1, p. 35–55, 2000. ISSN 07900627. DOI 10.1080/07900620048554. Citado nas pp. 110, 111 e 136.
- LOBINA, E.; HALL, D. Experience with private sector participation in Grenoble, France, and lessons on strengthening public water operations. **Utilities Policy**, v. 15, n. 2, p. 93–109, 2007. ISSN 09571787. DOI 10.1016/j.jup.2007.01.004. Citado na p. 33.
- MANN, P. C.; MIKESELL, J. L. Ownership and water system operation1. **JAWRA Journal of the American Water Resources Association**, v. 12, n. 5, p. 995–1004, 1976. DOI <https://doi.org/10.1111/j.1752-1688.1976.tb00216.x>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-1688.1976.tb00216.x>. Citado nas pp. 70 e 75.
- MARTÍNEZ-ESPIÑEIRA, R.; PÉREZ-URDIALES, M. Water affordability challenges in Latin America and the Caribbean : Accounting for coping costs due to reliance on multiple , non-exclusive sources. **World Development**, Elsevier Ltd, v. 186, n. November 2024, p. 106825, 2025. ISSN 0305-750X. DOI 10.1016/j.worlddev.2024.106825. Disponível em: <https://doi.org/10.1016/j.worlddev.2024.106825>. Citado na p. 108.
- MASSARUTTO, A. Servant of too many masters: Residential water pricing and the challenge of sustainability. **Utilities Policy**, Elsevier Ltd, v. 63, n. February, p. 101018, 2020. ISSN 09571787. DOI 10.1016/j.jup.2020.101018. Disponível em: <https://doi.org/10.1016/j.jup.2020.101018>. Citado na p. 28.
- MASSARUTTO, A.; GRASSETTI, L.; Lambardi di San Miniato, M.; MOLETTA, M. Efficient firms are all alike, but every inefficient firm is such in its own way: Heterogeneity of costs determinants in the Italian water sector. **Utilities Policy**, Elsevier Ltd, v. 84, n. November 2022, p. 101646, 2023. ISSN 09571787. DOI 10.1016/j.jup.2023.101646. Disponível em: <https://doi.org/10.1016/j.jup.2023.101646>. Citado nas pp. 73, 77, 78 e 103.
- MAZIOTIS, A.; MOLINOS-SENANTE, M. The impact of model specification and environmental variables on measuring the overall technical efficiency of water and sewerage services: Evidence from Chile. **Structural Change and Economic Dynamics**, Elsevier B.V., v. 61, p. 191–198, 2022. ISSN 0954349X. DOI 10.1016/j.strueco.2022.02.014. Disponível em: <https://doi.org/10.1016/j.strueco.2022.02.014>. Citado nas pp. 76 e 78.
- MAZIOTIS, A.; MOLINOS-SENANTE, M. Water-energy nexus: Benchmarking the energy efficiency of drinking water treatment plants accounting for heterogeneity. **Sustainable**

- Development**, John Wiley and Sons Ltd, 2026. ISSN 10991719. DOI 10.1002/sd.70797. Citado nas pp. 19 e 20.
- MAZIOTIS, A.; SALA-GARRIDO, R.; MOCHOLI-ARCE, M.; MOLINOS-SENANTE, M. Estimating potential economic savings of water utilities due to improvements on technical and allocative inefficiencies. **Socio-Economic Planning Sciences**, Elsevier Ltd, v. 98, n. November 2023, 2025. ISSN 00380121. DOI 10.1016/j.seps.2024.102142. Citado nas pp. 76 e 78.
- MCKENZIE, D.; MOOKHERJEE, D. The Distributive Impact of Privatization in Latin America: Evidence from Four Countries. **Economía**, v. 3, n. 2, p. 161–218, 2003. ISSN 1533-6239. DOI 10.1353/eco.2003.0006. Disponível em: <https://economia.lse.ac.uk/articles/10.1353/eco.2003.0006>. Citado nas pp. 53, 58 e 69.
- MEGGINSON, W. L.; NETTER, J. M. From State to Market: A Survey of Empirical Studies on Privatization. **Journal of Economic Literature**, v. 39, n. 2, p. 321–389, jun 2001. ISSN 0022-0515. DOI 10.1257/jel.39.2.321. Disponível em: <https://pubs.aeaweb.org/doi/10.1257/jel.39.2.321>. Citado nas pp. 53, 58, 69, 107 e 112.
- MEIRELES, F. **deflateBR: Deflate Nominal Brazilian Reais**. [S.l.], 2018. R package version 1.1.2. DOI 10.32614/CRAN.package.deflateBR. Disponível em: <https://CRAN.R-project.org/package=deflateBR>. Citado nas pp. 82 e 119.
- MENDES, A. T.; ALVES, C. de M. A. Asymmetries in water and sanitation services and how they are shaping urban drainage and stormwater management in brazil. **Utilities Policy**, Elsevier Ltd, v. 98, 2 2026. ISSN 09571787. DOI 10.1016/j.jup.2025.102114. Citado na p. 19.
- MERAN, G.; SIEHLOW, M.; HIRSCHHAUSEN, C. von. **The Economics of Water**. Cham: Springer International Publishing, 2021. v. 18. 48–50 p. (Springer Water, 4). ISSN 0893-7850. ISBN 978-3-030-48484-2. DOI 10.1007/978-3-030-48485-9. Disponível em: <http://link.springer.com/10.1007/978-3-030-48485-9>. Citado nas pp. 18, 22, 27 e 28.
- MOLINOS-SENANTE, M.; MAZIOTIS, A. Cost efficiency of english and welsh water companies: a meta-stochastic frontier analysis. **Water Resources Management**, Springer Netherlands, v. 33, p. 3041–3055, 7 2019. ISSN 15731650. DOI 10.1007/s11269-019-02287-8. Citado nas pp. 77, 78 e 103.
- MOLINOS-SENANTE, M.; MAZIOTIS, A.; VILLEGAS, A. Performance analysis of Chilean water companies after the privatization of the industry: the influence of ownership. **Water International**, Routledge, v. 47, n. 1, p. 114–131, 2022. ISSN 02508060. DOI 10.1080/02508060.2021.1999761. Disponível em: <https://doi.org/10.1080/02508060.2021.1999761>. Citado nas pp. 76 e 78.

- MOLINOS-SENANTE, M.; SALA-GARRIDO, R. The impact of privatization approaches on the productivity growth of the water industry: A case study of Chile. **Environmental Science and Policy**, v. 50, p. 166–179, 2015. ISSN 18736416. DOI 10.1016/j.envsci.2015.02.015. Citado nas pp. 31, 32 e 112.
- MOTTA, R. S. da; MOREIRA, A. Efficiency and regulation in the sanitation sector in Brazil. **Utilities Policy**, Elsevier BV, v. 14, p. 185–195, 2006. ISSN 09571787. DOI 10.1016/j.jup.2006.03.002. Citado nas pp. 79, 81, 101, 102 e 104.
- NARZETTI, D. A.; MARQUES, R. C. Access to Water and Sanitation Services in Brazilian Vulnerable Areas: The Role of Regulation and Recent Institutional Reform. **Water**, v. 13, n. 6, p. 787, mar 2021. ISSN 2073-4441. DOI 10.3390/w13060787. Disponível em: <https://www.mdpi.com/2073-4441/13/6/787>. Citado na p. 39.
- NARZETTI, D. A.; MARQUES, R. C. Policies and incentives for developing universal access to water and sanitation for vulnerable families. **Water Policy**, v. 24, n. 3, p. 485–499, mar 2022. ISSN 1366-7017. DOI 10.2166/wp.2022.227. Disponível em: <https://iwaponline.com/wp/article/24/3/485/87564/Policies-and-incentives-for-developing-universal>. Citado na p. 40.
- NICKSON, A.; VARGAS, C. The limitations of water regulation: The failure of the Cochabamba concession in Bolivia. **Bulletin of Latin American Research**, v. 21, n. 1, p. 99–120, 2002. ISSN 02613050. DOI 10.1111/1470-9856.00034. Citado na p. 36.
- NUNES, C. M.; MARQUES, R. C. Privatization pathways in Brazilian water utilities: The role of financial standing. **Utilities Policy**, Elsevier Ltd, v. 97, p. 102052, 12 2025. ISSN 09571787. DOI 10.1016/j.jup.2025.102052. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0957178725001675>. Citado nas pp. 68, 69, 111, 112 e 113.
- NYATHIKALA, S. A.; JAMASB, T.; LLORCA, M.; KULSHRESTHA, M. Utility governance, incentives, and performance: Evidence from India's urban water sector. **Utilities Policy**, Elsevier Ltd, v. 82, n. March, p. 101534, 2023. ISSN 09571787. DOI 10.1016/j.jup.2023.101534. Disponível em: <https://doi.org/10.1016/j.jup.2023.101534>. Citado nas pp. 75, 76 e 78.
- OGATA, R.; MATSUMOTO, S.; OTSUKA, T.; HATTORI, Y.; KITAWAKI, H. **Influence of customer satisfaction and perceived water service quality on urban water supply management: A systematic review**. [S.l.]: Elsevier Ltd, 2026. DOI 10.1016/j.clrc.2026.100389. Citado nas pp. 18 e 21.
- OLIVEIRA, A. R. de. Private Provision of Water Service in Brazil: Impacts on Access and Affordability. **Munich Personal RePEc Archive Private**, n. 11149, 2008. Disponível em: <https://mpra.ub.uni-muenchen.de/11149/>. Citado nas pp. 38 e 113.
- OLIVEIRA, G.; OLIVEIRA, R. P.; ANDRADE, J. M.; PEREIRA, F. R. G. A review of case law in the water and sanitation sector: A lack of consensus, the Sabesp's case, and

- what to expect in the near future¹. **Revista de Defesa da Concorrência**, Administrative Council for Economic Defense (CADE), v. 13, p. 10–35, 2025. ISSN 23182253. DOI 10.52896/rdc.v13i1.1903. Citado na p. 20.
- OLIVEIRA, W. T. de; SAIANI, C. C. S. Trade-off Custo-Qualidade na Provisão de Saneamento Básico no Brasil. **Estudos Econômicos (São Paulo)**, v. 52, n. 4, p. 769–808, dec 2022. ISSN 1980-5357. DOI 10.1590/1980-53575243wocs. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-41612022000400769&tlng=pt. Citado nas pp. 19, 24, 28, 29, 66, 67, 69, 70, 71, 111, 113 e 142.
- ONDA, K. S.; TEWARI, M. Water systems in California: Ownership, geography, and affordability. **Utilities Policy**, Elsevier Ltd, v. 72, n. July, p. 101279, 2021. ISSN 09571787. DOI 10.1016/j.jup.2021.101279. Disponível em: <https://doi.org/10.1016/j.jup.2021.101279>. Citado na p. 36.
- OSTROM, E. **Governing the commons: The evolution of institutions for collective action**. [S.l.]: Cambridge university press, 1990. Citado nas pp. 23 e 58.
- PAUL, C. J. M. **Cost Structure and the Measurement of Economic Performance: Productivity, Utilization, Cost Economics, and Related Performance Indicators**. New York: Springer Science & Business Media, 1999. DOI 10.1007/978-1-4615-5093-8. Citado nas pp. 82 e 88.
- PERARD, E. Economic and financial aspects of the sanitation challenge: A practitioner approach. **Utilities Policy**, Elsevier, v. 52, n. March, p. 22–26, 2018. ISSN 09571787. DOI 10.1016/j.jup.2018.03.007. Disponível em: <https://doi.org/10.1016/j.jup.2018.03.007>. Citado nas pp. 18 e 29.
- POLASKY, S.; NELSON, E.; PENNINGTON, D.; JOHNSON, K. A. The impact of land-use change on ecosystem services, biodiversity and returns to landowners: A case study in the state of Minnesota. **Environmental and Resource Economics**, v. 48, n. 2, p. 219–242, 2011. ISSN 09246460. DOI 10.1007/s10640-010-9407-0. Citado na p. 53.
- PORCHER, S. The ‘hidden costs’ of water provision: New evidence from the relationship between contracting-out and price in French water public services. **Utilities Policy**, Elsevier Ltd, v. 48, p. 166–175, 2017. ISSN 09571787. DOI 10.1016/j.jup.2017.08.002. Disponível em: <https://doi.org/10.1016/j.jup.2017.08.002>. Citado nas pp. 32 e 33.
- PRICE, J. I.; RENZETTI, S.; DUPONT, D.; ADAMOWICZ, W.; EMELKO, M. B. Production costs, inefficiency, and source water quality: A stochastic cost frontier analysis of canadian water utilities. **Land Economics**, v. 93, p. 1–11, 2 2017. ISSN 0023-7639. DOI 10.3368/le.93.1.1. Disponível em: <http://le.uwpress.org/lookup/doi/10.3368/le.93.1.1>. Citado nas pp. 77 e 78.

- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2024. Disponível em: <https://www.R-project.org/>. Citado na p. 51.
- ROSENBAUM, P. R.; RUBIN, D. B. The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, v. 70, n. 1, p. 41, apr 1983. ISSN 00063444. DOI 10.2307/2335942. Disponível em: <https://www.jstor.org/stable/2335942?origin=crossref>. Citado na p. 116.
- RUESTER, S.; ZSCHILLE, M. The impact of governance structure on firm performance: An application to the German water distribution sector. **Utilities Policy**, Elsevier Ltd, v. 18, n. 3, p. 154–162, 2010. ISSN 09571787. DOI 10.1016/j.jup.2010.03.003. Disponível em: <http://dx.doi.org/10.1016/j.jup.2010.03.003>. Citado nas pp. 33 e 34.
- RUFINO, M. A.; CAVALCANTE, P. R. N. The effect of capital ownership on the relationship between the regulatory process and companies' abnormal return. **Revista Contabilidade & Finanças**, v. 35, n. 94, 2024. ISSN 1808-057X. DOI 10.1590/1808-057x20231839.en. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-70772024000100509&tlng=en. Citado nas pp. 19, 67, 69 e 70.
- SAAL, D.; PARKER, D. Productivity and Price Performance in the Privatized Water and Sewerage Companies of England and Wales. **Journal of Regulatory Economics**, v. 20, n. 1, p. 61–90, 2001. ISSN 0922680X (ISSN). DOI 10.1023/A:1011162214995. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0035599932&doi=10.1023%2fA%3a1011162214995&partnerID=40&md5=b3982f0351b2bb02f85bce9c13faf76c>. Citado na p. 30.
- SAAL, D.; PARKER, D.; WEYMAN-JONE, T. Determining the contribution of technical change, efficiency change and scale change to productivity growth in the privatized English and Welsh water and sewerage industry: 1985-2000. **JOURNAL OF PRODUCTIVITY ANALYSIS**, v. 28, n. 1-2, p. 127–139, out. 2007. ISSN 0895-562X. DOI 10.1007/s11123-007-0040-z. Citado nas pp. 53, 77, 78 e 103.
- SAAL, D. S.; PARKER, D.; WEYMAN-JONES, T. Determining the contribution of technical change, efficiency change and scale change to productivity growth in the privatized English and Welsh water and sewerage industry: 1985-2000. **Journal of Productivity Analysis**, v. 28, n. 1-2, p. 127–139, 2007. ISSN 0895562X. DOI 10.1007/s11123-007-0040-z. Citado na p. 30.
- SABBIONI, G. Efficiency in the brazilian sanitation sector. **Utilities Policy**, Elsevier BV, v. 16, p. 11–20, 2008. ISSN 09571787. DOI 10.1016/j.jup.2007.06.003. Citado nas pp. 79, 81, 104 e 105.
- SAIANI, C.; AZEVEDO, P. F. de. Is privatization of sanitation services good for health? **Utilities Policy**, Elsevier, v. 52, n. October 2017, p. 27–36, 2018. ISSN 09571787. DOI

- 10.1016/j.jup.2018.03.003. Citado nas pp. 19, 29, 38, 111, 112, 114, 117, 135, 136, 138, 140 e 141.
- SAMPAIO, P. R.; SAMPAIO, R. S. The challenges of regulating water and sanitation tariffs under a three-level shared-authority federalism model: The case of Brazil. **Utilities Policy**, Elsevier Ltd, v. 64, 6 2020. ISSN 09571787. DOI 10.1016/j.jup.2020.101049. Citado nas pp. 80 e 103.
- SATHIRATHAI, S.; BARBIER, E. Valuing mangrove conservation in Southern Thailand. **Contemporary Economic Policy**, v. 19, n. 2, p. 109–122, 2001. ISSN 10743529 (ISSN). Publisher: Oxford University Press. DOI 10.1111/j.1465-7287.2001.tb00054.x. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-23044527662&doi=10.1111%2fj.1465-7287.2001.tb00054.x&partnerID=40&md5=f38c1083b54069e08908fe74965fa84b>. Citado na p. 53.
- SHARKEY, W. W. **The Theory of Natural Monopoly**. Cambridge University Press, 1982. 224 p. ISBN 9780521243940. DOI 10.1017/CBO9780511571817. Disponível em: <https://www.cambridge.org/core/product/identifier/9780511571817/type/book>. Citado nas pp. 23 e 24.
- SHLEIFER, A. A theory of yardstick competition. **The RAND Journal of Economics**, v. 16, p. 319, 1985. ISSN 07416261. DOI 10.2307/2555560. Disponível em: <http://doi.wiley.com/10.2307/2555560>. Citado nas pp. 73, 75, 80, 103 e 105.
- SILVA, F. G.; LEDO, B. C. A. ASPECTOS REGULATÓRIOS E SEUS EFEITOS NO SETOR DE ÁGUA E ESGOTO DO BRASIL. **Revista de Defesa da Concorrência**, v. 11, n. 1, p. 127–148, jul 2023. ISSN 2318-2253. DOI 10.52896/rdc.v11i1.1045. Disponível em: <https://revista.cade.gov.br/index.php/revistadedefesadaconcorrencia/article/view/1045>. Citado nas pp. 74, 79, 80, 81, 83, 87, 101, 102, 103 e 104.
- SILVA, F. G.; LEDO, B. C. A. PANORAMA DAS ESTRUTURAS TARIFÁRIAS NO SANEAMENTO BÁSICO. **Revista de Administração, Contabilidade e Economia da Fundace**, v. 14, n. 4, oct 2023. ISSN 2178-7638. DOI 10.13059/racef.v14i4.1120. Disponível em: <https://www.fundace.org.br/revistaracef/index.php/racef/article/view/1120>. Citado nas pp. 39, 73 e 74.
- SOUSA, C. O. M. de; FÁVERO, L. P. L.; FOUTO, N. M. M. D. Ownership in times of water scarcity: are inhabitants supplied by private utilities consuming more water? **AQUA — Water Infrastructure, Ecosystems and Society**, v. 00, n. 0, p. 1–11, jan 2025. ISSN 2709-8028. DOI 10.2166/aqua.2025.240. Disponível em: <https://iwaponline.com/aqua/article/doi/10.2166/aqua.2025.240/106969/Ownership-in-times-of-water-scarcity-are>. Citado nas pp. 68 e 69.
- SOUSA, L. V. de C.; TORRES, M. de O. Climate and the productive structure of the urban water industry. **Urban Water Journal**, Taylor and Francis Ltd., v. 19, p. 673–685, 8

2022. ISSN 1573-062X. DOI 10.1080/1573062X.2022.2075763. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1573062X.2022.2075763>. Citado nas pp. 82, 87 e 88.
- SUEYOSHI, T.; GOTO, M. Efficiency-based rank assessment for electric power industry: a combined use of data envelopment analysis (dea) and dea-discriminant analysis (da). **Energy Economics**, Elsevier, v. 34, n. 3, p. 634–644, 2012. Citado na p. 58.
- SUÁREZ-VARELA, M.; GARCÍA-VALIÑAS, M. de los Ángeles; GONZÁLEZ-GÓMEZ, F.; PICAZO-TADEO, A. J. Ownership and performance in water services revisited: Does private management really outperform public? **Water Resources Management**, Springer Netherlands, v. 31, p. 2355–2373, 6 2017. ISSN 15731650. DOI 10.1007/s11269-016-1495-3. Citado na p. 76.
- TAN, J. The pitfalls of water privatization: Failure and reform in malaysia. **World Development**, v. 40, p. 2552–2563, 12 2012. ISSN 0305750X. DOI 10.1016/j.worlddev.2012.05.012. Citado na p. 76.
- TEEPLES, R.; GLYER, D. Cost of water delivery systems: specification and ownership effects. **The Review of Economics and Statistics**, JSTOR, p. 399–408, 1987. Citado nas pp. 53 e 69.
- TIROLE, J. **The Theory of Industrial Organization**. London, England; Cambridge, Massachusetts: The MIT Press, 1988. 479 p. ISBN 0262200716. Citado nas pp. 9, 23, 24, 25, 26 e 27.
- TORRES, M.; PAUL, C. J. M. Driving forces for consolidation or fragmentation of the us water utility industry: A cost function approach with endogenous output. **Journal of Urban Economics**, v. 59, p. 104–120, 1 2006. ISSN 00941190. DOI 10.1016/j.jue.2005.09.003. Citado nas pp. 82, 83, 87, 88, 104 e 105.
- TOSUN, J.; TRIEBSKORN, R. Civil society and the governance of water services: German political parties' reactions to right2water. **Water (Switzerland)**, v. 12, n. 3, 2020. ISSN 20734441. DOI 10.3390/w12030743. Citado na p. 34.
- TOURINHO, M.; SANTOS, P. R.; PINTO, F. T.; CAMANHO, A. S. Performance assessment of water services in Brazilian municipalities: An integrated view of efficiency and access. **Socio-Economic Planning Sciences**, Elsevier Ltd, v. 79, n. May 2021, p. 101139, 2022. ISSN 00380121. DOI 10.1016/j.seps.2021.101139. Disponível em: <https://doi.org/10.1016/j.seps.2021.101139>. Citado nas pp. 39, 66, 67 e 69.
- TUPPER, H. C.; RESENDE, M. Efficiency and regulatory issues in the brazilian water and sewage sector: An empirical study. **Utilities Policy**, Elsevier BV, v. 12, p. 29–40, 2004. ISSN 09571787. DOI 10.1016/j.jup.2003.11.001. Citado nas pp. 73, 75, 79, 80, 81 e 103.

- United Nations. **The Human Right to Water and Sanitation**. 2010. G.A. Res. 64/292, UN DOC. A/RES/64/292 (August 3, 2010). Citado na p. 18.
- United Nations. **The Sustainable Development Goals Report**. New York, 2024. 48 p. Disponível em: <https://unstats.un.org/sdgs/report/2024>. Citado na p. 18.
- VICKERS, J.; YARROW, G. K. **Privatization: An economic analysis**. [S.l.]: MIT press, 1988. v. 18. Citado na p. 53.
- Vitor de Carvalho Sousa, L.; de Oliveira Torres, M. Climate and the productive structure of the urban water industry. **Urban Water Journal**, Taylor & Francis, v. 19, n. 7, p. 673–685, aug 2022. ISSN 1573-062X. DOI 10.1080/1573062X.2022.2075763. Disponível em: <https://doi.org/10.1080/1573062X.2022.2075763><https://www.tandfonline.com/doi/full/10.1080/1573062X.2022.2075763>. Citado na p. 83.
- WAIT, I. W.; PETRIE, W. A. Comparison of water pricing for publicly and privately owned water utilities in the United States. **Water International**, Routledge, v. 42, n. 8, p. 967–980, 2017. ISSN 02508060. DOI 10.1080/02508060.2017.1406782. Disponível em: <https://doi.org/10.1080/02508060.2017.1406782>. Citado na p. 36.
- WALLSTEN, S.; KOSEC, K. The effects of ownership and benchmark competition: An empirical analysis of U.S. water systems. **International Journal of Industrial Organization**, v. 26, n. 1, p. 186–205, 2008. ISSN 01677187 (ISSN). DOI 10.1016/j.ijindorg.2006.11.001. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-37349005207&doi=10.1016%2fj.ijindorg.2006.11.001&partnerID=40&md5=773d5044069d99cb7fde7dd94d02cfb4>. Citado na p. 58.
- WALLSTEN, S.; KOSEC, K. The effects of ownership and benchmark competition: An empirical analysis of U.S. water systems. **International Journal of Industrial Organization**, v. 26, n. 1, p. 186–205, 2008. ISSN 01677187. DOI 10.1016/j.ijindorg.2006.11.001. Citado na p. 36.
- WALTER, M.; CULLMANN, A.; HIRSCHHAUSEN, C. von; WAND, R.; ZSCHILLE, M. Quo vadis efficiency analysis of water distribution? A comparative literature review. **Utilities Policy**, Elsevier Ltd, v. 17, n. 3-4, p. 225–232, 2009. ISSN 09571787. DOI 10.1016/j.jup.2009.05.002. Disponível em: <http://dx.doi.org/10.1016/j.jup.2009.05.002>. Citado nas pp. 30, 74, 75 e 76.
- WARNER, M. E. Key issues in water privatization and remunicipalization. **Utilities Policy**, v. 73, n. September, 2021. ISSN 09571787. DOI 10.1016/j.jup.2021.101300. Citado nas pp. 19 e 36.
- WARNER, M. E. Pragmatic Municipalism: Privatization and Remunicipalisation in the US. **Local Government Studies**, Routledge, v. 50, n. 4, p. 677–695, 2024. ISSN 17439388.

DOI 10.1080/03003930.2022.2162884. Disponível em: <https://doi.org/10.1080/03003930.2022.2162884>. Citado nas pp. 19 e 37.

WARNER, M. E.; BEL, G. Competition or monopoly? Comparing privatization of local public services in the US and Spain. **Public Administration**, v. 86, n. 3, p. 723–735, 2008. ISSN 00333298. DOI 10.1111/j.1467-9299.2008.00700.x. Citado na p. 36.

WEIL, D. N. Accounting for the effect of health on economic growth. **The Quarterly Journal of Economics**, v. 122, p. 1265–1306, 8 2007. ISSN 0033-5533. DOI 10.1162/qjec.122.3.1265. Disponível em: <https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.122.3.1265>. Citado na p. 73.

World Health Organization WEDC. **How much water is needed in emergencies**. Geneva, 2013. 1–4 p. Disponível em: <https://cdn.who.int/media/docs/default-source/wash-documents/who-tn-09-how-much-water-is-needed.pdf>. Citado na p. 28.

WORTHINGTON, A. C. A review of frontier approaches to efficiency and productivity measurement in urban water utilities. **Urban Water Journal**, v. 11, p. 55–73, 1 2014. ISSN 1573062X. DOI 10.1080/1573062X.2013.765488. Citado nas pp. 74, 75 e 76.

ZHAO, J. Productivity change in the privatized water sector in china (1999–2006). **Journal of Productivity Analysis**, Springer, v. 53, p. 227–241, 4 2020. ISSN 15730441. DOI 10.1007/s11123-019-00572-7. Citado na p. 76.

Apêndice A – Tabelas longas

Tabela A.1 – Harmonização dos Indicadores de Saneamento: Equivalências entre SINISA (2023-2024) e SNIS (2017-2022). Brasil, 2025.

Código SINISA	Código SNIS	Nome do Indicador	Grupo	Unidade
IAG1002	IN001	Densidade de economias de água por ligação	Estruturais e Operacionais	econ./ligação
IFA2002	IN003	Despesa total média de água incluindo trib...	Administrativo e Financeiro	R\$/m³
IFA1001	IN005	Receita operacional direta média de usuári...	Administrativo e Financeiro	R\$/m³
IFA2006	IN008	Despesa média com pessoal próprio do servi...	Administrativo e Financeiro	R\$/empregado.ano
IAG1003	IN009	Incidência da hidrometração de água	Estruturais e Operacionais	Percentual
IAG2001	IN010	Micromedição de água em relação ao volume ...	Estruturais e Operacionais	Percentual
IAG2003	IN011	Macromedição do volume de água de entrada ...	Estruturais e Operacionais	Percentual
IFA2001	IN012	Desempenho financeiro do serviço de abaste...	Administrativo e Financeiro	Percentual
IAG2012	IN013	Perdas de faturamento de água	Estruturais e Operacionais	Percentual
IAG2004	IN014	Consumo micromedido de água por economia	Estruturais e Operacionais	m³/economia/mês
IAG2005	IN017	Consumo de água faturado por economia	Estruturais e Operacionais	m³/economia/mês
IAG1001	IN020	Extensão de rede de distribuição de água p...	Estruturais e Operacionais	m/ligação
IAG2006	IN022	Consumo total médio per capita de água	Estruturais e Operacionais	l/hab/dia
IAG0002	IN023	Atendimento da população urbana com rede d...	Atendimento	Percentual
IAG2008	IN025	Volume de água disponibilizado para distri...	Estruturais e Operacionais	m³/economia/mês
IFA2004	IN026	Despesa de exploração média de água	Administrativo e Financeiro	R\$/m³
IFA2005	IN027	Despesa de exploração média de água por ec...	Administrativo e Financeiro	R\$/economia.ano
IAG2011	IN028	Volume faturado de água em relação ao volu...	Estruturais e Operacionais	Percentual
IFA1004	IN029	Evasão de receitas do serviço de água	Administrativo e Financeiro	percentual
IFA2011	IN030	Margem da despesa de exploração de água	Administrativo e Financeiro	Percentual
IFA2012	IN031	Margem da despesa com pessoal próprio do s...	Administrativo e Financeiro	Percentual
IFA2013	IN032	Margem da despesa com pessoal total do ser...	Administrativo e Financeiro	Percentual
IFA2014	IN033	Margem da despesa com serviço da dívida de...	Administrativo e Financeiro	Percentual
IFA2008	IN035	Incidência da despesa de pessoal próprio n...	Administrativo e Financeiro	Percentual
IFA200G	IN036	Incidência da despesa de pessoal total nas...	Administrativo e Financeiro	Percentual
IFA2010	IN037	Incidência da despesa de energia elétrica ...	Administrativo e Financeiro	Percentual
IFA1003	IN042	Participação da receita operacional indire...	Administrativo e Financeiro	percentual

Tabela A.1 – Harmonização dos Indicadores de Saneamento: Equivalências entre SINISA (2023-2024) e SNIS (2017-2022). Brasil, 2025. *(continued)*

Código SINISA	Código SNIS	Nome do Indicador	Grupo	Unidade
IAG1004	IN043	Incidência das economias residenciais ativ...	Estruturais e Operacionais	Percentual
IAG2002	IN044	Micromedição do volume de água consumido	Estruturais e Operacionais	Percentual
IAG2013	IN049	Perdas totais de água na distribuição	Estruturais e Operacionais	Percentual
IAG2014	IN050	Perdas totais lineares de água na rede de ...	Estruturais e Operacionais	m³/km/dia
IAG2015	IN051	Perdas totais de água por ligação	Estruturais e Operacionais	l/lig/dia
IAG2010	IN052	Consumo de água em relação ao volume dispo...	Estruturais e Operacionais	Percentual
IAG200G	IN053	Consumo total médio de água por economia	Estruturais e Operacionais	m³/economia/mês
IFA1005	IN054	Dias de faturamento comprometidos com cont...	Administrativo e Financeiro	dias
IAG0001	IN055	Atendimento da população total com rede de...	Atendimento	Percentual
IAG2023	IN058	Consumo médio de energia elétrica no servi...	Estruturais e Operacionais	KWh/m³
IFA2007	IN060	Despesa média de energia elétrica do servi...	Administrativo e Financeiro	R\$/KWh
IFB3001	IN061	Liquidez corrente	Balanço	-
IFB3002	IN062	Liquidez geral	Balanço	-
IFB3003	IN063	Grau de endividamento	Balanço	-
IFB3004	IN064	Margem operacional com depreciação	Balanço	Percentual
IFB3005	IN065	Margem líquida com depreciação	Balanço	Percentual
IFB3008	IN066	Retorno sobre o patrimônio líquido	Balanço	Percentual
IFB300G	IN067	Composição das exigibilidades	Balanço	Percentual
IFB3006	IN068	Margem operacional sem depreciação	Balanço	Percentual
IFB3007	IN069	Margem líquida sem depreciação	Balanço	Percentual
IAG3001	IN071	Economias ativas de água atingidas por par...	Qualidade	economias/paralisação
IAG3004	IN072	Duração média das paralisações	Qualidade	horas/paralisação
IAG3002	IN073	Economias ativas de água atingidas por int...	Qualidade	economias/interrupção sistemática
IAG3005	IN074	Duração média das interrupções sistemáticas	Qualidade	horas/interrupção sistemática
IAG3003	IN083	Tempo médio de reparo de vazamentos de água	Qualidade	horas/reparo

Tabela A.2 – Validação da Metodologia: Correlação e RMSE entre Indicadores Calculados e SNIS Original (municípios com prestador único). Brasil, 2025.

Código	N pares	Correlação	RMSE (%)	Bias (%)	p-valor	Válido?
IN003	26700	1.0000	0.05	0.00	0.189	Sim
IN005	26455	1.0000	0.04	86.09	0.907	Sim
IN012	27774	1.0000	0.00	51.45	0.951	Sim
IN026	26703	1.0000	0.06	0.00	0.569	Sim
IN029	26885	1.0000	0.26	200.83	0.502	Sim
IN030	26799	1.0000	0.00	0.00	0.214	Sim
IN031	26798	1.0000	0.00	10.15	0.269	Sim
IN032	26794	1.0000	0.00	4.66	0.115	Sim
IN033	26797	1.0000	0.01	673.20	0.218	Sim
IN035	27774	1.0000	0.01	0.14	0.394	Sim
IN036	27770	1.0000	0.00	0.00	0.705	Sim
IN037	27775	1.0000	0.01	30.83	0.546	Sim
IN042	26883	1.0000	0.06	330.91	0.084	Sim
IN054	24723	1.0000	0.00	15.39	0.528	Sim
IN055	27778	1.0000	0.03	0.00	0.108	Sim
IN060	24864	1.0000	0.38	515.67	0.350	Sim
IN010	27642	1.0000	0.00	64.87	0.715	Sim
IN013	27642	1.0000	0.01	31.68	0.607	Sim
IN028	27642	1.0000	0.00	0.00	0.983	Sim
IN049	27637	1.0000	0.01	66.87	0.222	Sim
IN052	27642	1.0000	0.00	0.00	0.642	Sim
IN058	25433	1.0000	0.31	702.85	0.523	Sim
IN071	12442	1.0000	0.00	0.00	0.000	Sim
IN072	14490	1.0000	0.01	0.00	0.000	Sim
IN073	6109	1.0000	0.00	0.00	0.001	Sim
IN074	7532	1.0000	0.00	0.00	0.000	Sim
IN083	22705	1.0000	0.01	371.09	0.771	Sim
IN044	27641	0.9992	1.48	46.69	0.224	Sim
IN023	23306	0.9970	1.46	0.01	0.318	Sim
IN009	27624	0.9945	3.64	0.48	0.000	Sim
IN053	27622	0.9922	6.99	-0.86	0.000	Sim
IN017	27409	0.9903	6.76	13.04	0.000	Sim
IN050	27763	0.9886	24.49	269.06	0.000	Não
IN020	27769	0.9772	23.21	1.03	0.000	Não
IN025	27622	0.9657	21.52	-0.86	0.000	Não
IN051	27617	0.9594	50.97	44.47	0.355	Não
IN027	27031	0.9254	29.13	-0.68	0.000	Não
IN022	27632	0.9228	21.07	0.21	0.057	Não
IN001	27624	0.8921	10.54	0.08	0.057	Não
IN014	25971	0.8254	34.42	7.31	0.000	Não
IN008	27166	0.6580	93.94	2.54	0.000	Não
IN043	26428	0.5719	9.64	-0.86	0.000	Não

Tabela A.2 – Validação da Metodologia: Correlação e RMSE entre Indicadores Calculados e SNIS Original (municípios com prestador único). Brasil, 2025. (*continued*)

Código	N pares	Correlação	RMSE (%)	Bias (%)	p-valor	Válido?
IN011	27572	0.5052	137.91	-182014810.76	0.000	Não

Note:

Critério de validação: correlação $\geq 0,95$ e RMSE relativo $\leq 20\%$.

Comparação restrita a municípios com prestador único (agregado municipal = registro do prestador).

Indicadores válidos: 32 de 43.

Apêndice B – Códigos de programação

B.1 Bibliometria

Código B.1 – Bibliometria

```

1  # =====
2  # ANLISE BIBLIOMTRICA COMPLETA - SANEAMENTO E PRIVATIZAO
3  # =====
4
5  # Carregar bibliotecas necessárias
6  library(bibliometrix)
7  library(tidyverse)
8  library(igraph)
9  library(networkD3)
10 library(viridis)
11 library(scales)
12 library(ggrepel)
13 library(patchwork)
14
15 # Configurar diretório
16 setwd("C:/Users/Pedro Ricelly/Google Drive/Doutorado/Tese/Artigo 1/atualização 23-11-2025")
17
18
19 # Limpar ambiente
20 rm(list = ls())
21 gc()
22
23 # =====
24 # 1. IMPORTAO E PREPARAO DOS DADOS
25 # =====
26
27 # Importar bases de dados
28 scopus <- convert2df("scopus_export_Nov 23-2025_d6c0c57e-1107-4a7c-80de-ece9292aa908.csv",
29                      dbsource = "scopus", format = "csv")
30 wos <- convert2df("savedrecs.txt", dbsource = "vos", format = "plaintext")
31
32 # Adicionar identificadores únicos
33 scopus <- scopus %>% mutate(id = paste("SCOPUS", row_number(), sep = "_"))
34 wos <- wos %>% mutate(id = paste("WOS", row_number(), sep = "_"))
35
36 # Mesclar bases
37 database <- mergeDbSources(scopus, wos)
38 cat("\n Bases mescladas:", nrow(database), "documentos únicos\n")
39
40 # Ajustar citações
41 database <- database %>%
42   select(-CR) %>%
43   rename(CR = CR_raw)
44
45 # Salvar base processada
46 save(database, file = "database.RData")
47
48 # =====
49 # 2. ANLISE DESCRITIVA BSICA
50 # =====
51
52 Res <- biblioAnalysis(database, sep = ";")
53 DS <- summary(object = Res, k = 10, pause = FALSE)
54

```

```

55 # Estatísticas gerais
56 cat("\n=====")
57 cat("\n  ESTATÍSTICAS GERAIS DA AMOSTRA")
58 cat("\n=====")
59 cat("Total de documentos:", nrow(database), "\n")
60 cat("Período:", min(database$PY, na.rm = TRUE), "-", max(database$PY, na.rm = TRUE), "\n")
61
62 # Calcular idade média dos documentos
63 current_year <- as.numeric(format(Sys.Date(), "%Y"))
64 avg_age <- round(mean(current_year - database$PY, na.rm = TRUE), 1)
65
66 cat("Total de autores:", Res$nAuthors, "\n")
67 cat("Total de periódicos:", length(unique(database$SO)), "\n")
68 cat("Idade média dos documentos:", avg_age, "anos\n")
69 cat("Média de citações por documento:", round(mean(database$TC, na.rm = TRUE), 2), "\n")
70 cat("Média de citações por ano por documento:", round(mean(database$TC/((current_year -
    ↪ database$PY) + 1), na.rm = TRUE), 2), "\n")
71
72 # =====
73 # 3. GRFICOS - AUTORES MAIS PRODUTIVOS
74 # =====
75
76 # Preparar dados
77 top_authors <- as.data.frame(DS$MostProdAuthors)
78 colnames(top_authors) <- make.unique(colnames(top_authors)) # Tornar nomes únicos
79
80 top_authors <- top_authors %>%
81   slice_head(n = 10) %>%
82   mutate(Articles = as.numeric(Articles))
83
84 # Identificar coluna de autores (primeira coluna ou com "Author" no nome)
85 author_col <- colnames(top_authors)[1]
86
87 # Gráfico
88 g1 <- ggplot(top_authors, aes(y = reorder(.data[[author_col]], Articles), x = Articles)) +
89   geom_bar(stat = "identity", fill = "steelblue", alpha = 0.8) +
90   geom_text(aes(label = Articles), hjust = -0.3, size = 4) +
91   labs(title = "Top 10 Autores Mais Produtivos",
92        y = "Autor",
93        x = "Número de Publicações") +
94   theme_minimal(base_size = 14) +
95   theme(
96     plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
97     axis.text.y = element_text(size = 12),
98     panel.grid.minor = element_blank()
99   ) +
100   scale_x_continuous(expand = expansion(mult = c(0, 0.15)))
101
102 ggsave("fig_top_authors.png", g1, width = 10, height = 6, dpi = 300, bg = "white")
103 cat("\n Gráfico salvo: fig_top_authors.png\n")
104
105 # =====
106 # 4. ARTIGOS MAIS CITADOS
107 # =====
108
109 top_papers <- database %>%
110   arrange(desc(TC)) %>%
111   slice_head(n = 10) %>%
112   select(AU, PY, TI, TC) %>%
113   mutate(
114     First_Author = str_split(AU, ";", simplify = TRUE)[,1],
115     Short_Title = str_trunc(TI, 60),
116     Label = paste0(First_Author, " (", PY, ")")
117   )
118
119 g2 <- ggplot(top_papers, aes(y = reorder(Label, TC), x = TC)) +
120   geom_bar(stat = "identity", fill = "coral", alpha = 0.8) +

```

```

121   geom_text(aes(label = TC), hjust = -0.3, size = 4) +
122   labs(title = "Top 10 Artigos Mais Citados",
123        y = "Artigo (Primeiro Autor e Ano)",
124        x = "Citações Totais") +
125   theme_minimal(base_size = 14) +
126   theme(
127     plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
128     axis.text.y = element_text(size = 11),
129     panel.grid.minor = element_blank()
130   ) +
131   scale_x_continuous(expand = expansion(mult = c(0, 0.15)))
132
133   ggsave("fig_top_papers.png", g2, width = 12, height = 7, dpi = 300, bg = "white")
134   cat(" Gráfico salvo: fig_top_papers.png\n")
135
136   # =====
137   # 5. PRODUO CIENTÍFICA POR PAÍS
138   # =====
139
140   # Extrair países dos autores correspondentes
141   countries <- metaTagExtraction(database, Field = "AU_CO", sep = ";")
142   country_prod <- as.data.frame(table(unlist(strsplit(countries$AU_CO, ";"))))
143   colnames(country_prod) <- c("Country", "Freq")
144
145   top_countries <- country_prod %>%
146     filter(Country != "") %>%
147     arrange(desc(Freq)) %>%
148     slice_head(n = 10)
149
150   # Traduzir nomes
151   top_countries <- top_countries %>%
152     mutate(Country_PT = case_when(
153       Country == "USA" ~ "EUA",
154       Country == "UNITED KINGDOM" ~ "Reino Unido",
155       Country == "FRANCE" ~ "França",
156       Country == "CHINA" ~ "China",
157       Country == "BRAZIL" ~ "Brasil",
158       Country == "ITALY" ~ "Itália",
159       Country == "AUSTRALIA" ~ "Austrália",
160       Country == "BELGIUM" ~ "Bélgica",
161       Country == "CANADA" ~ "Canadá",
162       Country == "DENMARK" ~ "Dinamarca",
163       Country == "GERMANY" ~ "Alemanha",
164       Country == "INDIA" ~ "Índia",
165       Country == "SOUTH AFRICA" ~ "frica do Sul",
166       Country == "BOLIVIA" ~ "Bolívia",
167       Country == "SPAIN" ~ "Espanha",
168       Country == "NETHERLANDS" ~ "Holanda",
169       Country == "THAILAND" ~ "Tailândia",
170       TRUE ~ Country
171     ))
172
173   g3 <- ggplot(top_countries, aes(y = reorder(Country_PT, Freq), x = Freq)) +
174     geom_bar(stat = "identity", fill = "darkgreen", alpha = 0.7) +
175     geom_text(aes(label = Freq), hjust = -0.3, size = 4) +
176     labs(title = "Top 10 Países Mais Produtivos",
177          subtitle = "Por país do autor correspondente",
178          y = "País",
179          x = "Número de Artigos") +
180     theme_minimal(base_size = 14) +
181     theme(
182       plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
183       plot.subtitle = element_text(hjust = 0.5, size = 12),
184       panel.grid.minor = element_blank()
185     ) +
186     scale_x_continuous(expand = expansion(mult = c(0, 0.15)))
187

```

```

188 ggsave("fig_countries_production.png", g3, width = 10, height = 6, dpi = 300, bg = "white")
189 cat(" Gráfico salvo: fig_countries_production.png\n")
190
191 # =====
192 # 6. CITAES POR PAÍS (NMERO E MDIA)
193 # =====
194
195 # Extrair países primeiro
196 database_with_countries <- metaTagExtraction(database, Field = "AU_CO", sep = ";")
197
198 # Calcular citações por país
199 countries_citations <- database_with_countries %>%
200   filter(!is.na(AU_CO) & AU_CO != "") %>%
201   mutate(Country = str_split(AU_CO, ";")) %>%
202   unnest(Country) %>%
203   mutate(Country = str_trim(Country)) %>%
204   filter(Country != "") %>%
205   group_by(Country) %>%
206   summarise(
207     Total_Citations = sum(TC, na.rm = TRUE),
208     Avg_Citations = mean(TC, na.rm = TRUE),
209     N_Papers = n()
210   ) %>%
211   arrange(desc(Total_Citations)) %>%
212   slice_head(n = 10)
213
214 # Traduzir
215 countries_citations <- countries_citations %>%
216   mutate(Country_PT = case_when(
217     Country == "USA" ~ "EUA",
218     Country == "UNITED KINGDOM" ~ "Reino Unido",
219     Country == "FRANCE" ~ "França",
220     Country == "CHINA" ~ "China",
221     Country == "BRAZIL" ~ "Brasil",
222     Country == "ITALY" ~ "Itália",
223     Country == "AUSTRALIA" ~ "Austrália",
224     Country == "GERMANY" ~ "Alemanha",
225     Country == "DENMARK" ~ "Dinamarca",
226     Country == "THAILAND" ~ "Tailândia",
227     Country == "SPAIN" ~ "Espanha",
228     Country == "NETHERLANDS" ~ "Holanda",
229     Country == "BELGIUM" ~ "Bélgica",
230     Country == "CANADA" ~ "Canadá",
231     Country == "BOLIVIA" ~ "Bolívia",
232     TRUE ~ Country
233   ))
234
235 # Gráfico com dois eixos Y
236 g4 <- ggplot(countries_citations, aes(x = reorder(Country_PT, Total_Citations))) +
237   geom_bar(aes(y = Total_Citations, fill = "Número de Citações"),
238     stat = "identity", alpha = 0.7) +
239   geom_line(aes(y = Avg_Citations * max(Total_Citations) / max(Avg_Citations),
240     color = "Média de Citações por Artigo", group = 1),
241     linewidth = 1.5) +
242   geom_point(aes(y = Avg_Citations * max(Total_Citations) / max(Avg_Citations)),
243     color = "red", size = 3) +
244   scale_y_continuous(
245     name = "Número de Citações",
246     sec.axis = sec_axis(~ . * max(countries_citations$Avg_Citations) /
247       ↪ max(countries_citations$Total_Citations),
248       name = "Média de Citações por Artigo")
249   ) +
250   scale_fill_manual(values = c("Número de Citações" = "steelblue")) +
251   scale_color_manual(values = c("Média de Citações por Artigo" = "red")) +
252   labs(x = "País", title = "Citações por País") +
253   theme_minimal(base_size = 14) +
254   theme(

```

```

254     plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
255     axis.text.x = element_text(angle = 45, hjust = 1, size = 12),
256     axis.title.y.left = element_text(color = "steelblue", face = "bold"),
257     axis.title.y.right = element_text(color = "red", face = "bold"),
258     legend.position = "top",
259     legend.title = element_blank(),
260     panel.grid.minor = element_blank()
261 )
262
263 ggsave("fig_countries_citations.png", g4, width = 12, height = 7, dpi = 300, bg = "white")
264 cat(" Gráfico salvo: fig_countries_citations.png\n")
265
266 # =====
267 # 7. AFILIAES MAIS PRODUTIVAS
268 # =====
269
270 affiliations_df <- as.data.frame(Res$Affiliations) %>%
271   setNames(c("Affiliation", "Articles")) %>%
272   mutate(Affiliation = case_when(
273     grepl("UNIVERSITY OF CALIFORNIA BERKELEY", Affiliation, ignore.case = TRUE) ~ "UNIVERSITY
274       ↳ OF CALIFORNIA",
275     grepl("UNIVERSITY OF CALIFORNIA SYSTEM", Affiliation, ignore.case = TRUE) ~ "UNIVERSITY
276       ↳ OF CALIFORNIA",
277     grepl("UNIVERSITE TOULOUSE 1 CAPITOLE", Affiliation, ignore.case = TRUE) ~ "UNIVERSITE DE
278       ↳ TOULOUSE",
279     TRUE ~ Affiliation
280   )) %>%
281   filter(Affiliation != "NOTREPORTED") %>%
282   group_by(Affiliation) %>%
283   summarise(Articles = sum(Articles)) %>%
284   arrange(desc(Articles)) %>%
285   slice_head(n = 10)
286
287 g5 <- ggplot(affiliations_df, aes(y = reorder(Affiliation, Articles), x = Articles)) +
288   geom_bar(stat = "identity", fill = "purple", alpha = 0.7) +
289   geom_text(aes(label = Articles), hjust = -0.3, size = 4) +
290   labs(title = "Top 10 Afiliações Mais Produtivas",
291        y = "Afiliação",
292        x = "Número de Publicações") +
293   theme_minimal(base_size = 14) +
294   theme(
295     plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
296     axis.text.y = element_text(size = 11),
297     panel.grid.minor = element_blank()
298   ) +
299   scale_x_continuous(expand = expansion(mult = c(0, 0.15)))
300
301 ggsave("fig_affiliations.png", g5, width = 12, height = 7, dpi = 300, bg = "white")
302 cat(" Gráfico salvo: fig_affiliations.png\n")
303
304 # =====
305 # 8. REFERÊNCIAS MAIS CITADAS
306 # =====
307
308 topCited <- citations(database, field = "article", sep = ";")
309
310 # Criar dataframe das top referências
311 top_refs <- data.frame(
312   Reference = names(topCited$Cited[1:11]),
313   Citations = as.numeric(topCited$Cited[1:11])
314 )
315
316 cat("\n=====")
317 cat("\n  TOP 10 REFERÊNCIAS MAIS CITADAS")
318 cat("\n=====")
319 print(top_refs, row.names = FALSE)
320

```

```

318 # Gráfico das referências mais citadas (opcional)
319 top_refs <- top_refs %>%
320   mutate(Short_Ref = str_trunc(Reference, 60))
321
322 g_refs <- ggplot(top_refs, aes(y = reorder(Short_Ref, Citations), x = Citations)) +
323   geom_bar(stat = "identity", fill = "darkorange", alpha = 0.7) +
324   geom_text(aes(label = Citations), hjust = -0.3, size = 3.5) +
325   labs(title = "Top 10 Referências Mais Citadas",
326        y = "Referência",
327        x = "Número de Citações") +
328   theme_minimal(base_size = 12) +
329   theme(
330     plot.title = element_text(face = "bold", hjust = 0.5, size = 14),
331     axis.text.y = element_text(size = 10),
332     panel.grid.minor = element_blank()
333   ) +
334   scale_x_continuous(expand = expansion(mult = c(0, 0.15)))
335
336 ggsave("fig_top_references.png", g_refs, width = 12, height = 7, dpi = 300, bg = "white")
337 cat(" Gráfico salvo: fig_top_references.png\n")
338
339 # =====
340 # 9. EVOLUO TEMPORAL DA PRODUO
341 # =====
342
343 prod_year <- database %>%
344   group_by(PY) %>%
345   summarise(N_Papers = n()) %>%
346   filter(!is.na(PY))
347
348 g6 <- ggplot(prod_year, aes(x = PY, y = N_Papers)) +
349   geom_line(color = "darkblue", linewidth = 1.2) +
350   geom_point(color = "darkblue", size = 2.5) +
351   geom_smooth(method = "loess", se = TRUE, color = "red", linetype = "dashed", alpha = 0.2) +
352   labs(title = "Evolução Temporal da Produção Científica",
353        x = "Ano de Publicação",
354        y = "Número de Artigos") +
355   theme_minimal(base_size = 14) +
356   theme(
357     plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
358     panel.grid.minor = element_blank()
359   ) +
360   scale_x_continuous(breaks = seq(min(prod_year$PY), max(prod_year$PY), by = 5))
361
362 ggsave("fig_production_evolution.png", g6, width = 12, height = 6, dpi = 300, bg = "white")
363 cat(" Gráfico salvo: fig_production_evolution.png\n")
364
365 # =====
366 # 10. NUVEM DE PALAVRAS-CHAVE
367 # =====
368
369 # Extrair palavras-chave
370 keywords <- database %>%
371   select(DE, ID) %>%
372   mutate(Keywords = paste(DE, ID, sep = "; ")) %>%
373   pull(Keywords) %>%
374   str_split("; ") %>%
375   unlist() %>%
376   str_to_upper() %>%
377   str_trim() %>%
378   .[. != "" & !is.na(.)]
379
380 keyword_freq <- as.data.frame(table(keywords)) %>%
381   setNames(c("Keyword", "Freq")) %>%
382   arrange(desc(Freq)) %>%
383   slice_head(n = 30)
384

```

```

385 g7 <- ggplot(keyword_freq, aes(y = reorder(Keyword, Freq), x = Freq)) +
386   geom_bar(stat = "identity", fill = "orange", alpha = 0.7) +
387   labs(title = "Top 30 Palavras-Chave Mais Frequentes",
388        y = "Palavra-Chave",
389        x = "Frequência") +
390   theme_minimal(base_size = 12) +
391   theme(
392     plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
393     axis.text.y = element_text(size = 10),
394     panel.grid.minor = element_blank()
395   )
396
397 ggsave("fig_keywords.png", g7, width = 10, height = 10, dpi = 300, bg = "white")
398 cat(" Gráfico salvo: fig_keywords.png\n")
399
400 # =====
401 # 11. EVOLUO TEMTICA DOS RESUMOS (FIGURA 3 - MAIS IMPORTANTE)
402 # =====
403
404 biblioshiny()
405
406 # =====
407 # 11b. TREND TOPICS - TERMOS MAIS RELEVANTES AO LONGO DO TEMPO
408 # =====
409
410 cat("\n>>> Gerando análise de Trend Topics...\n")
411
412 # Análise de tendências de palavras-chave
413 trend_topics <- fieldByYear(database,
414                             field = "ID", # Keywords Plus
415                             timespan = c(2014,2025),
416                             min.freq = 5, # Frequência mínima
417                             n.items = 1, # Top 10 termos
418                             labelsiz = 4,
419                             graph = TRUE)
420
421 # Salvar gráfico gerado automaticamente
422 if(dev.cur() > 1) {
423   dev.copy(png, "fig_trend_topics_auto.png", width = 3500, height = 2500, res = 300)
424   dev.off()
425   cat(" Gráfico salvo: fig_trend_topics_auto.png\n")
426 }
427
428
429
430 # =====
431 # 12. REDE DE CO-AUTORIA
432 # =====
433
434 NetMatrix <- biblioNetwork(database, analysis = "collaboration", network = "authors", sep =
435   ↪ ";")
436
437 net <- networkPlot(NetMatrix, n = 30, Title = "Rede de Co-autoria",
438                   type = "fruchterman", size = TRUE, remove.multiple = FALSE,
439                   labelsiz = 0.7, cluster = "walktrap")
440
441 # Salvar rede
442 png("fig_coauthorship_network.png", width = 3000, height = 3000, res = 300)
443 networkPlot(NetMatrix, n = 30, Title = "Rede de Co-autoria",
444             type = "fruchterman", size = TRUE, remove.multiple = FALSE,
445             labelsiz = 0.7, cluster = "walktrap")
446 dev.off()
447 cat(" Gráfico salvo: fig_coauthorship_network.png\n")
448
449 # =====
450 # 13. ANLISE DE CO-CITAO
451 # =====

```

```

451 NetMatrix_cocit <- biblioNetwork(database, analysis = "co-citation", network = "references",
    ↪ sep = ";")
452 net_cocit <- networkPlot(NetMatrix_cocit, n = 30, Title = "Rede de Co-citação",
453                          type = "fruchterman", size = TRUE, remove.multiple = FALSE,
454                          labelsizes = 0.6, cluster = "walktrap")
455
456 png("fig_cocitation_network.png", width = 3000, height = 3000, res = 300)
457 networkPlot(NetMatrix_cocit, n = 30, Title = "Rede de Co-citação",
458             type = "fruchterman", size = TRUE, remove.multiple = FALSE,
459             labelsizes = 0.6, cluster = "walktrap")
460 dev.off()
461 cat(" Gráfico salvo: fig_cocitation_network.png\n")
462
463 # =====
464 # RESUMO FINAL
465 # =====
466
467 cat("\n=====")
468 cat("\n  ANLISE COMPLETA FINALIZADA")
469 cat("\n=====\\n")
470 cat("Total de gráficos gerados: 15\\n")
471 cat("  - Análises descritivas: 7\\n")
472 cat("  - Evolução temática: 2\\n")
473 cat("  - Trend Topics: 4\\n")
474 cat("  - Redes: 2\\n\\n")
475 cat("Todos os arquivos foram salvos no diretório de trabalho.\\n\\n")
476
477 # Lista de arquivos gerados
478 cat("Arquivos gerados:\\n")
479 cat("  1. fig_top_authors.png\\n")
480 cat("  2. fig_top_papers.png\\n")
481 cat("  3. fig_countries_production.png\\n")
482 cat("  4. fig_countries_citations.png\\n")
483 cat("  5. fig_affiliations.png\\n")
484 cat("  6. fig_production_evolution.png\\n")
485 cat("  7. fig_keywords.png\\n")
486 cat("  8. fig_thematic_evolution.png (FIGURA 3 - PRINCIPAL)\\n")
487 cat("  9. fig_keywords_evolution.png\\n")
488 cat(" 10. fig_trend_topics_auto.png\\n")
489 cat(" 11. fig_trend_topics_lines.png (TREND TOPICS)\\n")
490 cat(" 12. fig_trend_topics_area.png (TREND TOPICS)\\n")
491 cat(" 13. fig_trend_topics_heatmap.png (TREND TOPICS)\\n")
492 cat(" 14. fig_coauthorship_network.png\\n")
493 cat(" 15. fig_cocitation_network.png\\n\\n")
494
495 # Gerar relatório em texto
496 sink("relatorio_bibliometrico.txt")
497 cat("=====\\n")
498 cat("RELATRIO BIBLIOMTRICO - SANEAMENTO\\n")
499 cat("=====\\n\\n")
500 cat("Data da análise:", format(Sys.Date(), "%d/%m/%Y"), "\\n\\n")
501 print(DS)
502 cat("\\n\\n=====\\n")
503 cat("ANLISES ADICIONAIS\\n")
504 cat("=====\\n\\n")
505 cat("Evolução Temática: fig_thematic_evolution.png\\n")
506 cat("Trend Topics: fig_trend_topics_*.png\\n")
507 cat("Redes de colaboração e citação geradas.\\n")
508 sink()
509
510 cat(" Relatório salvo: relatorio_bibliometrico.txt\\n\\n")
511 cat("=====\\n")
512 cat("  Análise bibliométrica completa!\\n")
513 cat("=====\\n")

```

B.2 Fronteira Estocástica

Código B.2 – Fronteira Estocástica

```

1  # =====
2  # ANLISE SFA - SETOR DE GUA BRASIL (SNIS 2015-2022)
3  # =====
4  # Análise de Eficiência Técnica usando Stochastic Frontier Analysis
5  # Período: 2015-2022 (T = 8 anos)
6  # Fonte: SNIS Agregado
7  #
8  # ESPECIFICAO DA FUNO CUSTO (Kumbhakar & Lovell 2003, Cap. 2 e 4):
9  #
10 #    $vc(y, w, z) = \min \{ w'x : (y, x, z) \text{ GR } \}$ 
11 #
12 #   CUSTO (VC):   FN015
13 #   FN015 = despesas de exploração (pessoal, energia, químicos,
14 #   serviços de terceiros, água importada, esgoto exportado)
15 #   Exclui FN016 (depreciação) e FN019 (despesas financeiras)
16 #
17 #   OUTPUTS (volumes em 1.000 m/ano):
18 #   gua   = AG006
19 #   Esgoto= ES005 + ES006 + ES012
20 #
21 #   PREOS DOS INSUMOS VARIVEIS (normalizados por P_E):
22 #   P_L = FN010 / FN026
23 #   P_E = FN013 / (AG028 + ES028)   numerário
24 #   P_O = (VC FN010 FN013) / (AG005 + ES004)
25 #
26 #   CAPITAL QUASI-FIXO (z):
27 #   K = AG005 + ES004   (extensão total da rede, km)
28 # =====
29
30 # 1. CONFIGURAO INICIAL ----
31
32 rm(list = ls())
33 gc()
34
35 options(scipen = 999)
36
37 library(tidyverse)
38 library(readr)
39 library(frontier)
40 library(plm)
41 library(kableExtra)
42
43 # 2. DEFINIR DIRETRIOS ----
44
45 dir_projeto <- "C:/Users/Pedro Ricelly/Google Drive/Doutorado/Tese/Artigo 3/dados"
46 dir_outputs <- "C:/Users/Pedro Ricelly/Google Drive/Doutorado/Tese/Artigo 3/outputs"
47
48 if(!dir.exists(dir_outputs)) dir.create(dir_outputs, recursive = TRUE)
49
50 # 3. FUNO AUXILIAR ----
51
52 para_numerico <- function(x) {
53   x <- gsub("\\.", "", x)
54   x <- gsub(",", ".", x)
55   as.numeric(x)
56 }
57
58 # 4. CARREGAR DADOS SNIS ----
59
60 arquivo_csv <- file.path(dir_projeto, "Agregado-20260108185004.csv")
61 Sys.setenv("VROOM_CONNECTION_SIZE" = 2097152)
62

```

```

63 dados_snis <- read_delim(
64   arquivo_csv,
65   delim = ";",
66   locale = locale(encoding = "UTF-16LE"),
67   show_col_types = FALSE,
68   guess_max = 5000
69 )
70
71 dados_snis <- dados_snis[-nrow(dados_snis), ]
72
73 # 5. CONSTRUIR VARIVEIS (Kumbhakar & Lovell 2003) ----
74
75 dados_sfa <- dados_snis %>%
76   select(
77     cod_municipio = `Código do Município`,
78     municipio = `Município`,
79     estado = `Estado`,
80     ano = `Ano de Referência`,
81     cod_prestador = `Código do Prestador`,
82     prestador = `Prestador`,
83     abrangencia = `Abrangência`,
84     tipo_servico = `Tipo de Serviço`,
85     natureza_juridica = `Natureza Jurídica`,
86     # Custos
87     fn010 = `FN010 - Despesa com pessoal próprio`,
88     fn013 = `FN013 - Despesa com energia elétrica`,
89     fn015 = `FN015 - Despesas de Exploração (DEX), sendo FN015 = FN010 + FN011 + FN013 +
90     ↪ FN014 + FN020 + FN039 + FN021 + FN027`,
91     fn026 = `FN026 - Quantidade total de empregados próprios`,
92     # Outputs água (volumes)
93     ag006 = `AG006 - Volume de água produzido`,
94     # Outputs esgoto (volumes)
95     es005 = `ES005 - Volume de esgotos coletado`,
96     es006 = `ES006 - Volume de esgotos tratado`,
97     es012 = `ES012 - Volume de esgoto bruto exportado`,
98     # Capital quasi-fixo
99     ag005 = `AG005 - Extensão da rede de água`,
100    es004 = `ES004 - Extensão da rede de esgotos`,
101    # Energia
102    ag028 = `AG028 - Consumo total de energia elétrica nos sistemas de água`,
103    es028 = `ES028 - Consumo total de energia elétrica nos sistemas de esgotos`,
104    # Controles
105    ag001 = `AG001 - População total atendida com abastecimento de água`,
106    ag002 = `AG002 - Quantidade de ligações ativas de água`,
107    ag004 = `AG004 - Quantidade de ligações ativas de água micromedidas`,
108    es026 = `ES026 - População urbana atendida com esgotamento sanitário`,
109    ge06b = `G06B - População urbana residente do(s) município(s) com esgotamento sanitário`,
110    in013 = `IN013 - Índice de perdas faturamento`,
111    # Auxiliares para Treatment
112    es013 = `ES013 - Volume de esgotos bruto importado`,
113    es014 = `ES014 - Volume de esgoto importado tratado nas instalações do importador`,
114    es015 = `ES015 - Volume de esgoto bruto exportado tratado nas instalações do importador`
115  ) %>%
116  filter(!is.na(ano), ano != "---") %>%
117  mutate(across(fn010:es015, para_numerico))
118
119 # 5.1 DEFLACIONAR VARIVEIS MONETRIAS (IPCA, base = dezembro/2015) ----
120
121 library(deflateBR)
122
123 # deflateBR exige coluna de data no formato Date
124 # Usando dezembro de cada ano como mês de referência
125 dados_sfa <- dados_sfa %>%
126   mutate(data_ref = as.Date(paste0(ano, "-12-01")))
127
128 # Deflacionar variáveis monetárias individualmente
129 dados_sfa <- dados_sfa %>%

```

```

129   mutate(
130     fn010 = ipca(fn010, data_ref, "12/2015"),
131     fn013 = ipca(fn013, data_ref, "12/2015"),
132     fn015 = ipca(fn015, data_ref, "12/2015")
133   )
134
135   cat(" Variáveis monetárias deflacionadas (IPCA, base = dezembro/2015)\n")
136
137   dados_sfa %>%
138     group_by(ano) %>%
139     summarise(fn015_medio = mean(fn015, na.rm = TRUE)) %>%
140     print()
141
142   # 6. CRIAR VARIVEIS DERIVADAS ----
143
144   dados_sfa <- dados_sfa %>%
145     mutate(
146       VC          = fn015,
147       # Y_agua: volume produzido
148       Y_agua      = ag006,
149       # Y_esgo: volume coletado + tratado + exportado bruto
150       # Corresponde a ES005 + ES006 + ES012 (cf. Tab. de variáveis do artigo)
151       Y_esgo      = rowSums(cbind(es005, es006, es012), na.rm = FALSE),
152       K           = ag005 + es004,
153       P_l         = if_else(fn026 > 0, fn010 / fn026, NA_real_),
154       P_e         = if_else((ag028 + es028) > 0,
155                             fn013 / (ag028 + es028), NA_real_),
156       custo_outros = VC - fn010 - fn013,
157       P_o         = if_else((ag005 + es004) > 0,
158                             (VC - fn010 - fn013) / (ag005 + es004), NA_real_),
159       Dens        = if_else(ag005 > 0, ag002 / ag005, NA_real_),
160       Micro        = if_else(ag002 > 0, ag004 / ag002, NA_real_),
161       Loss         = in013,
162       Sewerage     = if_else(ge06b > 0, es026 / ge06b * 100, NA_real_),
163       Treatment_prop = if_else((es005 + es013) > 0,
164                                (es006 + es014 + es015) / (es005 + es013),
165                                NA_real_),
166       Treatment    = if_else(!is.na(Treatment_prop) & Treatment_prop > 0.20, 1, 0),
167       Independent  = if_else(natureza_juridica %in%
168                             c("Administração pública direta",
169                               "rgão da administração direta"), 0, 1),
170       Private      = if_else(grepl("Privada|privada|PRIVADA", natureza_juridica), 1, 0),
171       Local        = if_else(abrangencia == "Local", 1, 0),
172       WS           = if_else(str_trim(tipo_servico) == "gua e Esgoto", 1, 0),
173       NO           = if_else(estado %in% c("AC", "AM", "AP", "PA", "RO", "RR", "TO"), 1, 0),
174       NE           = if_else(estado %in% c("AL", "BA", "CE", "MA", "PB", "PE", "PI", "RN", "SE"), 1, 0),
175       SE           = if_else(estado %in% c("ES", "MG", "RJ", "SP"), 1, 0),
176       SO           = if_else(estado %in% c("PR", "RS", "SC"), 1, 0),
177       MI           = if_else(estado %in% c("DF", "GO", "MS", "MT"), 1, 0)
178     )
179
180   # 7. LIMPAR DADOS ----
181
182   dados_sfa_clean <- dados_sfa %>%
183     filter(
184       !is.na(VC),      VC > 0,
185       !is.na(Y_agua),  Y_agua > 0,
186       !is.na(Y_esgo),  Y_esgo > 0,
187       !is.na(K),       K > 0,
188       !is.na(P_l),     P_l > 0,
189       !is.na(P_e),     P_e > 0,
190       !is.na(P_o),     P_o > 0,
191       custo_outros > 0,
192       !is.na(Dens),    Dens > 0,
193       !is.na(Micro),   Micro > 0,
194       Loss >= 0, Loss < 100
195     )

```

```

196
197 # 8. CRIAR DUMMY PARA MUNICÍPIOS COM MLTIPLS PRESTADORES ----
198
199 municipios_multiplos <- dados_sfa_clean %>%
200   group_by(ano, cod_municipio) %>%
201   filter(n_distinct(cod_prestador) > 1) %>%
202   distinct(cod_municipio) %>%
203   pull(cod_municipio)
204
205 dados_sfa_final <- dados_sfa_clean %>%
206   mutate(
207     Multi = if_else(cod_municipio %in% unique(municipios_multiplos), 1, 0)
208   )
209
210 cat("Municípios com múltiplos prestadores:", length(unique(municipios_multiplos)), "\n")
211 cat("Observações com Multi = 1:", sum(dados_sfa_final$Multi), "\n")
212 cat("Observações com Multi = 0:", sum(dados_sfa_final$Multi == 0), "\n")
213 cat("Total de observações:", nrow(dados_sfa_final), "\n")
214
215 # 9. IMPUTAR SEWERAGE PARA 2022 ----
216
217 cat("\n=== IMPUTAO DE SEWERAGE (ANO 2022) ===\n")
218
219 cat("Observações de 2022 com Sewerage NA antes da imputação:",
220     sum(dados_sfa_final$ano == "2022" & is.na(dados_sfa_final$Sewerage)), "\n")
221
222 # Método 1: LOCF valor de 2021
223 dados_sfa_final <- dados_sfa_final %>%
224   arrange(cod_prestador, ano) %>%
225   group_by(cod_prestador) %>%
226   mutate(
227     Sewerage_original = Sewerage,
228     Sewerage_imputado = if_else(
229       ano == "2022" & is.na(Sewerage),
230       lag(Sewerage, order_by = ano),
231       Sewerage
232     ),
233     metodo_imputacao = if_else(
234       ano == "2022" & is.na(Sewerage) & !is.na(lag(Sewerage, order_by = ano)),
235       "LOCF_2021", NA_character_
236     )
237   ) %>%
238   ungroup()
239
240 # Método 2: média do prestador
241 dados_sfa_final <- dados_sfa_final %>%
242   group_by(cod_prestador) %>%
243   mutate(
244     Sewerage_media_prestador = mean(Sewerage_original, na.rm = TRUE),
245     Sewerage_imputado = if_else(
246       ano == "2022" & is.na(Sewerage_imputado) & !is.na(Sewerage_media_prestador),
247       Sewerage_media_prestador, Sewerage_imputado
248     ),
249     metodo_imputacao = if_else(
250       ano == "2022" & is.na(metodo_imputacao) &
251       !is.na(Sewerage_media_prestador) & is.na(Sewerage_original),
252       "Media_Prestador", metodo_imputacao
253     )
254   ) %>%
255   ungroup()
256
257 # Método 3: média regional
258 dados_sfa_final <- dados_sfa_final %>%
259   mutate(regiao = case_when(
260     NO == 1 ~ "Norte",    NE == 1 ~ "Nordeste",
261     SE == 1 ~ "Sudeste",  SO == 1 ~ "Sul",
262     MI == 1 ~ "Centro-Oeste"

```

```

263   )) %>%
264   group_by(regiao, ano) %>%
265   mutate(Sewerage_media_regiao = mean(Sewerage_original, na.rm = TRUE)) %>%
266   ungroup() %>%
267   mutate(
268     Sewerage_imputado = if_else(
269       ano == "2022" & is.na(Sewerage_imputado) & !is.na(Sewerage_media_regiao),
270       Sewerage_media_regiao, Sewerage_imputado
271     ),
272     metodo_imputacao = if_else(
273       ano == "2022" & is.na(metodo_imputacao) &
274       !is.na(Sewerage_media_regiao) & is.na(Sewerage_original),
275       "Media_Regional", metodo_imputacao
276     )
277   )
278
279 dados_sfa_final <- dados_sfa_final %>%
280   mutate(Sewerage = Sewerage_imputado)
281
282 log_imputacao <- dados_sfa_final %>%
283   filter(ano == "2022" & !is.na(metodo_imputacao)) %>%
284   select(cod_prestador, prestador, estado, regiao,
285     Sewerage_original, Sewerage_imputado, metodo_imputacao)
286
287 write_csv(log_imputacao, file.path(dir_outputs, "log_imputacao_sewerage_2022.csv"))
288
289 dados_sfa_final <- dados_sfa_final %>%
290   select(-Sewerage_original, -Sewerage_imputado, -metodo_imputacao,
291     -Sewerage_media_prestador, -Sewerage_media_regiao, -regiao)
292
293 # 10. CRIAR PAINEL BALANCEADO ----
294
295 n_anos_completo <- dados_sfa_final %>%
296   count(cod_prestador) %>%
297   pull(n) %>%
298   max()
299
300 dados_sfa_painel <- dados_sfa_final %>%
301   group_by(cod_prestador) %>%
302   filter(n() == n_anos_completo) %>%
303   ungroup() %>%
304   arrange(cod_prestador, ano)
305
306 cat("\n=== PAINEL BALANCEADO (INICIAL) ===\n")
307 cat("Observações:", nrow(dados_sfa_painel), "\n")
308 cat("Prestadores:", n_distinct(dados_sfa_painel$cod_prestador), "\n")
309 cat("Anos:", paste(sort(unique(dados_sfa_painel$ano)), collapse = ", "), "\n")
310
311 # 11. ANLISE DE COBERTURA POPULACIONAL ----
312
313 pop_brasil <- data.frame(
314   ano = as.character(2015:2022),
315   pop_total_br = c(204450649, 206081432, 207660929, 208494900,
316     209469323, 210147125, 211755692, 213196304)
317 )
318
319 cobertura_pop <- dados_sfa_painel %>%
320   group_by(ano) %>%
321   summarise(
322     pop_agua_amostra = sum(ag001, na.rm = TRUE),
323     n_prestadores = n_distinct(cod_prestador),
324     .groups = "drop"
325   ) %>%
326   left_join(pop_brasil, by = "ano") %>%
327   mutate(perc_cobertura = (pop_agua_amostra / pop_total_br) * 100)
328
329 cat("\n=== COBERTURA POPULACIONAL (PAINEL INICIAL) ===\n")

```

```

330 print(cobertura_pop)
331
332 # 12. ESTATÍSTICAS DESCRITIVAS (TABELAS 2, 3 E 4) ----
333
334 cat("\n=== GERANDO ESTATÍSTICAS DESCRITIVAS ===\n")
335
336 # Função auxiliar: descritivas por grupo
337 descritivas_grupo <- function(df, sufixo) {
338   df %>%
339     select(VC, Y_agua, Y_esgo, K, P_l, P_e, P_o, Dens, Micro, Loss) %>%
340     summarise(across(everything(), list(
341       Media = ~mean(., na.rm = TRUE),
342       DP = ~sd(., na.rm = TRUE),
343       Min = ~min(., na.rm = TRUE),
344       Max = ~max(., na.rm = TRUE)
345     ), .names = "{.col}_{.fn}")) %>%
346     pivot_longer(everything(),
347       names_to = c("Variável", "Estatística"),
348       names_pattern = "(.+)_(.+)") %>%
349     pivot_wider(names_from = Estatística, values_from = value) %>%
350     rename_with(~paste0(., sufixo), c(Media, DP, Min, Max))
351 }
352
353 rotulos_variaveis <- tibble(
354   Variável = c("VC", "Y_agua", "Y_esgo", "K", "P_l", "P_e", "P_o", "Dens", "Micro", "Loss"),
355   Label = c(
356     "Custo Variável (R\\$ 1.000)",
357     "Produto gua (1.000 m/ano)",
358     "Produto Esgoto (1.000 m/ano)",
359     "Extensão da Rede quasi-fixo (km)",
360     "Preço do Trabalho (R\\$/empregado)",
361     "Preço da Energia (R\\$/kWh)",
362     "Preço de Outros Insumos (R\\$/km)",
363     "Densidade da Rede (lig/km)",
364     "Micromedicação (\\%)",
365     "Índice de Perdas (\\%)"
366   )
367 )
368
369 desc_privado <- descritivas_grupo(dados_sfa_painel %>% filter(Private == 1), "_priv")
370 desc_naoprivado <- descritivas_grupo(dados_sfa_painel %>% filter(Private == 0), "_npri")
371
372 n_priv <- sum(dados_sfa_painel$Private == 1)
373 n_npri <- sum(dados_sfa_painel$Private == 0)
374
375 variaveis_desc_long <- desc_privado %>%
376   left_join(desc_naoprivado, by = "Variável") %>%
377   left_join(rotulos_variaveis, by = "Variável") %>%
378   select(Label,
379     Media_priv, DP_priv, Min_priv, Max_priv,
380     Media_npri, DP_npri, Min_npri, Max_npri) %>%
381   rename(Variável = Label)
382
383 write_csv(variaveis_desc_long,
384   file.path(dir_outputs, "tabela_2_descritivas_originais.csv"))
385
386 tabela_2_latex <- kable(
387   variaveis_desc_long,
388   format = "latex", booktabs = TRUE, digits = 2, escape = FALSE,
389   col.names = c("Variável",
390     "Média", "DP", "Mín.", "Máx.",
391     "Média", "DP", "Mín.", "Máx."),
392   caption = "Estatísticas Descritivas das Variáveis por Tipo de Prestador (Valores
393     ↪ Originais)",
394   label = "desc_stats_original",
395   align = c("l", rep("r", 8))
396 ) %>%

```

```

396 kable_styling(latex_options = c("hold_position", "scale_down"), font_size = 9) %>%
397 add_header_above(c(
398   " "                               = 1,
399   setNames(4, sprintf("Privado (N = %d)", n_priv)),
400   setNames(4, sprintf("Não Privado (N = %d)", n_npri))
401 )) %>%
402 footnote(
403   general = paste0(
404     "Custo Variável = FN015 (deflacionado, IPCA, base dez/2015). ",
405     "Produtos em 1.000 m\\\\"textsuperscript{3}/ano. ",
406     "Extensão da Rede (K) é o capital quasi-fixo. ",
407     "Total de observações: ", nrow(dados_sfa_painel), " (",
408     n_distinct(dados_sfa_painel$cod_prestador), " prestadores, ",
409     n_distinct(dados_sfa_painel$ano), " anos, 2015--2022).",
410   ),
411   general_title = "Nota:", footnote_as_chunk = TRUE, escape = FALSE
412 )
413
414 writeLines(as.character(tabela_2_latex),
415   file.path(dir_outputs, "tabela_2_descritivas_originais.tex"))
416 cat(" Tabela 2 (Descritivas Originais Privado vs Não Privado) salva\n")
417 cat("\n=== GERANDO ESTATÍSTICAS DESCRITIVAS ===\n")
418
419 # TABELA 2: Valores Originais em português
420 variaveis_desc <- dados_sfa_painel %>%
421   select(VC, Y_agua, Y_esgo, K, P_l, P_e, P_o, Dens, Micro, Loss) %>%
422   summarise(across(everything(), list(
423     Media = ~mean(., na.rm = TRUE),
424     DP     = ~sd(., na.rm = TRUE),
425     Min    = ~min(., na.rm = TRUE),
426     Max    = ~max(., na.rm = TRUE)
427   ), .names = "{.col}_{.fn}"))
428
429 variaveis_desc_long <- variaveis_desc %>%
430   pivot_longer(everything(),
431     names_to = c("Variável", "Estatística"),
432     names_pattern = "(.)_(.)" %>%
433   pivot_wider(names_from = Estatística, values_from = value) %>%
434   mutate(Variável = case_when(
435     Variável == "VC" ~ "Custo Variável (R$ 1.000)",
436     Variável == "Y_agua" ~ "Produto gua (1.000 m/ano)",
437     Variável == "Y_esgo" ~ "Produto Esgoto (1.000 m/ano)",
438     Variável == "K" ~ "Extensão da Rede quasi-fixo (km)",
439     Variável == "P_l" ~ "Preço do Trabalho (R$/empregado)",
440     Variável == "P_e" ~ "Preço da Energia (R$/kWh)",
441     Variável == "P_o" ~ "Preço de Outros Insumos (R$/km)",
442     Variável == "Dens" ~ "Densidade da Rede (lig/km)",
443     Variável == "Micro" ~ "Micromedição (%)",
444     Variável == "Loss" ~ "Índice de Perdas (%)",
445     TRUE ~ Variável
446   ))
447
448 write_csv(variaveis_desc_long,
449   file.path(dir_outputs, "tabela_2_descritivas_originais.csv"))
450
451 tabela_2_latex <- kable(
452   variaveis_desc_long,
453   format = "latex", booktabs = TRUE, digits = 2,
454   col.names = c("Variável", "Média", "DP", "Mín.", "Máx."),
455   caption = "Estatísticas Descritivas das Variáveis (Valores Originais)",
456   label = "desc_stats_original"
457 ) %>%
458 kable_styling(latex_options = c("hold_position")) %>%
459 footnote(
460   general = paste0(
461     "N = ", nrow(dados_sfa_painel), " observações de ",
462     n_distinct(dados_sfa_painel$cod_prestador), " prestadores ao longo de ",

```

```

463     n_distinct(dados_sfa_painel$ano), " anos (2015-2022). ",
464     "Custo Variável = FN015. Produtos em 1.000 m\\\\"textsuperscript{3}/ano. ",
465     "Extensão da Rede (K) é o capital quasi-fixo."
466   ),
467   general_title = "Nota:", footnote_as_chunk = TRUE, escape = FALSE
468 )
469
470 writeLines(as.character(tabela_2_latex),
471           file.path(dir_outputs, "tabela_2_descritivas_originais.tex"))
472 cat(" Tabela 2 (Descritivas Originais) salva\n")
473
474 # TABELA 3: Descritivas Log Privado vs Não Privado
475 descritivas_log_grupo <- function(df, sufixo) {
476   df %>%
477     select(lnVC, lnY_agua, lnY_esgo, lnK, lnP_l, lnP_o, lnDens, lnMicro, Loss) %>%
478     summarise(across(everything(), list(
479       Media = ~mean(., na.rm = TRUE),
480       DP     = ~sd(., na.rm = TRUE),
481       Min    = ~min(., na.rm = TRUE),
482       Max    = ~max(., na.rm = TRUE)
483     )), .names = "{.col}_{.fn}") %>%
484     pivot_longer(everything(),
485                 names_to = c("Variável", "Estatística"),
486                 names_pattern = "(.+)_(.+)") %>%
487     pivot_wider(names_from = Estatística, values_from = value) %>%
488     rename_with(~paste0(., sufixo), c(Media, DP, Min, Max))
489 }
490
491 rotulos_log <- tibble(
492   Variável = c("lnVC", "lnY_agua", "lnY_esgo", "lnK", "lnP_l", "lnP_o", "lnDens", "lnMicro", "Loss"),
493   Label    = c(
494     "$\\ln(CV / P_E)$",
495     "$\\ln(Y_{\\text{água}})$",
496     "$\\ln(Y_{\\text{esgoto}})$",
497     "$\\ln(K)$",
498     "$\\ln(P_L / P_E)$",
499     "$\\ln(P_O / P_E)$",
500     "$\\ln(\\text{Densidade})$",
501     "$\\ln(\\text{Micromedicação})$",
502     "Índice de Perdas (\\%) "
503   )
504 )
505
506 log_privado <- descritivas_log_grupo(dados_sfa_painel %>% filter(Private == 1), "_priv")
507 log_naoprivado <- descritivas_log_grupo(dados_sfa_painel %>% filter(Private == 0), "_npri")
508
509 variaveis_log_long <- log_privado %>%
510   left_join(log_naoprivado, by = "Variável") %>%
511   left_join(rotulos_log, by = "Variável") %>%
512   select(Label,
513          Media_priv, DP_priv, Min_priv, Max_priv,
514          Media_npri, DP_npri, Min_npri, Max_npri) %>%
515   rename(Variável = Label)
516
517 write_csv(variaveis_log_long,
518           file.path(dir_outputs, "tabela_3_descritivas_log.csv"))
519
520 tabela_3_latex <- kable(
521   variaveis_log_long,
522   format = "latex", booktabs = TRUE, digits = 4, escape = FALSE,
523   col.names = c("Variável",
524                 "Média", "DP", "Mín.", "Máx.",
525                 "Média", "DP", "Mín.", "Máx."),
526   caption = "Estatísticas Descritivas das Variáveis por Tipo de Prestador (Transformação
527     ↪ Logarítmica)",
527   label = "desc_stats_log",
528   align = c("l", rep("r", 8))

```

```

529 ) %>%
530   kable_styling(latex_options = c("hold_position", "scale_down"), font_size = 9) %>%
531   add_header_above(c(
532     " "                               = 1,
533     setNames(4, sprintf("Privado (N = %d)", n_priv)),
534     setNames(4, sprintf("Não Privado (N = %d)", n_npri))
535   )) %>%
536   footnote(
537     general = c(
538       "Todas as variáveis, exceto o índice de Perdas, estão em forma logarítmica.",
539       "Custos e preços normalizados pelo preço da energia ($P_E$) como numerário.",
540       "Variáveis centralizadas na média amostral antes da estimação."
541     ),
542     general_title = "Nota:", footnote_as_chunk = TRUE, escape = FALSE
543   )
544
545 writeLines(as.character(tabela_3_latex),
546           file.path(dir_outputs, "tabela_3_descritivas_log.tex"))
547 cat(" Tabela 3 (Descritivas Log Privado vs Não Privado) salva\n")
548
549
550 cat("\n=== TRANSFORMANDO VARIVEIS ===\n")
551
552 dados_sfa_painel <- dados_sfa_painel %>%
553   mutate(
554     VC_norm = VC / P_e,
555     P_l_norm = P_l / P_e,
556     P_o_norm = P_o / P_e
557   )
558
559 n_antes <- nrow(dados_sfa_painel)
560
561 dados_sfa_painel <- dados_sfa_painel %>%
562   filter(
563     VC_norm > 0, P_l_norm > 0, P_o_norm > 0,
564     Y_agua > 0, Y_esgo > 0, K > 0,
565     Dens > 0, !is.na(Sewerage)
566   )
567
568 n_depois <- nrow(dados_sfa_painel)
569 cat(sprintf("\nObservações removidas por valores inválidos: %d (%.1f%%)\n",
570           n_antes - n_depois, (n_antes - n_depois) / n_antes * 100))
571
572 prestadores_removidos <- dados_sfa_painel %>%
573   group_by(cod_prestador) %>%
574   summarise(n_anos = n_distinct(ano)) %>%
575   filter(n_anos < n_anos_completo)
576
577 if(nrow(prestadores_removidos) > 0) {
578   dados_sfa_painel <- dados_sfa_painel %>%
579     group_by(cod_prestador) %>%
580     filter(n() == n_anos_completo) %>%
581     ungroup()
582 }
583
584 # Transformações log e centralização na média
585 dados_sfa_painel <- dados_sfa_painel %>%
586   mutate(
587     lnVC = log(VC_norm),
588     lnY_agua = log(Y_agua),
589     lnY_esgo = log(Y_esgo),
590     lnK = log(K),
591     lnP_l = log(P_l_norm),
592     lnP_o = log(P_o_norm),
593     lnDens = log(Dens),
594     # lnMicro: log(Micro + 0.01) para evitar log(0) quando Micro = 0.
595     # A constante 0,01 é escolhida para ser negligenciável dado que

```

```

596 # a taxa média de micromedição da amostra é 96% (mediana 100%).
597 lnMicro = log(Micro + 0.01)
598 ) %>%
599 mutate(
600   lnVC_c = lnVC - mean(lnVC, na.rm = TRUE),
601   lnY_agua_c = lnY_agua - mean(lnY_agua, na.rm = TRUE),
602   lnY_esgo_c = lnY_esgo - mean(lnY_esgo, na.rm = TRUE),
603   lnK_c = lnK - mean(lnK, na.rm = TRUE),
604   lnP_l_c = lnP_l - mean(lnP_l, na.rm = TRUE),
605   lnP_o_c = lnP_o - mean(lnP_o, na.rm = TRUE)
606 ) %>%
607 mutate(
608   lnP_l_sq = 0.5 * lnP_l_c^2,
609   lnP_o_sq = 0.5 * lnP_o_c^2,
610   lnPl_Po = lnP_l_c * lnP_o_c,
611   lnY_agua_sq = 0.5 * lnY_agua_c^2,
612   lnY_esgo_sq = 0.5 * lnY_esgo_c^2,
613   lnYa_Ye = lnY_agua_c * lnY_esgo_c,
614   lnK_sq = 0.5 * lnK_c^2,
615   lnPl_Ya = lnP_l_c * lnY_agua_c,
616   lnPl_Ye = lnP_l_c * lnY_esgo_c,
617   lnPo_Ya = lnP_o_c * lnY_agua_c,
618   lnPo_Ye = lnP_o_c * lnY_esgo_c,
619   lnPl_K = lnP_l_c * lnK_c,
620   lnPo_K = lnP_o_c * lnK_c,
621   lnYa_K = lnY_agua_c * lnK_c,
622   lnYe_K = lnY_esgo_c * lnK_c
623 )
624
625 # Verificar infinitos e NAs
626 vars_log <- c("lnVC", "lnY_agua", "lnY_esgo", "lnK", "lnP_l", "lnP_o", "lnDens", "lnMicro")
627
628 tem_problemas <- FALSE
629 for(var in vars_log) {
630   n_inf <- sum(is.infinite(dados_sfa_painel[[var]]))
631   n_na <- sum(is.na(dados_sfa_painel[[var]]))
632   if(n_inf > 0 | n_na > 0) {
633     cat(sprintf("%s: %d Inf, %d NA\n", var, n_inf, n_na))
634     tem_problemas <- TRUE
635   }
636 }
637
638 if(tem_problemas) {
639   dados_sfa_painel <- dados_sfa_painel %>%
640     filter(across(all_of(vars_log), is.finite)) %>%
641     group_by(cod_prestador) %>%
642     filter(n() == n_anos_completo) %>%
643     ungroup()
644 }
645
646 cat(sprintf("\nPainel final: %d obs | %d prestadores | %d anos\n",
647           nrow(dados_sfa_painel),
648           n_distinct(dados_sfa_painel$cod_prestador),
649           n_distinct(dados_sfa_painel$ano)))
650
651 # TABELA 3: Descritivas Log em português
652 variaveis_log <- dados_sfa_painel %>%
653   select(lnVC, lnY_agua, lnY_esgo, lnK, lnP_l, lnP_o, lnDens, lnMicro, Loss) %>%
654   summarise(across(everything(), list(
655     Media = ~mean(., na.rm = TRUE),
656     DP = ~sd(., na.rm = TRUE),
657     Min = ~min(., na.rm = TRUE),
658     Max = ~max(., na.rm = TRUE)
659   ), .names = "{.col}_{.fn}"))
660
661 variaveis_log_long <- variaveis_log %>%
662   pivot_longer(everything(),

```

```

663     names_to = c("Variável", "Estatística"),
664     names_pattern = "(.+)_(.+)$" %>%
665 pivot_wider(names_from = Estatística, values_from = value) %>%
666 mutate(Variável = case_when(
667   Variável == "lnVC" ~ "ln(Custo Variável / P\\textsubscript{E})",
668   Variável == "lnY_agua" ~ "ln(Produto gua)",
669   Variável == "lnY_esgo" ~ "ln(Produto Esgoto)",
670   Variável == "lnK" ~ "ln(Extensão da Rede)",
671   Variável == "lnP_l" ~ "ln(Preço do Trabalho / P\\textsubscript{E})",
672   Variável == "lnP_o" ~ "ln(Preço de Outros Insumos / P\\textsubscript{E})",
673   Variável == "lnDens" ~ "ln(Densidade da Rede)",
674   Variável == "lnMicro" ~ "ln(Micromedicação)",
675   Variável == "Loss" ~ "Índice de Perdas (%)",
676   TRUE ~ Variável
677 ))
678
679 write_csv(variaveis_log_long,
680   file.path(dir_outputs, "tabela_3_descritivas_log.csv"))
681
682 tabela_3_latex <- kable(
683   variaveis_log_long,
684   format = "latex", booktabs = TRUE, digits = 4, escape = FALSE,
685   col.names = c("Variável", "Média", "DP", "Mín.", "Máx."),
686   caption = "Estatísticas Descritivas das Variáveis (Transformação Logarítmica)",
687   label = "desc_stats_log"
688 ) %>%
689 kable_styling(latex_options = c("hold_position")) %>%
690 footnote(
691   general = c(
692     "Todas as variáveis, exceto o Índice de Perdas, estão em forma logarítmica.",
693     "Custos e preços normalizados pelo preço da energia (P\\textsubscript{E}) como
694     ↪ número.",
695     "Variáveis centralizadas na média amostral antes da estimação."
696   ),
697   general_title = "Nota:", footnote_as_chunk = TRUE, escape = FALSE
698 )
699 writeLines(as.character(tabela_3_latex),
700   file.path(dir_outputs, "tabela_3_descritivas_log.tex"))
701 cat(" Tabela 3 (Descritivas Log) salva\\n")
702
703 # TABELA 4: Categóricas em português
704 categoricas_dist <- dados_sfa_painel %>%
705 summarise(
706   Treatment = sum(Treatment == 1) / n() * 100,
707   Independent = sum(Independent == 1) / n() * 100,
708   Private = sum(Private == 1) / n() * 100,
709   Local = sum(Local == 1) / n() * 100,
710   Multi = sum(Multi == 1) / n() * 100,
711   North = sum(NO == 1) / n() * 100,
712   Northeast = sum(NE == 1) / n() * 100,
713   Southeast = sum(SE == 1) / n() * 100,
714   South = sum(SO == 1) / n() * 100,
715   Midwest = sum(MI == 1) / n() * 100
716 ) %>%
717 pivot_longer(everything(), names_to = "Variável", values_to = "Percentual") %>%
718 mutate(
719   N = round(Percentual * nrow(dados_sfa_painel) / 100),
720   Variável = case_when(
721     Variável == "Treatment" ~ "Tratamento de Esgoto (>20\\%)",
722     Variável == "Independent" ~ "Gestão Independente",
723     Variável == "Private" ~ "Propriedade Privada",
724     Variável == "Local" ~ "Prestador Local",
725     Variável == "Multi" ~ "Município com Múltiplos Prestadores",
726     Variável == "North" ~ "Região Norte",
727     Variável == "Northeast" ~ "Região Nordeste",
728     Variável == "Southeast" ~ "Região Sudeste",

```

```

729     Variável == "South"         ~ "Região Sul",
730     Variável == "Midwest"      ~ "Região Centro-Oeste"
731   )
732   ) %>%
733   select(Variável, N, Percentual)
734
735   tabela_4_latex <- kable(
736     categoricas_dist,
737     format      = "latex", booktabs = TRUE, digits = 2, escape = FALSE,
738     col.names   = c("Variável", "N", "Percentual (\\%)"),
739     caption     = "Distribuição das Variáveis Categóricas",
740     label       = "categorical_dist"
741   ) %>%
742   kable_styling(latex_options = c("hold_position")) %>%
743   add_header_above(c(" " = 1, "Observações" = 2)) %>%
744   pack_rows("Variáveis Institucionais", 1, 5) %>%
745   pack_rows("Variáveis Regionais", 6, 10) %>%
746   footnote(
747     general = paste0("Total de observações: ", nrow(dados_sfa_painel)),
748     general_title = "Nota:", footnote_as_chunk = TRUE
749   )
750
751   writeLines(as.character(tabela_4_latex),
752     file.path(dir_outputs, "tabela_4_categoricas.tex"))
753   cat(" Tabela 4 (Categóricas) salva\\n")
754
755   # Matriz de correlação
756   correlacao <- dados_sfa_painel %>%
757     select(lnY_agua, lnY_esgo, lnK, lnP_l, lnP_o, lnDens, lnMicro, Loss) %>%
758     cor(use = "complete.obs") %>%
759     round(3)
760
761   write.csv(correlacao, file.path(dir_outputs, "matriz_correlacao.csv"))
762   cat(" Matriz de correlação salva\\n")
763
764   # Converter para pdata.frame
765   dados_sfa_painel_plm <- pdata.frame(dados_sfa_painel,
766     index = c("cod_prestador", "ano"))
767
768   # 14. FRMULAS DOS MODELOS ----
769
770   # Cobb-Douglas
771   formula_cd <- lnVC ~
772     lnP_l_c + lnP_o_c +
773     lnY_agua_c + lnY_esgo_c + lnK_c +
774     lnDens + lnMicro + Loss + Sewerage +
775     Treatment + Independent + Private + Local + Multi +
776     NO + NE + SE + SO
777
778   # Translog
779   formula_tl <- lnVC ~
780     lnP_l_c + lnP_o_c +
781     lnY_agua_c + lnY_esgo_c + lnK_c +
782     lnP_l_sq + lnP_o_sq + lnPl_Po +
783     lnY_agua_sq + lnY_esgo_sq + lnYa_Ye +
784     lnK_sq +
785     lnPl_Ya + lnPl_Ye +
786     lnPo_Ya + lnPo_Ye +
787     lnPl_K + lnPo_K +
788     lnYa_K + lnYe_K +
789     lnDens + lnMicro + Loss + Sewerage +
790     Treatment + Independent + Private + Local + Multi +
791     NO + NE + SE + SO
792
793   # 15. ESTIMAR 4 MODELOS COM FRONTIER ----
794
795   cat("\\n=== ESTIMANDO 4 MODELOS SFA ===\\n")

```

```

796
797 cat("\n1. Cobb-Douglas Tempo-Invariante...\n")
798 modelo_1_cd_ti <- frontier::sfa(
799   formula = formula_cd, data = dados_sfa_painel_plm,
800   ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = FALSE
801 )
802
803 cat("2. Cobb-Douglas Tempo-Variante...\n")
804 modelo_2_cd_tv <- frontier::sfa(
805   formula = formula_cd, data = dados_sfa_painel_plm,
806   ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = TRUE
807 )
808
809 cat("3. Translog Tempo-Invariante...\n")
810 modelo_3_tl_ti <- frontier::sfa(
811   formula = formula_tl, data = dados_sfa_painel_plm,
812   ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = FALSE
813 )
814
815 cat("4. Translog Tempo-Variante...\n")
816 modelo_4_tl_tv <- frontier::sfa(
817   formula = formula_tl, data = dados_sfa_painel_plm,
818   ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = TRUE
819 )
820
821 cat("\n Todos os modelos estimados com sucesso!\n")
822
823 # 16. COMPARAO DOS MODELOS ----
824
825 lista_modelos <- list(
826   "CD Tempo-Invariante" = modelo_1_cd_ti,
827   "CD Tempo-Variante"   = modelo_2_cd_tv,
828   "TL Tempo-Invariante" = modelo_3_tl_ti,
829   "TL Tempo-Variante"   = modelo_4_tl_tv
830 )
831
832 comparacao <- data.frame(
833   Modelo   = names(lista_modelos),
834   N_obs    = sapply(lista_modelos, nobs),
835   N_params = sapply(lista_modelos, function(m) length(coef(m))),
836   LogLik   = sapply(lista_modelos, function(m) as.numeric(logLik(m)))
837 )
838 comparacao$AIC <- -2 * comparacao$LogLik + 2 * comparacao$N_params
839 comparacao$BIC <- -2 * comparacao$LogLik + log(comparacao$N_obs) * comparacao$N_params
840
841 lr_cd      <- 2 * (comparacao$LogLik[2] - comparacao$LogLik[1])
842 lr_tl      <- 2 * (comparacao$LogLik[4] - comparacao$LogLik[3])
843 lr_spec_ti <- 2 * (comparacao$LogLik[3] - comparacao$LogLik[1])
844 lr_spec_tv <- 2 * (comparacao$LogLik[4] - comparacao$LogLik[2])
845
846 cat("\n=== COMPARAO DOS MODELOS ===\n")
847 print(comparacao)
848
849 # 17. EXTRAIR EFICIÊNCIAS ----
850
851 cat("\n=== EXTRAINDO EFICIÊNCIAS ===\n")
852
853 dados_sfa_painel$ef_cd_ti <- as.numeric(frontier::eficiencias(modelo_1_cd_ti, asInData =
854   ↪ TRUE))
855 dados_sfa_painel$ef_cd_tv <- as.numeric(frontier::eficiencias(modelo_2_cd_tv, asInData =
856   ↪ TRUE))
857 dados_sfa_painel$ef_tl_ti <- as.numeric(frontier::eficiencias(modelo_3_tl_ti, asInData =
858   ↪ TRUE))
859 dados_sfa_painel$ef_tl_tv <- as.numeric(frontier::eficiencias(modelo_4_tl_tv, asInData =
860   ↪ TRUE))
861
862 cat(" Eficiências extraídas com sucesso!\n")

```

```

859
860 estatisticas_ef <- data.frame(
861   Modelo = names(lista_modelos),
862   N       = sapply(c("ef_cd_ti", "ef_cd_tv", "ef_tl_ti", "ef_tl_tv"),
863                     function(x) sum(!is.na(dados_sfa_painel[[x]]))),
864   Media   = sapply(c("ef_cd_ti", "ef_cd_tv", "ef_tl_ti", "ef_tl_tv"),
865                     function(x) mean(dados_sfa_painel[[x]], na.rm = TRUE)),
866   Mediana = sapply(c("ef_cd_ti", "ef_cd_tv", "ef_tl_ti", "ef_tl_tv"),
867                     function(x) median(dados_sfa_painel[[x]], na.rm = TRUE)),
868   DP      = sapply(c("ef_cd_ti", "ef_cd_tv", "ef_tl_ti", "ef_tl_tv"),
869                     function(x) sd(dados_sfa_painel[[x]], na.rm = TRUE)),
870   Min     = sapply(c("ef_cd_ti", "ef_cd_tv", "ef_tl_ti", "ef_tl_tv"),
871                     function(x) min(dados_sfa_painel[[x]], na.rm = TRUE)),
872   Max     = sapply(c("ef_cd_ti", "ef_cd_tv", "ef_tl_ti", "ef_tl_tv"),
873                     function(x) max(dados_sfa_painel[[x]], na.rm = TRUE))
874 )
875
876 ef_por_ano <- dados_sfa_painel %>%
877   group_by(ano) %>%
878   summarise(
879     ef_media   = mean(ef_tl_tv, na.rm = TRUE),
880     ef_mediana = median(ef_tl_tv, na.rm = TRUE),
881     ef_q25     = quantile(ef_tl_tv, 0.25, na.rm = TRUE),
882     ef_q75     = quantile(ef_tl_tv, 0.75, na.rm = TRUE),
883     n = n(), .groups = "drop"
884   )
885
886 ef_por_regiao <- dados_sfa_painel %>%
887   mutate(regiao = case_when(
888     NO == 1 ~ "Norte", NE == 1 ~ "Nordeste",
889     SE == 1 ~ "Sudeste", SO == 1 ~ "Sul",
890     MI == 1 ~ "Centro-Oeste"
891   )) %>%
892   group_by(regiao) %>%
893   summarise(
894     n = n(),
895     n_prestadores = n_distinct(cod_prestador),
896     ef_media = mean(ef_tl_tv, na.rm = TRUE),
897     .groups = "drop"
898   ) %>%
899   arrange(desc(ef_media))
900
901 # 18. TABELAS LATEX (TABELAS 5 E 6) em português ----
902
903 # Função auxiliar para extrair coeficientes
904 extrair_coef <- function(modelo) {
905   coef_est <- coef(modelo)
906   coef_se <- sqrt(diag(vcov(modelo)))
907   data.frame(
908     estimate = coef_est,
909     std.error = coef_se,
910     statistic = coef_est / coef_se,
911     p.value = 2 * (1 - pnorm(abs(coef_est / coef_se)))
912   )
913 }
914
915 # Função auxiliar para extrair parâmetros de variância
916 extrair_params_variancia <- function(modelo) {
917   params <- coef(modelo)
918   sigma_sq <- if("sigmaSq" %in% names(params)) params["sigmaSq"] else NA
919   gamma <- if("gamma" %in% names(params)) params["gamma"] else NA
920   if(!is.na(sigma_sq) && !is.na(gamma)) {
921     lambda <- sqrt(gamma / (1 - gamma))
922     sigma_u <- sqrt(gamma * sigma_sq)
923     sigma_v <- sqrt((1 - gamma) * sigma_sq)
924   } else { lambda <- sigma_u <- sigma_v <- NA }
925   list(sigma_sq=sigma_sq, gamma=gamma, lambda=lambda, sigma_u=sigma_u, sigma_v=sigma_v)

```

```

926 }
927
928 coef_tv <- extrair_coef(modelo_4_tl_tv)
929
930 criar_linha_unico <- function(nome_display, var_name, coef_df) {
931   if(var_name %in% rownames(coef_df)) {
932     est <- coef_df[var_name, "estimate"]
933     se <- coef_df[var_name, "std.error"]
934     p <- coef_df[var_name, "p.value"]
935     stars <- ifelse(p < 0.001, "***",
936                    ifelse(p < 0.01, "**",
937                           ifelse(p < 0.05, "*", "")))
938     data.frame(
939       Variável = nome_display,
940       Estimativa = sprintf("%.4f%s", est, stars),
941       EP = sprintf("%.4f", se),
942       stringsAsFactors = FALSE
943     )
944   } else {
945     data.frame(Variável = nome_display, Estimativa = "", EP = "",
946               stringsAsFactors = FALSE)
947   }
948 }
949
950 # Parâmetros da fronteira
951 vars_tabela <- list(
952   list("Intercepto", "(Intercept)"),
953   list("$\\ln(P_L / P_E)$", "lnP_l_c"),
954   list("$\\ln(P_O / P_E)$", "lnP_o_c"),
955   list("$\\ln(Y_{\\text{água}})$", "lnY_agua_c"),
956   list("$\\ln(Y_{\\text{esgoto}})$", "lnY_esgo_c"),
957   list("$\\ln(K)$", "lnK_c"),
958   list("$\\frac{1}{2}[\\ln(P_L / P_E)]^2$", "lnP_l_sq"),
959   list("$\\frac{1}{2}[\\ln(P_O / P_E)]^2$", "lnP_o_sq"),
960   list("$\\ln(P_L / P_E) \\times \\ln(P_O / P_E)$", "lnPl_Po"),
961   list("$\\frac{1}{2}[\\ln(Y_{\\text{água}})]^2$", "lnY_agua_sq"),
962   list("$\\frac{1}{2}[\\ln(Y_{\\text{esgoto}})]^2$", "lnY_esgo_sq"),
963   list("$\\ln(Y_{\\text{água}}) \\times \\ln(Y_{\\text{esgoto}})$", "lnYa_Ye"),
964   list("$\\frac{1}{2}[\\ln(K)]^2$", "lnK_sq"),
965   list("$\\ln(P_L / P_E) \\times \\ln(Y_{\\text{água}})$", "lnPl_Ya"),
966   list("$\\ln(P_L / P_E) \\times \\ln(Y_{\\text{esgoto}})$", "lnPl_Ye"),
967   list("$\\ln(P_O / P_E) \\times \\ln(Y_{\\text{água}})$", "lnPo_Ya"),
968   list("$\\ln(P_O / P_E) \\times \\ln(Y_{\\text{esgoto}})$", "lnPo_Ye"),
969   list("$\\ln(P_L / P_E) \\times \\ln(K)$", "lnPl_K"),
970   list("$\\ln(P_O / P_E) \\times \\ln(K)$", "lnPo_K"),
971   list("$\\ln(Y_{\\text{água}}) \\times \\ln(K)$", "lnYa_K"),
972   list("$\\ln(Y_{\\text{esgoto}}) \\times \\ln(K)$", "lnYe_K"),
973   list("$\\ln(\\text{Densidade})$", "lnDens"),
974   list("$\\ln(\\text{Micromedicação})$", "lnMicro"),
975   list("Índice de Perdas", "Loss"),
976   list("Cobertura de Esgoto", "Sewerage"),
977   list("Tratamento de Esgoto", "Treatment"),
978   list("Gestão Independente", "Independent"),
979   list("Propriedade Privada", "Private"),
980   list("Prestador Local", "Local"),
981   list("Múltiplos Prestadores", "Multi"),
982   list("Norte", "NO"),
983   list("Nordeste", "NE"),
984   list("Sudeste", "SE"),
985   list("Sul", "SO")
986 )
987
988 tabela_coef_unico <- do.call(rbind, lapply(vars_tabela, function(v) {
989   criar_linha_unico(v[[1]], v[[2]], coef_tv)
990 })))
991
992 params_4 <- extrair_params_variancia(modelo_4_tl_tv)

```

```

993
994 stats_unico <- data.frame(
995   Variável = c(
996     "\\sigma^2$", "\\gamma$",
997     "\\lambda = \\sigma_u / \\sigma_v$",
998     "\\sigma_u$", "\\sigma_v$",
999     "Observações", "Prestadores", "Anos",
1000     "Log-Verossimilhança", "AIC", "BIC",
1001     "Eficiência Técnica Média"
1002   ),
1003   Estimativa = c(
1004     sprintf("%.4f", params_4$sigma_sq),
1005     sprintf("%.4f", params_4$gamma),
1006     sprintf("%.4f", params_4$lambda),
1007     sprintf("%.4f", params_4$sigma_u),
1008     sprintf("%.4f", params_4$sigma_v),
1009     as.character(nrow(dados_sfa_painel)),
1010     as.character(n_distinct(dados_sfa_painel$cod_prestador)),
1011     as.character(n_distinct(dados_sfa_painel$ano)),
1012     sprintf("%.2f", as.numeric(logLik(modelo_4_tlt_tv))),
1013     sprintf("%.2f", comparacao$AIC[4]),
1014     sprintf("%.2f", comparacao$BIC[4]),
1015     sprintf("%.4f", mean(dados_sfa_painel$ef_tlt_tv, na.rm = TRUE))
1016   ),
1017   EP = rep("", 12),
1018   stringsAsFactors = FALSE
1019 )
1020
1021 tabela_5_final <- rbind(tabela_coef_unico, stats_unico)
1022
1023 tabela_5_latex <- kable(
1024   tabela_5_final,
1025   format = "latex", booktabs = TRUE, escape = FALSE,
1026   col.names = c("Variável", "Estimativa", "EP"),
1027   caption = "Resultados da Análise de Fronteira Estocástica Modelo Translog
1028     ↳ Tempo-Variante",
1029   label = "sfa_results",
1030   align = c("l", "c", "c")
1031 ) %>%
1032   kable_styling(latex_options = c("hold_position"), font_size = 10) %>%
1033   pack_rows("Parâmetros da Fronteira Preços", 2, 3) %>%
1034   pack_rows("Parâmetros da Fronteira Produtos", 4, 5) %>%
1035   pack_rows("Parâmetros da Fronteira Capital", 6, 6) %>%
1036   pack_rows("Termos Quadráticos Preços", 7, 9) %>%
1037   pack_rows("Termos Quadráticos Produtos e Capital", 10, 13) %>%
1038   pack_rows("Interações Preço  $\\times$  Produto", 14, 17) %>%
1039   pack_rows("Interações Preço  $\\times$  Capital", 18, 19) %>%
1040   pack_rows("Interações Produto  $\\times$  Capital", 20, 21) %>%
1041   pack_rows("Variáveis de Controle", 22, 25) %>%
1042   pack_rows("Variáveis Institucionais", 26, 30) %>%
1043   pack_rows("Variáveis Regionais", 31, 34) %>%
1044   pack_rows("Parâmetros de Variância", 35, 39) %>%
1045   pack_rows("Estatísticas do Modelo", 40, 46) %>%
1046   footnote(
1047     general = c(
1048       "Erros padrão entre parênteses. * p <$ 0,05; ** p <$ 0,01; *** p <$ 0,001.",
1049       "Variável dependente: ln(Custo Variável / P\\\\textsubscript{E}). CV = FN015.",
1050       "Preços normalizados pelo preço da energia (P\\\\textsubscript{E}) como numerário.",
1051       "Todas as variáveis contínuas centralizadas na média amostral.",
1052       paste0("Modelo selecionado pelos testes LR contra Cobb-Douglas (LR = ",
1053         sprintf("%.2f", lr_spec_tv), "; p <$ 0,001) e contra especificação
1054         ↳ tempo-invariante (LR = ",
1055         sprintf("%.2f", lr_tlt), "; p <$ 0,001).")
1056     ),
1057     general_title = "Notas:", footnote_as_chunk = TRUE, escape = FALSE
1058   )
1059

```

```

1058 writeLines(as.character(tabela_5_latex),
1059             file.path(dir_outputs, "tabela_5_resultados_sfa.tex"))
1060 cat(" Tabela 5 (TL Tempo-Variante) salva\n")
1061
1062 # TABELA 6: Estatísticas de eficiência em português
1063 tabela_6_latex <- kable(
1064   estatisticas_ef,
1065   format = "latex", booktabs = TRUE, digits = 4,
1066   col.names = c("Modelo", "N", "Média", "Mediana", "DP", "Mín.", "Máx."),
1067   caption = "Estatísticas de Eficiência Técnica por Modelo",
1068   label = "efficiency_stats"
1069 ) %>%
1070 kable_styling(latex_options = c("hold_position")) %>%
1071 footnote(
1072   general = "Escores de eficiência variam de 0 (completamente ineficiente) a 1 (plenamente
1073     ↪ eficiente).",
1074   general_title = "Nota:", footnote_as_chunk = TRUE
1075 )
1076
1077 writeLines(as.character(tabela_6_latex),
1078             file.path(dir_outputs, "tabela_6_estatisticas_eficiencia.tex"))
1079 cat(" Tabela 6 salva\n")
1080
1081 # Tabela de eficiência por região em português
1082 tabela_regiao_latex <- kable(
1083   ef_por_regiao,
1084   format = "latex", booktabs = TRUE, digits = 4,
1085   col.names = c("Região", "Observações", "Prestadores", "Eficiência Média"),
1086   caption = "Eficiência Técnica por Região Modelo Translog Tempo-Variante",
1087   label = "efficiency_region"
1088 ) %>%
1089 kable_styling(latex_options = c("hold_position"))
1090
1091 writeLines(as.character(tabela_regiao_latex),
1092             file.path(dir_outputs, "tabela_eficiencia_regiao.tex"))
1093
1094 # Tabela de eficiência ao longo do tempo em português
1095 tabela_tempo_latex <- kable(
1096   ef_por_ano %>% select(ano, ef_media, ef_mediana, n),
1097   format = "latex", booktabs = TRUE, digits = 4,
1098   col.names = c("Ano", "Eficiência Média", "Eficiência Mediana", "N"),
1099   caption = "Evolução da Eficiência Técnica ao Longo do Tempo Modelo Translog
1100     ↪ Tempo-Variante",
1101   label = "efficiency_time"
1102 ) %>%
1103 kable_styling(latex_options = c("hold_position"))
1104
1105 writeLines(as.character(tabela_tempo_latex),
1106             file.path(dir_outputs, "tabela_eficiencia_tempo.tex"))
1107
1108 cat(" Tabelas adicionais salvas\n")
1109
1110 # =====
1111 # TABELAS 8 E 9 TOP 5 MAIS EFICIENTES E TOP 5 MENOS EFICIENTES
1112 # =====
1113
1114 cat("\n=== GERANDO TABELAS DE RANKING DE EFICIÊNCIA ===\n")
1115
1116 # Calcular eficiência MDIA por prestador ao longo de todo o período
1117 # e agregar variáveis de caracterização
1118 ef_por_prestador <- dados_sfa_painel %>%
1119   mutate(
1120     regiao = case_when(
1121       NO == 1 ~ "Norte", NE == 1 ~ "Nordeste",
1122       SE == 1 ~ "Sudeste", SO == 1 ~ "Sul",
1123       MI == 1 ~ "Centro-Oeste"
1124     ),

```

```

1123   tipo_gestao = case_when(
1124     Private      == 1 ~ "Privado",
1125     Independent == 0 ~ "Adm. Direta",
1126     TRUE         ~ "Emp. Estatal"
1127   )
1128 ) %>%
1129 group_by(cod_prestador, prestador, estado, regioao, tipo_gestao) %>%
1130 summarise(
1131   ef_media      = mean(ef_tl_tv, na.rm = TRUE),
1132   ef_min        = min(ef_tl_tv, na.rm = TRUE),
1133   ef_max        = max(ef_tl_tv, na.rm = TRUE),
1134   pop_atendida = mean(ag001, na.rm = TRUE), # média do período
1135   abrangencia  = first(abrangencia),
1136   .groups      = "drop"
1137 ) %>%
1138 arrange(desc(ef_media))
1139
1140 # --- TOP 5 MAIS EFICIENTES ---
1141 top5_mais <- ef_por_prestador %>%
1142   slice_head(n = 5) %>%
1143   mutate(
1144     Rank          = paste0(row_number(), " "),
1145     pop_atendida_fmt = format(round(pop_atendida), big.mark = ".", scientific = FALSE),
1146     ef_media_fmt   = sprintf("%.4f", ef_media),
1147     ef_min_fmt     = sprintf("%.4f", ef_min),
1148     ef_max_fmt     = sprintf("%.4f", ef_max)
1149   ) %>%
1150   select(Rank, prestador, estado, regioao, tipo_gestao,
1151     pop_atendida_fmt, ef_media_fmt, ef_min_fmt, ef_max_fmt)
1152
1153 tabela_8_latex <- kable(
1154   top5_mais,
1155   format      = "latex", booktabs = TRUE, escape = TRUE,
1156   col.names   = c("Pos.", "Prestador", "UF", "Região",
1157     "Tipo de Gestão", "Pop. Atendida (hab.)",
1158     "Ef. Média", "Ef. Mín.", "Ef. Máx."),
1159   caption     = "Os Cinco Prestadores com Maior Eficiência Técnica Média (2015--2022)",
1160   label       = "top5_eficientes",
1161   align       = c("c", "l", "c", "c", "c", "r", "c", "c", "c")
1162 ) %>%
1163   kable_styling(
1164     latex_options = c("hold_position", "scale_down"),
1165     font_size     = 10
1166   ) %>%
1167   column_spec(2, width = "5.5cm") %>%
1168   footnote(
1169     general = paste0(
1170       "Eficiência técnica estimada pelo Modelo Translog Tempo-Variante (SFA). ",
1171       "Ef. Média = média da eficiência técnica anual no período 2015--2022. ",
1172       "População atendida com abastecimento de água (AG001), média do período."
1173     ),
1174     general_title = "Nota:", footnote_as_chunk = TRUE
1175   )
1176
1177 writeLines(as.character(tabela_8_latex),
1178   file.path(dir_outputs, "tabela_8_top5_mais_eficientes.tex"))
1179 cat(" Tabela 8 (Top 5 mais eficientes) salva\n")
1180
1181 # --- TOP 5 MENOS EFICIENTES ---
1182 top5_menos <- ef_por_prestador %>%
1183   slice_tail(n = 5) %>%
1184   arrange(ef_media) %>%
1185   mutate(
1186     Rank          = paste0(row_number(), " "),
1187     pop_atendida_fmt = format(round(pop_atendida), big.mark = ".", scientific = FALSE),
1188     ef_media_fmt   = sprintf("%.4f", ef_media),
1189     ef_min_fmt     = sprintf("%.4f", ef_min),

```

```

1190     ef_max_fmt      = sprintf("%.4f", ef_max)
1191   ) %>%
1192   select(Rank, prestador, estado, regioao, tipo_gestao,
1193          pop_atendida_fmt, ef_media_fmt, ef_min_fmt, ef_max_fmt)
1194
1195 tabela_9_latex <- kable(
1196   top5_menos,
1197   format      = "latex", booktabs = TRUE, escape = TRUE,
1198   col.names   = c("Pos.", "Prestador", "UF", "Região",
1199                  "Tipo de Gestão", "Pop. Atendida (hab.)",
1200                  "Ef. Média", "Ef. Mín.", "Ef. Máx."),
1201   caption     = "Os Cinco Prestadores com Menor Eficiência Técnica Média (2015--2022)",
1202   label       = "bottom5_eficientes",
1203   align       = c("c", "l", "c", "c", "c", "r", "c", "c", "c")
1204 ) %>%
1205 kable_styling(
1206   latex_options = c("hold_position", "scale_down"),
1207   font_size     = 10
1208 ) %>%
1209 column_spec(2, width = "5.5cm") %>%
1210 footnote(
1211   general = paste0(
1212     "Eficiência técnica estimada pelo Modelo Translog Tempo-Variante (SFA). ",
1213     "Ef. Média = média da eficiência técnica anual no período 2015--2022. ",
1214     "População atendida com abastecimento de água (AG001), média do período."
1215   ),
1216   general_title = "Nota:", footnote_as_chunk = TRUE
1217 )
1218
1219 writeLines(as.character(tabela_9_latex),
1220            file.path(dir_outputs, "tabela_9_top5_menos_eficientes.tex"))
1221 cat(" Tabela 9 (Top 5 menos eficientes) salva\n")
1222
1223 # Salvar também em CSV para referência
1224 write_csv(ef_por_prestador %>% slice_head(n = 5),
1225            file.path(dir_outputs, "ranking_top5_mais_eficientes.csv"))
1226 write_csv(ef_por_prestador %>% slice_tail(n = 5) %>% arrange(ef_media),
1227            file.path(dir_outputs, "ranking_top5_menos_eficientes.csv"))
1228
1229 cat(" Rankings de eficiência (CSV) salvos\n")
1230
1231 # =====
1232 # 19. FIGURA 1 MAPA DO BRASIL COM AMOSTRA DO MODELO PRINCIPAL
1233 # Círculos proporcionais população atendida (ag001)
1234 # =====
1235
1236 cat("\n=== GERANDO FIGURA 1: MAPA DO BRASIL ===\n")
1237
1238 library(sf)
1239 library(geobr)
1240 library(ggplot2)
1241 library(dplyr)
1242
1243 # -----
1244 # PASSO 1: Um registro por prestador, com população TOTAL atendida (média
1245 # do período) e a região já disponível nos dados do SNIS
1246 # -----
1247 dados_prestadores_mapa <- dados_sfa_painel %>%
1248   mutate(
1249     regioao = case_when(
1250       NO == 1 ~ "Norte",
1251       NE == 1 ~ "Nordeste",
1252       SE == 1 ~ "Sudeste",
1253       SO == 1 ~ "Sul",
1254       MI == 1 ~ "Centro-Oeste",
1255       TRUE   ~ NA_character_
1256     )

```

```

1257 ) %>%
1258 group_by(cod_prestador, prestador, cod_municipio, regioao) %>%
1259 summarise(
1260   pop_atendida = mean(ag001, na.rm = TRUE), # média do período
1261   .groups = "drop"
1262 ) %>%
1263 mutate(
1264   # SNIS usa 6 dígitos; IBGE/geobr usa 7 (6 + dígito verificador)
1265   # Estratégia: converter os 7 dígitos do geobr para 6 no join (truncar)
1266   cod_municipio_6 = as.integer(cod_municipio)
1267 )
1268
1269 # -----
1270 # PASSO 2: Shapefiles do Brasil
1271 # -----
1272 municipios_br <- geobr::read_municipality(year = 2020, simplified = TRUE)
1273 estados_br <- geobr::read_state(year = 2020, simplified = TRUE)
1274
1275 # Verificar nome da coluna de código no shapefile
1276 cat("\nColunas do shapefile municipios_br:\n")
1277 print(names(municipios_br))
1278
1279 # Detectar automaticamente o nome da coluna de código do município
1280 col_code <- intersect(names(municipios_br), c("code_muni", "code_municipality", "cd_mun",
1281   ↪ "codmun", "cod_mun"))
1282 if (length(col_code) == 0) stop("Coluna de código do município não encontrada no shapefile.
1283   ↪ Colunas: ", paste(names(municipios_br), collapse=" "))
1284 col_code <- col_code[1]
1285 cat(sprintf("Coluna de código detectada: %s\n", col_code))
1286 cat("Exemplos de valores:\n")
1287 print(head(sort(municipios_br[[col_code]]), 6))
1288
1289 # Criar chave de 6 dígitos no shapefile (truncar o dígito verificador)
1290 municipios_br <- municipios_br %>%
1291   mutate(code_muni_6 = as.integer(floor(.data[[col_code]] / 10)))
1292
1293 # Diagnóstico
1294 cat(sprintf("\nPrestadores a plotar: %d\n", nrow(dados_prestadores_mapa)))
1295 cat("Distribuição por região:\n")
1296 print(table(dados_prestadores_mapa$regiao, useNA = "always"))
1297 cat("\nExemplos cod_municipio_6 (SNIS 6 dig):\n")
1298 print(head(dados_prestadores_mapa$cod_municipio_6, 6))
1299 cat("\nExemplos code_muni_6 no shapefile (7 dig -> 6 dig):\n")
1300 print(head(sort(municipios_br$code_muni_6), 6))
1301 n_match <- sum(dados_prestadores_mapa$cod_municipio_6 %in% municipios_br$code_muni_6)
1302 cat(sprintf("\nPrestadores com match no shapefile: %d / %d\n",
1303   n_match, nrow(dados_prestadores_mapa)))
1304
1305 # -----
1306 # PASSO 3: Obter centroide de cada município-sede do prestador
1307 # -----
1308 municipios_amostra <- municipios_br %>%
1309   filter(code_muni_6 %in% dados_prestadores_mapa$cod_municipio_6) %>%
1310   st_centroid()
1311
1312 coords_muni <- st_coordinates(municipios_amostra)
1313 municipios_amostra$lon <- coords_muni[, 1]
1314 municipios_amostra$lat <- coords_muni[, 2]
1315
1316 # Join pela chave de 6 dígitos
1317 centroides_df <- municipios_amostra %>%
1318   st_drop_geometry() %>%
1319   select(code_muni_6, lon, lat)
1320
1321 dados_plot <- dados_prestadores_mapa %>%
1322   left_join(centroides_df, by = c("cod_municipio_6" = "code_muni_6"))

```

```

1322 n_sem_coords <- sum(is.na(dados_plot$lon))
1323 cat(sprintf("Prestadores sem coordenadas após join: %d\n", n_sem_coords))
1324
1325 dados_plot <- dados_plot %>% filter(!is.na(lon), !is.na(lat), !is.na(regiao))
1326 cat(sprintf("Prestadores com coordenadas e região: %d\n", nrow(dados_plot)))
1327
1328 # -----
1329 # PASSO 4: Plotar
1330 # -----
1331 cores_regiao <- c(
1332   "Norte"      = "#2E86AB",
1333   "Nordeste"   = "#E84855",
1334   "Sudeste"    = "#3BB273",
1335   "Sul"        = "#F4A261",
1336   "Centro-Oeste" = "#9B59B6"
1337 )
1338
1339 figura_1 <- ggplot() +
1340   # Base: municípios em cinza claro
1341   geom_sf(data = municipios_br, fill = "#F0F0F0", color = NA) +
1342   # Contorno dos estados
1343   geom_sf(data = estados_br, fill = NA, color = "#BBBBBB", linewidth = 0.4) +
1344   # Círculos coloridos por região
1345   geom_point(
1346     data = dados_plot %>% arrange(pop_atendida),
1347     aes(x = lon, y = lat,
1348         size = pop_atendida / 1000,
1349         color = regiao),
1350     alpha = 0.85, shape = 16
1351   ) +
1352   scale_size_area(
1353     name = "Pop. Atendida\n(mil hab.)",
1354     max_size = 16,
1355     breaks = c(10, 100, 500, 1000, 5000),
1356     labels = c("10 mil", "100 mil", "500 mil", "1 milhão", "5 milhões")
1357   ) +
1358   scale_color_manual(
1359     name = "Região",
1360     values = c(
1361       "Norte"      = "#0077B6",
1362       "Nordeste"   = "#D62828",
1363       "Sudeste"    = "#2DC653",
1364       "Sul"        = "#F4A100",
1365       "Centro-Oeste" = "#7B2FBE"
1366     )
1367   ) +
1368   guides(
1369     size = guide_legend(order = 1, override.aes = list(alpha = 0.85)),
1370     color = guide_legend(order = 2, override.aes = list(size = 4, alpha = 0.85))
1371   ) +
1372   labs(
1373     title = "Distribuição Geográfica da Amostra Modelo Principal (20152022)",
1374     subtitle = paste0(
1375       n_distinct(dados_plot$cod_prestador), " prestadores | ",
1376       n_distinct(dados_plot$cod_municipio_6), " municípios-sede | ",
1377       sprintf("%.1f", mean(cobertura_pop$perc_cobertura)), "% da pop. brasileira"
1378     ),
1379     caption = paste0(
1380       "Fonte: SNIS (20152022). Cada círculo representa um prestador de serviço, ",
1381       "posicionado no centroide do município-sede.\nTamanho proporcional população ",
1382       "média atendida com abastecimento de água no período (AG001)."
1383     )
1384   ) +
1385   theme_void(base_family = "serif") +
1386   theme(
1387     plot.background = element_rect(fill = "white", color = NA),
1388     panel.background = element_rect(fill = "white", color = NA),

```

```

1389   plot.title      = element_text(size = 13, face = "bold", hjust = 0.5,
1390                                   color = "grey10", margin = margin(b = 4)),
1391   plot.subtitle   = element_text(size = 10, hjust = 0.5, color = "grey40",
1392                                   margin = margin(b = 8)),
1393   plot.caption    = element_text(size = 8, hjust = 0, color = "grey50",
1394                                   margin = margin(t = 8)),
1395   legend.position = "right",
1396   legend.background = element_rect(fill = NA, color = NA),
1397   legend.title     = element_text(size = 9, face = "bold", color = "grey10"),
1398   legend.text      = element_text(size = 8, color = "grey20"),
1399   legend.key.size  = unit(0.5, "cm"),
1400   plot.margin      = margin(10, 10, 10, 10)
1401 )
1402
1403 ggsave(
1404   filename = file.path(dir_outputs, "figura_1_mapa_amostra.pdf"),
1405   plot     = figura_1,
1406   width    = 18, height = 16, units = "cm", dpi = 300
1407 )
1408 ggsave(
1409   filename = file.path(dir_outputs, "figura_1_mapa_amostra.png"),
1410   plot     = figura_1,
1411   width    = 18, height = 16, units = "cm", dpi = 300
1412 )
1413
1414 cat(" Figura 1 (Mapa da Amostra) salva\n")
1415
1416 # =====
1417 # 20. FIGURA 2  EVOLUO DA EFICIÊNCIA TCNICA AO LONGO DO TEMPO
1418 # =====
1419
1420 cat("\n=== GERANDO FIGURA 2: EVOLUO DA EFICIÊNCIA TCNICA ===\n")
1421
1422 library(ggplot2)
1423
1424 # Preparar dados por ano e por tipo institucional (para linhas secundárias)
1425 ef_por_ano_inst <- dados_sfa_painel %>%
1426   mutate(
1427     tipo = case_when(
1428       natureza_juridica == "Empresa privada"
1429       ~ "Privado",
1430       natureza_juridica == "Administração pública direta"
1431       ~ "Adm. Direta",
1432       natureza_juridica == "Autarquia"
1433       ~ "Autarquia",
1434       natureza_juridica %in% c(
1435         "Sociedade de economia mista com administração pública",
1436         "Empresa pública",
1437         "Sociedade de economia mista com administração predominantemente privada"
1438       ) ~ "Emp. Estatal/Mista"
1439     )
1440   ) %>%
1441   filter(!is.na(tipo)) %>%
1442   group_by(ano, tipo) %>%
1443   summarise(
1444     ef_media    = mean(ef_tl_tv, na.rm = TRUE),
1445     ef_mediana  = median(ef_tl_tv, na.rm = TRUE),
1446     ef_q25      = quantile(ef_tl_tv, 0.25, na.rm = TRUE),
1447     ef_q75      = quantile(ef_tl_tv, 0.75, na.rm = TRUE),
1448     n           = n(),
1449     .groups     = "drop"
1450   ) %>%
1451   mutate(ano_num = as.numeric(as.character(ano)))
1452
1453 ef_geral_num <- ef_por_ano %>%
1454   mutate(ano_num = as.numeric(as.character(ano)))
1455

```

```

1456 # Paleta quatro tipos, cores distintas e saturadas
1457 cores_inst <- c(
1458   "Emp. Estatal/Mista" = "#0077B6", # azul
1459   "Autarquia"         = "#E85D04", # laranja
1460   "Adm. Direta"       = "#9B2226", # vermelho escuro
1461   "Privado"           = "#2DC653"  # verde
1462 )
1463
1464 ltipos_inst <- c(
1465   "Emp. Estatal/Mista" = "solid",
1466   "Autarquia"         = "dashed",
1467   "Adm. Direta"       = "dotdash",
1468   "Privado"           = "dotted"
1469 )
1470
1471 figura_2 <- ggplot() +
1472   # Faixa interquartil geral (cinza suave)
1473   geom_ribbon(
1474     data = ef_geral_num,
1475     aes(x = ano_num, ymin = ef_q25, ymax = ef_q75),
1476     fill = "grey80", alpha = 0.45
1477   ) +
1478   # Linhas por tipo institucional
1479   geom_line(
1480     data = ef_por_ano_inst,
1481     aes(x = ano_num, y = ef_media, color = tipo, linetype = tipo),
1482     linewidth = 1.1
1483   ) +
1484   geom_point(
1485     data = ef_por_ano_inst,
1486     aes(x = ano_num, y = ef_media, color = tipo, shape = tipo),
1487     size = 3
1488   ) +
1489   # Linha geral (média) preta, mais grossa, por cima
1490   geom_line(
1491     data = ef_geral_num,
1492     aes(x = ano_num, y = ef_media),
1493     color = "black", linewidth = 1.5, linetype = "solid"
1494   ) +
1495   geom_point(
1496     data = ef_geral_num,
1497     aes(x = ano_num, y = ef_media),
1498     color = "black", size = 3, shape = 16
1499   ) +
1500   # Linha de referência: média geral do período
1501   geom_hline(
1502     yintercept = mean(dados_sfa_painel$ef_tl_tv, na.rm = TRUE),
1503     linetype = "longdash", color = "grey50", linewidth = 0.6
1504   ) +
1505   annotate(
1506     "text",
1507     x = 2015.1,
1508     y = mean(dados_sfa_painel$ef_tl_tv, na.rm = TRUE) + 0.003,
1509     label = sprintf("Média geral = %.3f", mean(dados_sfa_painel$ef_tl_tv, na.rm = TRUE)),
1510     size = 3, hjust = 0, color = "grey40", family = "serif"
1511   ) +
1512   scale_x_continuous(
1513     breaks = 2015:2022,
1514     labels = as.character(2015:2022)
1515   ) +
1516   scale_y_continuous(
1517     labels = scales::percent_format(accuracy = 0.1)
1518   ) +
1519   scale_color_manual(
1520     name = "Tipo de Gestão",
1521     values = cores_inst
1522   ) +

```

```

1523 scale_linetype_manual(
1524   name = "Tipo de Gestão",
1525   values = ltipos_inst
1526 ) +
1527 scale_shape_manual(
1528   name = "Tipo de Gestão",
1529   values = c(
1530     "Emp. Estatal/Mista" = 16,
1531     "Autarquia" = 17,
1532     "Adm. Direta" = 15,
1533     "Privado" = 18
1534   )
1535 ) +
1536 labs(
1537   title = "Evolução da Eficiência Técnica nos Serviços de água e Esgoto",
1538   subtitle = "Brasil, 2015-2022 Modelo Translog Tempo-Variante (SFA)",
1539   x = NULL,
1540   y = "Eficiência Técnica Média",
1541   caption = paste0(
1542     "Fonte: SNIS (2015-2022). Modelo: Fronteira Estocástica Translog Tempo-Variante.\n",
1543     "Linha preta espessa: média geral. Faixa cinza: intervalo interquartil (P25-P75). ",
1544     "Linha tracejada cinza: média do período completo."
1545   )
1546 ) +
1547 theme_classic(base_family = "serif", base_size = 11) +
1548 theme(
1549   plot.title = element_text(size = 13, face = "bold", hjust = 0.5),
1550   plot.subtitle = element_text(size = 10, hjust = 0.5, color = "grey40",
1551                                 margin = margin(b = 8)),
1552   plot.caption = element_text(size = 8, hjust = 0, color = "grey50"),
1553   axis.title.y = element_text(margin = margin(r = 8)),
1554   legend.position = "bottom",
1555   legend.title = element_text(size = 9, face = "bold"),
1556   legend.text = element_text(size = 8),
1557   legend.key.width = unit(1.5, "cm"),
1558   panel.grid.major.y = element_line(color = "grey92", linewidth = 0.4),
1559   panel.grid.major.x = element_blank(),
1560   plot.margin = margin(10, 15, 10, 10)
1561 )
1562
1563 ggsave(
1564   filename = file.path(dir_outputs, "figura_2_evolucao_eficiencia.pdf"),
1565   plot = figura_2,
1566   width = 18, height = 12, units = "cm", dpi = 300
1567 )
1568 ggsave(
1569   filename = file.path(dir_outputs, "figura_2_evolucao_eficiencia.png"),
1570   plot = figura_2,
1571   width = 18, height = 12, units = "cm", dpi = 300
1572 )
1573
1574 cat(" Figura 2 (Evolução da Eficiência) salva\n")
1575
1576 # 21. SALVAR RESULTADOS ----
1577
1578 write_csv(dados_sfa_painel, file.path(dir_outputs, "dados_sfa_painel_final.csv"))
1579 write_csv(comparacao, file.path(dir_outputs, "comparacao_modelos.csv"))
1580 write_csv(estatisticas_ef, file.path(dir_outputs, "estatisticas_eficiencia.csv"))
1581 write_csv(ef_por_ano, file.path(dir_outputs, "eficiencia_por_ano.csv"))
1582 write_csv(ef_por_regiao, file.path(dir_outputs, "eficiencia_por_regiao.csv"))
1583 write_csv(cobertura_pop, file.path(dir_outputs, "cobertura_populacional.csv"))
1584 saveRDS(lista_modelos, file.path(dir_outputs, "modelos_sfa.rds"))
1585
1586 cat("\n===== \n")
1587 cat("ARQUIVOS SALVOS\n")
1588 cat("===== \n")
1589 cat("\nDados:\n")

```

```

1590 cat(" 1. dados_sfa_painel_final.csv\n")
1591 cat(" 2. cobertura_populacional.csv\n")
1592 cat(" 3. matriz_correlacao.csv\n")
1593 cat(" 4. log_imputacao_sewerage_2022.csv\n")
1594 cat("\nResultados:\n")
1595 cat(" 5. comparacao_modelos.csv\n")
1596 cat(" 6. estatisticas_eficiencia.csv\n")
1597 cat(" 7. eficiencia_por_ano.csv\n")
1598 cat(" 8. eficiencia_por_regiao.csv\n")
1599 cat(" 9. modelos_sfa.rds\n")
1600 cat("\nTabelas LaTeX (em português):\n")
1601 cat(" 10. tabela_2_descritivas_originais.tex\n")
1602 cat(" 11. tabela_3_descritivas_log.tex\n")
1603 cat(" 12. tabela_4_categoricas.tex\n")
1604 cat(" 13. tabela_5_resultados_sfa.tex\n")
1605 cat(" 14. tabela_6_estatisticas_eficiencia.tex\n")
1606 cat(" 15. tabela_eficiencia_regiao.tex\n")
1607 cat(" 16. tabela_eficiencia_tempo.tex\n")
1608 cat(" 17. tabela_8_top5_mais_eficientes.tex\n")
1609 cat(" 18. tabela_9_top5_menos_eficientes.tex\n")
1610 cat("\nFiguras:\n")
1611 cat(" 19. figura_1_mapa_amostra.pdf / .png\n")
1612 cat(" 20. figura_2_evolucao_eficiencia.pdf / .png\n")
1613
1614 cat("\n===== \n")
1615 cat("RESUMO DOS RESULTADOS\n")
1616 cat("===== \n")
1617 cat(sprintf("Painel: %d obs | %d prestadores | %d anos (%s a %s)\n",
1618             nrow(dados_sfa_painel),
1619             n_distinct(dados_sfa_painel$cod_prestador),
1620             n_distinct(dados_sfa_painel$ano),
1621             min(dados_sfa_painel$ano), max(dados_sfa_painel$ano)))
1622 cat(sprintf("Cobertura: %.1f milhões (%.1f%% do Brasil)\n",
1623             mean(cobertura_pop$pop_agua_amostra) / 1e6,
1624             mean(cobertura_pop$perc_cobertura)))
1625 cat(sprintf("Melhor modelo: AIC = %s | BIC = %s\n",
1626             comparacao$Modelo[which.min(comparacao$AIC)],
1627             comparacao$Modelo[which.min(comparacao$BIC)]))
1628 cat(sprintf("Eficiência TL TV: média = %.1f%% | mediana = %.1f%% | DP = %.1f%%\n",
1629             mean(dados_sfa_painel$ef_tl_tv, na.rm=TRUE)*100,
1630             median(dados_sfa_painel$ef_tl_tv, na.rm=TRUE)*100,
1631             sd(dados_sfa_painel$ef_tl_tv, na.rm=TRUE)*100))
1632
1633 cat("\n===== \n")
1634 cat("ANÁLISE SFA CONCLUÍDA COM SUCESSO!\n")
1635 cat("===== \n")
1636
1637 # =====
1638 # TESTES DE ROBUSTEZ - VARIAÇÃO DA AMOSTRA TEMPORAL
1639 # =====
1640
1641 cat("\n===== \n")
1642 cat("TESTES DE ROBUSTEZ - VARIAÇÃO DA AMOSTRA TEMPORAL\n")
1643 cat("===== \n")
1644
1645 # 22. PREPARAR AMOSTRAS ALTERNATIVAS ----
1646
1647 anos_disponiveis <- sort(unique(dados_sfa_final$ano))
1648 cat("\nAnos disponíveis:", paste(anos_disponiveis, collapse = ", "), "\n")
1649
1650 periodos_teste <- list(
1651   "2015-2022 (8 anos)" = as.character(2015:2022),
1652   "2017-2022 (6 anos)" = as.character(2017:2022),
1653   "2018-2022 (5 anos)" = as.character(2018:2022),
1654   "2019-2022 (4 anos)" = as.character(2019:2022),
1655   "2020-2022 (3 anos)" = as.character(2020:2022)
1656 )

```

```

1657
1658 analise_completude <- lapply(names(periodos_teste), function(perodo) {
1659   anos <- periodos_teste[[perodo]]
1660   dados_temp <- dados_sfa_final %>% filter(ano %in% anos)
1661   prestadores_completos <- dados_temp %>%
1662     group_by(cod_prestador) %>%
1663     summarise(n_anos = n_distinct(ano), .groups = "drop") %>%
1664     filter(n_anos == length(anos))
1665   data.frame(
1666     Perodo = perodo, Anos = length(anos),
1667     Obs_Disponiveis = nrow(dados_temp),
1668     Prestadores_Completos = nrow(prestadores_completos),
1669     Obs_Painel_Balanceado = nrow(prestadores_completos) * length(anos),
1670     stringsAsFactors = FALSE
1671   )
1672 }) %>% bind_rows()
1673
1674 cat("\n=== ANLISE DE COMPLETUDE POR PERÍODO ===\n")
1675 print(analise_completude)
1676
1677 periodos_viáveis <- analise_completude %>%
1678   filter(Prestadores_Completos >= 200) %>%
1679   arrange(desc(Anos))
1680
1681 if(nrow(periodos_viáveis) < 2) {
1682   periodos_viáveis <- analise_completude %>%
1683     filter(Prestadores_Completos >= 150) %>%
1684     arrange(desc(Anos))
1685 }
1686
1687 # Função reutilizável para criar painel de robustez
1688 criar_painel_robustez <- function(anos_selecionados, nome_amostra) {
1689   cat(sprintf("\n=== CRIANDO PAINEL: %s ===\n", nome_amostra))
1690   n_anos_necessarios <- length(anos_selecionados)
1691
1692   dados_painel <- dados_sfa_final %>%
1693     filter(ano %in% anos_selecionados) %>%
1694     group_by(cod_prestador) %>%
1695     filter(n_distinct(ano) == n_anos_necessarios) %>%
1696     ungroup() %>%
1697     mutate(
1698       VC_norm = VC / P_e,
1699       P_l_norm = P_l / P_e,
1700       P_o_norm = P_o / P_e
1701     ) %>%
1702     filter(
1703       VC_norm > 0, P_l_norm > 0, P_o_norm > 0,
1704       Y_agua > 0, Y_esgo > 0, K > 0,
1705       Dens > 0, !is.na(Sewerage)
1706     ) %>%
1707     group_by(cod_prestador) %>%
1708     filter(n_distinct(ano) == n_anos_necessarios) %>%
1709     ungroup() %>%
1710     mutate(
1711       lnVC = log(VC_norm),
1712       lnY_agua = log(Y_agua),
1713       lnY_esgo = log(Y_esgo),
1714       lnK = log(K),
1715       lnP_l = log(P_l_norm),
1716       lnP_o = log(P_o_norm),
1717       lnDens = log(Dens),
1718       # lnMicro: log(Micro + 0.01)  mesma translação do modelo principal
1719       lnMicro = log(Micro + 0.01)
1720     ) %>%
1721     mutate(
1722       lnVC_c = lnVC - mean(lnVC, na.rm = TRUE),
1723       lnY_agua_c = lnY_agua - mean(lnY_agua, na.rm = TRUE),

```

```

1724     lnY_esgo_c = lnY_esgo - mean(lnY_esgo, na.rm = TRUE),
1725     lnK_c      = lnK      - mean(lnK,      na.rm = TRUE),
1726     lnP_l_c    = lnP_l    - mean(lnP_l,    na.rm = TRUE),
1727     lnP_o_c    = lnP_o    - mean(lnP_o,    na.rm = TRUE)
1728   ) %>%
1729   mutate(
1730     lnP_l_sq    = 0.5 * lnP_l_c^2,    lnP_o_sq    = 0.5 * lnP_o_c^2,
1731     lnPl_Po     = lnP_l_c * lnP_o_c,
1732     lnY_agua_sq = 0.5 * lnY_agua_c^2, lnY_esgo_sq = 0.5 * lnY_esgo_c^2,
1733     lnYa_Ye     = lnY_agua_c * lnY_esgo_c,
1734     lnK_sq      = 0.5 * lnK_c^2,
1735     lnPl_Ya     = lnP_l_c * lnY_agua_c, lnPl_Ye = lnP_l_c * lnY_esgo_c,
1736     lnPo_Ya     = lnP_o_c * lnY_agua_c, lnPo_Ye = lnP_o_c * lnY_esgo_c,
1737     lnPl_K      = lnP_l_c * lnK_c,      lnPo_K  = lnP_o_c * lnK_c,
1738     lnYa_K      = lnY_agua_c * lnK_c,    lnYe_K  = lnY_esgo_c * lnK_c
1739   ) %>%
1740   filter(if_all(c(lnVC, lnY_agua, lnY_esgo, lnK,
1741                  lnP_l, lnP_o, lnDens, lnMicro), is.finite)) %>%
1742   group_by(cod_prestador) %>%
1743   filter(n_distinct(ano) == n_anos_necessarios) %>%
1744   ungroup() %>%
1745   arrange(cod_prestador, ano)
1746
1747   cat(sprintf(" Painei final: %d obs | %d prestadores\n",
1748              nrow(dados_painel), n_distinct(dados_painel$cod_prestador)))
1749   return(dados_painel)
1750 }
1751
1752 # 23. CRIAR E ESTIMAR AMOSTRAS ALTERNATIVAS ----
1753
1754 amostra_1 <- dados_sfa_painel
1755 nome_1    <- "2015-2022"
1756
1757 periodos_alternativos <- periodos_viáveis %>%
1758   filter(Periodo != "2015-2022 (8 anos)") %>%
1759   head(2)
1760
1761 if(nrow(periodos_alternativos) >= 1) {
1762   periodo_2_info <- periodos_alternativos[1, ]
1763   anos_2 <- periodos_teste[[periodo_2_info$Periodo]]
1764   nome_2 <- gsub(" \\([0-9]+ anos\\)", "", periodo_2_info$Periodo)
1765   amostra_2 <- criar_painel_robustez(anos_2, nome_2)
1766 } else {
1767   stop("Não há períodos alternativos viáveis para análise de robustez!")
1768 }
1769
1770 if(nrow(periodos_alternativos) >= 2) {
1771   periodo_3_info <- periodos_alternativos[2, ]
1772   anos_3 <- periodos_teste[[periodo_3_info$Periodo]]
1773   nome_3 <- gsub(" \\([0-9]+ anos\\)", "", periodo_3_info$Periodo)
1774   amostra_3 <- criar_painel_robustez(anos_3, nome_3)
1775 } else {
1776   cat("\n Usando subsample de prestadores grandes como terceira amostra\n")
1777   prestadores_grandes <- dados_sfa_painel %>%
1778     group_by(cod_prestador) %>%
1779     summarise(Y_agua_medio = mean(Y_agua, na.rm = TRUE), .groups = "drop") %>%
1780     filter(Y_agua_medio >= median(Y_agua_medio, na.rm = TRUE)) %>%
1781     pull(cod_prestador)
1782   amostra_3 <- dados_sfa_painel %>% filter(cod_prestador %in% prestadores_grandes)
1783   nome_3 <- "2015-2022 (Grandes)"
1784 }
1785
1786 estimar_tl_tv <- function(dados, nome_amostra) {
1787   cat(sprintf("\n=== Estimando: %s | Obs: %d | Prestadores: %d | Anos: %d ===\n",
1788              nome_amostra, nrow(dados),
1789              n_distinct(dados$cod_prestador), n_distinct(dados$ano)))
1790   dados_plm <- pdata.frame(dados, index = c("cod_prestador", "ano"))

```

```

1791   modelo <- frontier::sfa(
1792     formula = formula_tl, data = dados_plm,
1793     ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = TRUE
1794   )
1795   cat(sprintf(" LogLik: %.2f\n", as.numeric(logLik(modelo))))
1796   return(modelo)
1797 }
1798
1799 modelo_rob_1 <- modelo_4_tl_tv
1800 modelo_rob_2 <- estimar_tl_tv(amostra_2, nome_2)
1801 modelo_rob_3 <- estimar_tl_tv(amostra_3, nome_3)
1802
1803 # 24. EXTRAIR EFICIÊNCIAS DOS MODELOS DE ROBUSTEZ ----
1804
1805 amostra_1$ef_rob <- as.numeric(frontier::efficiencias(modelo_rob_1, asInData = TRUE))
1806 amostra_2$ef_rob <- as.numeric(frontier::efficiencias(modelo_rob_2, asInData = TRUE))
1807 amostra_3$ef_rob <- as.numeric(frontier::efficiencias(modelo_rob_3, asInData = TRUE))
1808
1809 calcular_cobertura <- function(dados, nome) {
1810   pop_media_amostra <- dados %>%
1811     group_by(ano) %>%
1812     summarise(pop = sum(ag001, na.rm = TRUE), .groups = "drop") %>%
1813     pull(pop) %>% mean()
1814   pop_media_brasil <- pop_brasil %>%
1815     filter(ano %in% unique(dados$ano)) %>%
1816     pull(pop_total_br) %>% mean()
1817   cobertura_pct <- (pop_media_amostra / pop_media_brasil) * 100
1818   cat(sprintf("%s: %.1f mi (%.1f%%)\n", nome, pop_media_amostra/1e6, cobertura_pct))
1819   list(pop_media = pop_media_amostra, cobertura_pct = cobertura_pct)
1820 }
1821
1822 cob_1 <- calcular_cobertura(amostra_1, nome_1)
1823 cob_2 <- calcular_cobertura(amostra_2, nome_2)
1824 cob_3 <- calcular_cobertura(amostra_3, nome_3)
1825
1826 # 25. CRIAR TABELA DE ROBUSTEZ em português ----
1827
1828 extrair_coef_rob <- function(modelo) {
1829   coef_est <- coef(modelo)
1830   coef_se <- sqrt(diag(vcov(modelo)))
1831   data.frame(
1832     estimate = coef_est,
1833     std.error = coef_se,
1834     statistic = coef_est / coef_se,
1835     p.value = 2 * (1 - pnorm(abs(coef_est / coef_se)))
1836   )
1837 }
1838
1839 coef_rob_1 <- extrair_coef_rob(modelo_rob_1)
1840 coef_rob_2 <- extrair_coef_rob(modelo_rob_2)
1841 coef_rob_3 <- extrair_coef_rob(modelo_rob_3)
1842
1843 extrair_params_rob <- function(modelo) {
1844   params <- coef(modelo)
1845   sigma_sq <- if("sigmaSq" %in% names(params)) params["sigmaSq"] else NA
1846   gamma <- if("gamma" %in% names(params)) params["gamma"] else NA
1847   if(!is.na(sigma_sq) && !is.na(gamma)) {
1848     lambda <- sqrt(gamma / (1 - gamma))
1849     sigma_u <- sqrt(gamma * sigma_sq)
1850     sigma_v <- sqrt((1 - gamma) * sigma_sq)
1851   } else { lambda <- sigma_u <- sigma_v <- NA }
1852   list(sigma_sq=sigma_sq, gamma=gamma, lambda=lambda, sigma_u=sigma_u, sigma_v=sigma_v)
1853 }
1854
1855 params_rob_1 <- extrair_params_rob(modelo_rob_1)
1856 params_rob_2 <- extrair_params_rob(modelo_rob_2)
1857 params_rob_3 <- extrair_params_rob(modelo_rob_3)

```

```

1858
1859 criar_linha_simples <- function(nome_var, var_name, c1, c2, c3) {
1860   get_cs <- function(coef_df, var) {
1861     if(var %in% rownames(coef_df)) {
1862       est <- coef_df[var,"estimate"]; se <- coef_df[var,"std.error"]
1863       p <- coef_df[var,"p.value"]
1864       stars <- ifelse(p<0.001,"***",ifelse(p<0.01,"**",ifelse(p<0.05,"*","")))
1865       list(coef = sprintf("%.4f%s", est, stars), se = sprintf("%.4f", se))
1866     } else { list(coef = "", se = "") }
1867   }
1868   r1<-get_cs(c1,var_name); r2<-get_cs(c2,var_name); r3<-get_cs(c3,var_name)
1869   sprintf("%s & %s & %s & %s & %s & %s & %s \\\\",
1870     nome_var, r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se)
1871 }
1872
1873 latex_doc <- c(
1874   "\\begin{table}[htbp]",
1875   "\\centering",
1876   "\\caption{Testes de Robustez: Modelo Translog Tempo-Variante com Diferentes Amostras}",
1877   "\\label{tab:robustez}",
1878   "\\footnotesize",
1879   "\\begin{tabular}{@{}lcccccc@{}}",
1880   "\\hline\\hline",
1881   sprintf("& \\multicolumn{2}{c}{%s} & \\multicolumn{2}{c}{%s} & \\multicolumn{2}{c}{%s} \\",
1882     nome_1, nome_2, nome_3),
1883   "\\cline{2-3} \\cline{4-5} \\cline{6-7}",
1884   "Variável & Coef. & EP & Coef. & EP & Coef. & EP \\\\",
1885   "\\hline"
1886 )
1887
1888 vars_latex <- list(
1889   # O intercepto é incluído por convenção, mas sua comparação entre colunas
1890   # é limitada: cada amostra tem seu próprio ponto de centralização.
1891   list(secao = "Parâmetro",
1892     vars = c("(Intercept)"),
1893     nomes = c("Intercepto")),
1894   list(secao = "Preços dos Insumos (normalizados por $P_E$)",
1895     vars = c("lnP_l_c", "lnP_o_c"),
1896     nomes = c("$\\ln(P_L / P_E)$", "$\\ln(P_O / P_E)$")),
1897   list(secao = "Produtos",
1898     vars = c("lnY_agua_c", "lnY_esgo_c"),
1899     nomes = c("$\\ln(Y_{\\text{água}})$", "$\\ln(Y_{\\text{esgoto}})$")),
1900   list(secao = "Capital Quasi-Fixo",
1901     vars = c("lnK_c"),
1902     nomes = c("$\\ln(K)$")),
1903   list(secao = "Termos Quadráticos Preços",
1904     vars = c("lnP_l_sq", "lnP_o_sq", "lnPl_Po"),
1905     nomes = c("$\\frac{1}{2}[\\ln(P_L / P_E)]^2$",
1906       "$\\frac{1}{2}[\\ln(P_O / P_E)]^2$",
1907       "$\\ln(P_L / P_E) \\times \\ln(P_O / P_E)$")),
1908   list(secao = "Termos Quadráticos Produtos e Capital",
1909     vars = c("lnY_agua_sq", "lnY_esgo_sq", "lnYa_Ye", "lnK_sq"),
1910     nomes = c("$\\frac{1}{2}[\\ln(Y_{\\text{água}})]^2$",
1911       "$\\frac{1}{2}[\\ln(Y_{\\text{esgoto}})]^2$",
1912       "$\\ln(Y_{\\text{água}}) \\times \\ln(Y_{\\text{esgoto}})$",
1913       "$\\frac{1}{2}[\\ln(K)]^2$")),
1914   list(secao = "Interações: Preço $\\times$ Produto",
1915     vars = c("lnPl_Ya", "lnPl_Ye", "lnPo_Ya", "lnPo_Ye"),
1916     nomes = c("$\\ln(P_L / P_E) \\times \\ln(Y_{\\text{água}})$",
1917       "$\\ln(P_L / P_E) \\times \\ln(Y_{\\text{esgoto}})$",
1918       "$\\ln(P_O / P_E) \\times \\ln(Y_{\\text{água}})$",
1919       "$\\ln(P_O / P_E) \\times \\ln(Y_{\\text{esgoto}})$")),
1920   list(secao = "Interações: Preço $\\times$ Capital",
1921     vars = c("lnPl_K", "lnPo_K"),
1922     nomes = c("$\\ln(P_L / P_E) \\times \\ln(K)$",
1923       "$\\ln(P_O / P_E) \\times \\ln(K)$")),

```

```

1924 list(secao = "Interações: Produto  $\times$  Capital",
1925       vars = c("lnYa_K", "lnYe_K"),
1926       nomes = c("$\\ln(Y_{\\text{água}}) \\times \\ln(K)$",
1927                "$\\ln(Y_{\\text{esgoto}}) \\times \\ln(K)$")),
1928 list(secao = "Variáveis de Controle",
1929       vars = c("lnDens", "lnMicro", "Loss", "Sewerage"),
1930       nomes = c("$\\ln(\\text{Densidade})$",
1931                "$\\ln(\\text{Micromedição})$",
1932                "Índice de Perdas",
1933                "Cobertura de Esgoto")),
1934 list(secao = "Variáveis Institucionais",
1935       vars = c("Treatment", "Independent", "Private", "Local", "Multi"),
1936       nomes = c("Tratamento de Esgoto",
1937                "Gestão Independente",
1938                "Propriedade Privada",
1939                "Prestador Local",
1940                "Múltiplos Prestadores")),
1941 list(secao = "Variáveis Regionais (ref.: Centro-Oeste)",
1942       vars = c("NO", "NE", "SE", "SO"),
1943       nomes = c("Norte", "Nordeste", "Sudeste", "Sul"))
1944 )
1945
1946 for(secao_info in vars_latex) {
1947   latex_doc <- c(latex_doc,
1948                 sprintf("\\multicolumn{7}{l}{\\textit{%s}} \\\\ ",
1949                         secao_info$secao),
1950                 "[0.3em]")
1951   for(i in seq_along(secao_info$vars)) {
1952     latex_doc <- c(latex_doc,
1953                   criar_linha_simples(secao_info$nomes[i], secao_info$vars[i],
1954                                       coef_rob_1, coef_rob_2, coef_rob_3))
1955   }
1956   latex_doc <- c(latex_doc, "[0.5em]")
1957 }
1958
1959 latex_doc <- c(latex_doc,
1960               "\\hline",
1961               "\\multicolumn{7}{l}{\\textit{Parâmetros de Variância}} \\\\ [0.3em]",
1962               sprintf("$\\sigma^2$ & \\multicolumn{2}{c}{%.4f} & \\multicolumn{2}{c}{%.4f} &
1963               ↪ \\multicolumn{2}{c}{%.4f} \\\\ ",
1964               params_rob_1$sigma_sq, params_rob_2$sigma_sq, params_rob_3$sigma_sq),
1965               sprintf("$\\gamma$ & \\multicolumn{2}{c}{%.4f} & \\multicolumn{2}{c}{%.4f} &
1966               ↪ \\multicolumn{2}{c}{%.4f} \\\\ ",
1967               params_rob_1$gamma, params_rob_2$gamma, params_rob_3$gamma),
1968               sprintf("$\\lambda$ & \\multicolumn{2}{c}{%.4f} & \\multicolumn{2}{c}{%.4f} &
1969               ↪ \\multicolumn{2}{c}{%.4f} \\\\ ",
1970               params_rob_1$lambda, params_rob_2$lambda, params_rob_3$lambda),
1971               "\\hline",
1972               "\\multicolumn{7}{l}{\\textit{Características da Amostra}} \\\\ [0.3em]",
1973               sprintf("Observações & \\multicolumn{2}{c}{%d} & \\multicolumn{2}{c}{%d} &
1974               ↪ \\multicolumn{2}{c}{%d} \\\\ ",
1975               nrow(amostra_1), nrow(amostra_2), nrow(amostra_3)),
1976               sprintf("Prestadores & \\multicolumn{2}{c}{%d} & \\multicolumn{2}{c}{%d} &
1977               ↪ \\multicolumn{2}{c}{%d} \\\\ ",
1978               n_distinct(amostra_1$cod_prestador),
1979               n_distinct(amostra_2$cod_prestador),
1980               n_distinct(amostra_3$cod_prestador)),
1981               sprintf("Anos & \\multicolumn{2}{c}{%d} & \\multicolumn{2}{c}{%d} &
1982               ↪ \\multicolumn{2}{c}{%d} \\\\ ",
1983               n_distinct(amostra_1$ano),
1984               n_distinct(amostra_2$ano),
1985               n_distinct(amostra_3$ano)),
1986               sprintf("Cobertura Pop.\\ (\\%) & \\multicolumn{2}{c}{%.1f} &
1987               ↪ \\multicolumn{2}{c}{%.1f} & \\multicolumn{2}{c}{%.1f} \\\\ ",
1988               cob_1$cobertura_pct, cob_2$cobertura_pct, cob_3$cobertura_pct),
1989               sprintf("Log-Verossimilhança & \\multicolumn{2}{c}{%.2f} &
1990               ↪ \\multicolumn{2}{c}{%.2f} & \\multicolumn{2}{c}{%.2f} \\\\ ",

```

```

1983         as.numeric(logLik(modelo_rob_1)),
1984         as.numeric(logLik(modelo_rob_2)),
1985         as.numeric(logLik(modelo_rob_3))),
1986     "\\hline",
1987     "\\multicolumn{7}{l}{\\textit{Estatísticas de Eficiência}} \\\\ [0.3em]",
1988     sprintf("Eficiência Média & \\multicolumn{2}{c}{%.4f} &
↪ \\multicolumn{2}{c}{%.4f} & \\multicolumn{2}{c}{%.4f} \\\\",
1989         mean(amostra_1$ef_rob, na.rm=TRUE),
1990         mean(amostra_2$ef_rob, na.rm=TRUE),
1991         mean(amostra_3$ef_rob, na.rm=TRUE)),
1992     sprintf("Eficiência Mediana & \\multicolumn{2}{c}{%.4f} &
↪ \\multicolumn{2}{c}{%.4f} & \\multicolumn{2}{c}{%.4f} \\\\",
1993         median(amostra_1$ef_rob, na.rm=TRUE),
1994         median(amostra_2$ef_rob, na.rm=TRUE),
1995         median(amostra_3$ef_rob, na.rm=TRUE)),
1996     "\\hline\\hline",
1997     "\\end{tabular}",
1998     "\\vspace{0.3cm}",
1999     "\\begin{minipage}{\\textwidth}",
2000     "\\footnotesize",
2001     "\\textit{Notas:} Erros padrão entre parênteses. * p < 0,05; ** p < 0,01;
↪ *** p < 0,001. ",
2002     "\\sigma^2 = \\sigma_u^2 + \\sigma_v^2; \\gamma = \\sigma_u^2 / \\sigma^2;
↪ \\lambda = \\sigma_u / \\sigma_v. ",
2003     "Variável dependente: ln(Custo Variável / P\\_Energia). CV = FN015. ",
2004     "Todos os modelos utilizam painéis balanceados com série temporal completa
↪ para cada prestador.",
2005     "\\end{minipage}",
2006     "\\end{table}"
2007 )
2008
2009 writeLines(latex_doc, file.path(dir_outputs, "tabela_7_robustez_amostras.tex"))
2010 cat(" Tabela 7 (Robustez em português) salva\\n")
2011
2012 # 26. SALVAR RESULTADOS DE ROBUSTEZ ----
2013
2014 write_csv(analise_completude, file.path(dir_outputs, "analise_completude_periodos.csv"))
2015
2016 stats_comp <- data.frame(
2017     Amostra      = c(nome_1, nome_2, nome_3),
2018     Período      = c(paste(range(amostra_1$ano), collapse="-"),
2019                     paste(range(amostra_2$ano), collapse="-"),
2020                     paste(range(amostra_3$ano), collapse="-")),
2021     N_obs        = c(nrow(amostra_1), nrow(amostra_2), nrow(amostra_3)),
2022     N_prestadores = c(n_distinct(amostra_1$cod_prestador),
2023                     n_distinct(amostra_2$cod_prestador),
2024                     n_distinct(amostra_3$cod_prestador)),
2025     Ef_Media_pct = c(mean(amostra_1$ef_rob, na.rm=TRUE)*100,
2026                     mean(amostra_2$ef_rob, na.rm=TRUE)*100,
2027                     mean(amostra_3$ef_rob, na.rm=TRUE)*100)
2028 )
2029
2030 write_csv(stats_comp, file.path(dir_outputs, "comparacao_amostras_robustez.csv"))
2031
2032 cat("\\n=====\\n")
2033 cat("RESUMO DOS TESTES DE ROBUSTEZ\\n")
2034 cat("=====\\n")
2035 cat(sprintf("AMOSTRA 1: %s | Obs: %d | Prestadores: %d | Eficiência: %.1f%%\\n",
2036             nome_1, nrow(amostra_1), n_distinct(amostra_1$cod_prestador),
2037             mean(amostra_1$ef_rob, na.rm=TRUE)*100))
2038 cat(sprintf("AMOSTRA 2: %s | Obs: %d | Prestadores: %d | Eficiência: %.1f%%\\n",
2039             nome_2, nrow(amostra_2), n_distinct(amostra_2$cod_prestador),
2040             mean(amostra_2$ef_rob, na.rm=TRUE)*100))
2041 cat(sprintf("AMOSTRA 3: %s | Obs: %d | Prestadores: %d | Eficiência: %.1f%%\\n",
2042             nome_3, nrow(amostra_3), n_distinct(amostra_3$cod_prestador),
2043             mean(amostra_3$ef_rob, na.rm=TRUE)*100))
2044

```

```

2045 cat("\n TESTES DE ROBUSTEZ CONCLUÍDOS!\n")
2046 cat("=====\n")
2047
2048 # =====
2049 # 27. FIGURAS 3 E 4 EVOLUO DA EFICIÊNCIA NAS AMOSTRAS DE ROBUSTEZ
2050 #     Mesmo layout da Figura 2, aplicado s amostras 2 e 3
2051 # =====
2052
2053 cat("
2054 === GERANDO FIGURAS 3 E 4: EFICIÊNCIA NAS AMOSTRAS DE ROBUSTEZ ===
2055 ")
2056
2057 # Função reutilizável: recebe o data frame da amostra (com ef_rob) e metadados
2058 gerar_figura_ef <- function(dados, nome_amostra, numero_figura) {
2059
2060   # Classificação por natureza jurídica (mesma da Figura 2)
2061   ef_inst <- dados %>%
2062     mutate(
2063       tipo = case_when(
2064         natureza_juridica == "Empresa privada"
2065           ~ "Privado",
2066         natureza_juridica == "Administração pública direta"
2067           ~ "Adm. Direta",
2068         natureza_juridica == "Autarquia"
2069           ~ "Autarquia",
2070         natureza_juridica %in% c(
2071           "Sociedade de economia mista com administração pública",
2072           "Empresa pública",
2073           "Sociedade de economia mista com administração predominantemente privada"
2074         ) ~ "Emp. Estatal/Mista"
2075       )
2076     ) %>%
2077     filter(!is.na(tipo)) %>%
2078     group_by(ano, tipo) %>%
2079     summarise(
2080       ef_media = mean(ef_rob, na.rm = TRUE),
2081       ef_q25   = quantile(ef_rob, 0.25, na.rm = TRUE),
2082       ef_q75   = quantile(ef_rob, 0.75, na.rm = TRUE),
2083       n        = n(),
2084       .groups  = "drop"
2085     ) %>%
2086     mutate(ano_num = as.numeric(as.character(ano)))
2087
2088   ef_geral <- dados %>%
2089     group_by(ano) %>%
2090     summarise(
2091       ef_media = mean(ef_rob, na.rm = TRUE),
2092       ef_q25   = quantile(ef_rob, 0.25, na.rm = TRUE),
2093       ef_q75   = quantile(ef_rob, 0.75, na.rm = TRUE),
2094       n        = n(),
2095       .groups  = "drop"
2096     ) %>%
2097     mutate(ano_num = as.numeric(as.character(ano)))
2098
2099   media_perodo <- mean(dados$ef_rob, na.rm = TRUE)
2100   anos_range <- paste(range(as.character(dados$ano)), collapse = "")
2101
2102   cores_inst <- c(
2103     "Emp. Estatal/Mista" = "#0077B6",
2104     "Autarquia"         = "#E85D04",
2105     "Adm. Direta"       = "#9B2226",
2106     "Privado"           = "#2DC653"
2107   )
2108   ltipos_inst <- c(
2109     "Emp. Estatal/Mista" = "solid",
2110     "Autarquia"         = "dashed",
2111     "Adm. Direta"       = "dotdash",

```

```

2112     "Privado"           = "dotted"
2113   )
2114   shapes_inst <- c(
2115     "Emp. Estatal/Mista" = 16,
2116     "Autarquia"         = 17,
2117     "Adm. Direta"       = 15,
2118     "Privado"           = 18
2119   )
2120
2121   fig <- ggplot() +
2122     geom_ribbon(
2123       data = ef_geral,
2124       aes(x = ano_num, ymin = ef_q25, ymax = ef_q75),
2125       fill = "grey80", alpha = 0.45
2126     ) +
2127     geom_line(
2128       data = ef_inst,
2129       aes(x = ano_num, y = ef_media, color = tipo, linetype = tipo),
2130       linewidth = 1.1
2131     ) +
2132     geom_point(
2133       data = ef_inst,
2134       aes(x = ano_num, y = ef_media, color = tipo, shape = tipo),
2135       size = 3
2136     ) +
2137     geom_line(
2138       data = ef_geral,
2139       aes(x = ano_num, y = ef_media),
2140       color = "black", linewidth = 1.5, linetype = "solid"
2141     ) +
2142     geom_point(
2143       data = ef_geral,
2144       aes(x = ano_num, y = ef_media),
2145       color = "black", size = 3, shape = 16
2146     ) +
2147     geom_hline(
2148       yintercept = media_perodo,
2149       linetype = "longdash", color = "grey50", linewidth = 0.6
2150     ) +
2151     annotate(
2152       "text",
2153       x = min(ef_geral$ano_num) + 0.1,
2154       y = media_perodo + 0.003,
2155       label = sprintf("Média geral = %.3f", media_perodo),
2156       size = 3, hjust = 0, color = "grey40", family = "serif"
2157     ) +
2158     scale_x_continuous(
2159       breaks = sort(unique(ef_geral$ano_num)),
2160       labels = as.character(sort(unique(ef_geral$ano_num)))
2161     ) +
2162     scale_y_continuous(
2163       labels = scales::percent_format(accuracy = 0.1)
2164     ) +
2165     scale_color_manual(name = "Tipo de Gestão", values = cores_inst) +
2166     scale_linetype_manual(name = "Tipo de Gestão", values = ltipos_inst) +
2167     scale_shape_manual(name = "Tipo de Gestão", values = shapes_inst) +
2168     labs(
2169       title = "Evolução da Eficiência Técnica nos Serviços de gua e Esgoto",
2170       subtitle = sprintf("Brasil, %s Amostra de Robustez: %s (%d prestadores)",
2171                           anos_range, nome_amostra,
2172                           n_distinct(dados$cod_prestador)),
2173       x = NULL,
2174       y = "Eficiência Técnica Média",
2175       caption = paste0(
2176         "Fonte: SNIS (", anos_range, "). Modelo: Fronteira Estocástica Translog
2177         ↪ Tempo-Variante.

```

```

2178     "Linha preta espessa: média geral. Faixa cinza: intervalo interquartil (P25P75). ",
2179     "Linha tracejada cinza: média do período completo."
2180 )
2181 ) +
2182 theme_classic(base_family = "serif", base_size = 11) +
2183 theme(
2184     plot.title           = element_text(size = 13, face = "bold", hjust = 0.5),
2185     plot.subtitle        = element_text(size = 10, hjust = 0.5, color = "grey40",
2186                                     margin = margin(b = 8)),
2187     plot.caption         = element_text(size = 8, hjust = 0, color = "grey50"),
2188     axis.title.y         = element_text(margin = margin(r = 8)),
2189     legend.position      = "bottom",
2190     legend.title         = element_text(size = 9, face = "bold"),
2191     legend.text          = element_text(size = 8),
2192     legend.key.width     = unit(1.5, "cm"),
2193     panel.grid.major.y   = element_line(color = "grey92", linewidth = 0.4),
2194     panel.grid.major.x   = element_blank(),
2195     plot.margin          = margin(10, 15, 10, 10)
2196 )
2197
2198 # Salvar
2199 nome_arquivo <- sprintf("figura_%d_eficiencia_robustez_%s",
2200                         numero_figura,
2201                         gsub("[^a-zA-Z0-9]", "_", nome_amostra))
2202
2203 ggsave(
2204     filename = file.path(dir_outputs, paste0(nome_arquivo, ".pdf")),
2205     plot = fig, width = 18, height = 12, units = "cm", dpi = 300
2206 )
2207 ggsave(
2208     filename = file.path(dir_outputs, paste0(nome_arquivo, ".png")),
2209     plot = fig, width = 18, height = 12, units = "cm", dpi = 300
2210 )
2211
2212 cat(sprintf(" Figura %d (%s) salva
2213 ", numero_figura, nome_amostra))
2214 return(fig)
2215 }
2216
2217 # Gerar para amostra 2 e amostra 3
2218 figura_3 <- gerar_figura_ef(amostra_2, nome_2, 3)
2219 figura_4 <- gerar_figura_ef(amostra_3, nome_3, 4)
2220
2221 cat(" Figuras 3 e 4 (Robustez) salvas
2222 ")
2223
2224 # =====
2225 #  MODELOS ALTERNATIVOS PARA TABELA DE ROBUSTEZ DE ESPECIFICACAO
2226 #  (Variáveis institucionais mantidas em todas as especificações)
2227 #  =====
2228
2229 cat("\n=== ESTIMANDO MODELOS ALTERNATIVOS (SEM CONTROLES E/OU REGIONAIS) ===\n")
2230 cat("(Variáveis institucionais mantidas em todas as especificações)\n")
2231
2232 # Fórmula COMPLETA (já estimada)
2233 formula_completa <- lnVC ~
2234     lnP_l_c + lnP_o_c +
2235     lnY_agua_c + lnY_esgo_c + lnK_c +
2236     lnP_l_sq + lnP_o_sq + lnPl_Po +
2237     lnY_agua_sq + lnY_esgo_sq + lnYa_Ye +
2238     lnK_sq +
2239     lnPl_Ya + lnPl_Ye +
2240     lnPo_Ya + lnPo_Ye +
2241     lnPl_K + lnPo_K +
2242     lnYa_K + lnYe_K +
2243     lnDens + lnMicro + Loss + Sewerage + # controles estruturais
2244     Treatment + Independent + Private + Local + Multi + # institucionais (todas)

```

```

2245 NO + NE + SE + SO # regionais
2246
2247 # Fórmula SEM CONTROLES ESTRUTURAIS (mantém regionais e institucionais)
2248 formula_sem_controles <- lnVC ~
2249   lnP_l_c + lnP_o_c +
2250   lnY_agua_c + lnY_esgo_c + lnK_c +
2251   lnP_l_sq + lnP_o_sq + lnPl_Po +
2252   lnY_agua_sq + lnY_esgo_sq + lnYa_Ye +
2253   lnK_sq +
2254   lnPl_Ya + lnPl_Ye +
2255   lnPo_Ya + lnPo_Ye +
2256   lnPl_K + lnPo_K +
2257   lnYa_K + lnYe_K +
2258   Treatment + Independent + Private + Local + Multi + # institucionais (todas)
2259 NO + NE + SE + SO # regionais
2260
2261 # Fórmula PARCIMONIOSA (sem controles e sem regionais, apenas institucionais)
2262 formula_parcimoniosa <- lnVC ~
2263   lnP_l_c + lnP_o_c +
2264   lnY_agua_c + lnY_esgo_c + lnK_c +
2265   lnP_l_sq + lnP_o_sq + lnPl_Po +
2266   lnY_agua_sq + lnY_esgo_sq + lnYa_Ye +
2267   lnK_sq +
2268   lnPl_Ya + lnPl_Ye +
2269   lnPo_Ya + lnPo_Ye +
2270   lnPl_K + lnPo_K +
2271   lnYa_K + lnYe_K +
2272   Treatment + Independent + Private + Local + Multi # institucionais (todas)
2273
2274 cat("\n1. Estimando modelo COMPLETO (já estimado)...\n")
2275 modelo_completo <- modelo_4_tl_tv
2276
2277 cat("2. Estimando modelo SEM CONTROLES ESTRUTURAIS...\n")
2278 dados_plm <- pdata.frame(dados_sfa_painel, index = c("cod_prestador", "ano"))
2279 modelo_sem_controles <- frontier::sfa(
2280   formula = formula_sem_controles, data = dados_plm,
2281   ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = TRUE
2282 )
2283
2284 cat("3. Estimando modelo PARCIMONIOSO (sem controles e sem regionais)...\n")
2285 modelo_parcimonioso <- frontier::sfa(
2286   formula = formula_parcimoniosa, data = dados_plm,
2287   ineffDecrease = FALSE, truncNorm = FALSE, timeEffect = TRUE
2288 )
2289
2290 cat(" Todos os modelos alternativos estimados com sucesso!\n")
2291
2292 # Extrair coeficientes e erros padrão
2293 coef_completo <- extrair_coef(modelo_completo)
2294 coef_sem_controles <- extrair_coef(modelo_sem_controles)
2295 coef_parcimonioso <- extrair_coef(modelo_parcimonioso)
2296
2297 # Extrair parâmetros de variância
2298 params_completo <- extrair_params_variancia(modelo_completo)
2299 params_sem_controles <- extrair_params_variancia(modelo_sem_controles)
2300 params_parcimonioso <- extrair_params_variancia(modelo_parcimonioso)
2301
2302 # Eficiências
2303 dados_sfa_painel$ef_completo <- as.numeric(frontier::eficiencias(modelo_completo, asInData =
2304   ↪ TRUE))
2305 dados_sfa_painel$ef_sem_controles <- as.numeric(frontier::eficiencias(modelo_sem_controles,
2306   ↪ asInData = TRUE))
2307 dados_sfa_painel$ef_parcimonioso <- as.numeric(frontier::eficiencias(modelo_parcimonioso,
2308   ↪ asInData = TRUE))
2309
2310 # =====
2311 # TABELA DE RESULTADOS COM 3 ESPECIFICAES (TODAS AS VARIVEIS)

```

```

2309 # =====
2310
2311 cat("\n=== GERANDO TABELA COMPARATIVA (3 ESPECIFICAÇÕES TODAS AS VARÍVEIS) ===\n")
2312
2313
2314 # Função para extrair coeficiente com estrelas
2315 extrair_coef_stars <- function(coef_df, var_name) {
2316   if(var_name %in% rownames(coef_df)) {
2317     est <- coef_df[var_name, "estimate"]
2318     se <- coef_df[var_name, "std.error"]
2319     p <- coef_df[var_name, "p.value"]
2320     stars <- ifelse(p < 0.001, "***", ifelse(p < 0.01, "**", ifelse(p < 0.05, "*", "")))
2321     list(coef = sprintf("%.4f%s", est, stars), se = sprintf("%.4f", se))
2322   } else {
2323     list(coef = "", se = "")
2324   }
2325 }
2326
2327 # Função para criar linha da tabela para 3 especificações
2328 criar_linha_3esp <- function(nome_var, var_name, coef1, coef2, coef3) {
2329   r1 <- extrair_coef_stars(coef1, var_name)
2330   r2 <- extrair_coef_stars(coef2, var_name)
2331   r3 <- extrair_coef_stars(coef3, var_name)
2332
2333   # Ajuste de espaçamento para variáveis categóricas (com \hspace{1em})
2334   if(grepl("hspace", nome_var)) {
2335     sprintf("%s & %s & %s & %s & %s & %s & %s \\\\",
2336       nome_var, r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se)
2337   } else {
2338     sprintf("%s & %s & %s & %s & %s & %s & %s \\\\",
2339       nome_var, r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se)
2340   }
2341 }
2342
2343 # Lista completa de variáveis (mesma ordem da tabela original)
2344 # Cada elemento: c(nome_display, nome_var_no_coef)
2345 vars_completas <- list(
2346   # 1 Intercepto
2347   list("Intercepto", "(Intercept)"),
2348   # 2-3 Parâmetros da fronteira Preços
2349   list("\hspace{1em}$\\ln(P_L / P_E)$", "lnP_l_c"),
2350   list("\hspace{1em}$\\ln(P_O / P_E)$", "lnP_o_c"),
2351   # 4-5 Parâmetros da fronteira Produtos
2352   list("\hspace{1em}$\\ln(Y_{\\text{água}})$", "lnY_agua_c"),
2353   list("\hspace{1em}$\\ln(Y_{\\text{esgoto}})$", "lnY_esgo_c"),
2354   # 6 Parâmetros da fronteira Capital
2355   list("\hspace{1em}$\\ln(\\overline{K})$", "lnK_c"),
2356   # 7-9 Termos Quadráticos Preços (_jj e _jk)
2357   list("\hspace{1em}$\\frac{1}{2}[\\ln(P_L / P_E)]^2$",
2358     "lnP_l_sq"),
2359   list("\hspace{1em}$\\frac{1}{2}[\\ln(P_O / P_E)]^2$",
2360     "lnP_o_sq"),
2361   list("\hspace{1em}$\\ln(P_L / P_E) \\times \\ln(P_O / P_E)$",
2362     "lnPl_Po"),
2363   # 10-13 Termos Quadráticos Produtos e Capital
2364   list("\hspace{1em}$\\frac{1}{2}[\\ln(Y_{\\text{água}})]^2$",
2365     "lnY_agua_sq"),
2366   list("\hspace{1em}$\\frac{1}{2}[\\ln(Y_{\\text{esgoto}})]^2$",
2367     "lnY_esgo_sq"),
2368   list("\hspace{1em}$\\ln(Y_{\\text{água}}) \\times \\ln(Y_{\\text{esgoto}})$",
2369     "lnYa_Ye"),
2370   list("\hspace{1em}$\\frac{1}{2}[\\ln(\\overline{K})]^2$",
2371     "lnK_sq"),
2372   # 14-17 Interações Preço Produto
2373   list("\hspace{1em}$\\ln(P_L / P_E) \\times \\ln(Y_{\\text{água}})$", "lnPl_Ya"),
2374   list("\hspace{1em}$\\ln(P_L / P_E) \\times \\ln(Y_{\\text{esgoto}})$", "lnPl_Ye"),
2375   list("\hspace{1em}$\\ln(P_O / P_E) \\times \\ln(Y_{\\text{água}})$", "lnPo_Ya"),

```

```

2376 list("\\hspace{1em}$\\ln(P_O / P_E) \\times \\ln(Y_{\\text{esgoto}})$", "lnPo_Ye"),
2377 # 18-19 Interações Preço Capital
2378 list("\\hspace{1em}$\\ln(P_L / P_E) \\times \\ln(\\overline{K})$", "lnPl_K"),
2379 list("\\hspace{1em}$\\ln(P_O / P_E) \\times \\ln(\\overline{K})$", "lnPo_K"),
2380 # 20-21 Interações Produto Capital
2381 list("\\hspace{1em}$\\ln(Y_{\\text{água}}) \\times \\ln(\\overline{K})$", "lnYa_K"),
2382 list("\\hspace{1em}$\\ln(Y_{\\text{esgoto}}) \\times \\ln(\\overline{K})$", "lnYe_K"),
2383 # 22-25 Controles Estruturais (apenas na especificação completa)
2384 list("\\hspace{1em}$\\ln(\\text{Densidade})$", "lnDens"),
2385 list("\\hspace{1em}$\\ln(\\text{Micromedicação})$", "lnMicro"),
2386 list("\\hspace{1em}Índice de Perdas", "Loss"),
2387 list("\\hspace{1em}Cobertura de Esgoto", "Sewerage"),
2388 # 26-30 Dummies Institucionais
2389 list("\\hspace{1em}Tratamento de Esgoto", "Treatment"),
2390 list("\\hspace{1em}Gestão Independente", "Independent"),
2391 list("\\hspace{1em}Propriedade Privada", "Private"),
2392 list("\\hspace{1em}Prestador Local", "Local"),
2393 list("\\hspace{1em}Múltiplos Prestadores", "Multi"),
2394 # 31-34 Dummies Regionais (ref.: Centro-Oeste)
2395 list("\\hspace{1em}Norte", "NO"),
2396 list("\\hspace{1em}Nordeste", "NE"),
2397 list("\\hspace{1em}Sudeste", "SE"),
2398 list("\\hspace{1em}Sul", "SO")
2399 )
2400
2401 # Coeficientes disponíveis em cada especificação
2402 # (para preencher com "---" quando não incluídos)
2403 variaveis_apenas_completa <- c("lnDens", "lnMicro", "Loss", "Sewerage")
2404 variaveis_apenas_completa_e_2 <- c("NO", "NE", "SE", "SO")
2405
2406 # Construir documento LaTeX linha por linha
2407 latex_3esp <- c(
2408   "\\begin{table}[htbp]",
2409   "\\centering",
2410   "\\caption{Resultados da Análise de Fronteira Estocástica Comparação de Especificações}",
2411   "\\label{tab:sfa_comparacao_especificacoes}",
2412   "\\fontsize{10}{12}\\selectfont",
2413   "\\begin{tabular}{@{}lcccc@{}}",
2414   "\\toprule",
2415   "& \\multicolumn{2}{c}{(1) Completa} & \\multicolumn{2}{c}{(2) Sem controles} &
2416   & \\multicolumn{2}{c}{(3) Parcimoniosa} & \\",
2417   "& \\multicolumn{2}{c}{(com regionais)} & \\multicolumn{2}{c}{(apenas regionais)} &
2418   & \\multicolumn{2}{c}{(sem regionais)} & \\",
2419   "\\cline{2-3} \\cline{4-5} \\cline{6-7}",
2420   "Variável & Estimativa & EP & Estimativa & EP & Estimativa & EP \\",
2421   "\\midrule"
2422 )
2423
2424 # Adicionar Intercepto
2425 r1 <- extrair_coef_stars(coef_completo, "(Intercept)")
2426 r2 <- extrair_coef_stars(coef_sem_controles, "(Intercept)")
2427 r3 <- extrair_coef_stars(coef_parcimonioso, "(Intercept)")
2428 latex_3esp <- c(latex_3esp,
2429   sprintf("Intercepto & %s & %s & %s & %s & %s & %s \\",
2430     r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2431 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2432
2433 # Adicionar Parâmetros da Fronteira Preços
2434 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Parâmetros da Fronteira Preços}}
2435   & \\")
2436 for(v in vars_completas[2:3]) {
2437   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2438   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2439   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2440   latex_3esp <- c(latex_3esp,
2441     sprintf("%s & %s & %s & %s & %s & %s & %s \\",
2442       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))

```

```

2440 }
2441
2442 # Adicionar Parâmetros da Fronteira Produtos
2443 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2444 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Parâmetros da Fronteira Produtos}}
    ↳ \\\")
2445 for(v in vars_completas[4:5]) {
2446   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2447   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2448   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2449   latex_3esp <- c(latex_3esp,
2450     sprintf("%s & %s & %s & %s & %s & %s & %s \\\",
2451       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2452 }
2453
2454 # Adicionar Parâmetros da Fronteira Capital
2455 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2456 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Parâmetros da Fronteira Capital}}
    ↳ \\\")
2457 v <- vars_completas[[6]]
2458 r1 <- extrair_coef_stars(coef_completo, v[[2]])
2459 r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2460 r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2461 latex_3esp <- c(latex_3esp,
2462   sprintf("%s & %s & %s & %s & %s & %s & %s \\\",
2463     v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2464
2465 # Adicionar Termos Quadráticos Preços
2466 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2467 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Termos Quadráticos Preços}} \\\")
2468 for(v in vars_completas[7:9]) {
2469   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2470   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2471   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2472   latex_3esp <- c(latex_3esp,
2473     sprintf("%s & %s & %s & %s & %s & %s & %s \\\",
2474       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2475 }
2476
2477 # Adicionar Termos Quadráticos Produtos e Capital
2478 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2479 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Termos Quadráticos Produtos e
    ↳ Capital}} \\\")
2480 for(v in vars_completas[10:13]) {
2481   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2482   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2483   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2484   latex_3esp <- c(latex_3esp,
2485     sprintf("%s & %s & %s & %s & %s & %s & %s \\\",
2486       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2487 }
2488
2489 # Adicionar Interações Preço Produto
2490 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2491 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Interações Preço $\\times$
    ↳ Produto}} \\\")
2492 for(v in vars_completas[14:17]) {
2493   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2494   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2495   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2496   latex_3esp <- c(latex_3esp,
2497     sprintf("%s & %s & %s & %s & %s & %s & %s \\\",
2498       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2499 }
2500
2501 # Adicionar Interações Preço Capital
2502 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")

```

```

2503 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Interações Preço $\\times$"}
      ↪ Capital)} \\")
2504 for(v in vars_completas[18:19]) {
2505   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2506   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2507   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2508   latex_3esp <- c(latex_3esp,
2509     sprintf("%s & %s & %s & %s & %s & %s & %s \\",
2510       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2511 }
2512
2513 # Adicionar Interações Produto Capital
2514 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2515 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Interações Produto $\\times$"}
      ↪ Capital)} \\")
2516 for(v in vars_completas[20:21]) {
2517   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2518   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2519   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2520   latex_3esp <- c(latex_3esp,
2521     sprintf("%s & %s & %s & %s & %s & %s & %s \\",
2522       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2523 }
2524
2525 # Adicionar Controles Estruturais (apenas na especificação 1)
2526 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2527 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Controles Estruturais}} \\")
2528 for(v in vars_completas[22:25]) {
2529   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2530   # Especificações 2 e 3: não incluem controles estruturais
2531   if(v[[2]] %in% variaveis_apenas_completa) {
2532     latex_3esp <- c(latex_3esp,
2533       sprintf("%s & %s & %s & --- & --- & --- & --- \\",
2534         v[[1]], r1$coef, r1$se))
2535   } else {
2536     r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2537     r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2538     latex_3esp <- c(latex_3esp,
2539       sprintf("%s & %s & %s & %s & %s & %s & %s \\",
2540         v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2541   }
2542 }
2543
2544 # Adicionar Dummies Institucionais
2545 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2546 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Dummies Institucionais}} \\")
2547 for(v in vars_completas[26:30]) {
2548   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2549   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2550   r3 <- extrair_coef_stars(coef_parcimonioso, v[[2]])
2551   latex_3esp <- c(latex_3esp,
2552     sprintf("%s & %s & %s & %s & %s & %s & %s \\",
2553       v[[1]], r1$coef, r1$se, r2$coef, r2$se, r3$coef, r3$se))
2554 }
2555
2556 # Adicionar Dummies Regionais (apenas nas especificações 1 e 2)
2557 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2558 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Dummies Regionais (ref.:"}
      ↪ Centro-Oeste)} \\")
2559 for(v in vars_completas[31:34]) {
2560   r1 <- extrair_coef_stars(coef_completo, v[[2]])
2561   r2 <- extrair_coef_stars(coef_sem_controles, v[[2]])
2562   # Especificação 3: não inclui regionais
2563   latex_3esp <- c(latex_3esp,
2564     sprintf("%s & %s & %s & %s & %s & --- & --- \\",
2565       v[[1]], r1$coef, r1$se, r2$coef, r2$se))
2566 }

```

```

2567
2568 # Adicionar Parâmetros de Variância
2569 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2570 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Parâmetros de Variância}} \\")
2571 latex_3esp <- c(latex_3esp,
2572   sprintf("\\hspace{1em}$\\sigma^2$ & %.4f & & %.4f & & %.4f & \\",
2573     params_completo$sigma_sq, params_sem_controles$sigma_sq,
2574     ↪ params_parcimonioso$sigma_sq),
2575   sprintf("\\hspace{1em}$\\gamma$ & %.4f & & %.4f & & %.4f & \\",
2576     params_completo$gamma, params_sem_controles$gamma,
2577     ↪ params_parcimonioso$gamma),
2578   sprintf("\\hspace{1em}$\\lambda = \\sigma_u / \\sigma_v$ & %.4f & & %.4f & &
2579     ↪ %.4f & \\",
2580     params_completo$lambda, params_sem_controles$lambda,
2581     ↪ params_parcimonioso$lambda),
2582   sprintf("\\hspace{1em}$\\sigma_u$ & %.4f & & %.4f & & %.4f & \\",
2583     params_completo$sigma_u, params_sem_controles$sigma_u,
2584     ↪ params_parcimonioso$sigma_u),
2585   sprintf("\\hspace{1em}$\\sigma_v$ & %.4f & & %.4f & & %.4f & \\",
2586     params_completo$sigma_v, params_sem_controles$sigma_v,
2587     ↪ params_parcimonioso$sigma_v))
2588
2589 # Adicionar Estatísticas do Modelo
2590 latex_3esp <- c(latex_3esp, "\\addlinespace[0.3em]")
2591 latex_3esp <- c(latex_3esp, "\\multicolumn{7}{l}{\\textbf{Estatísticas do Modelo}} \\")
2592 latex_3esp <- c(latex_3esp,
2593   sprintf("\\hspace{1em}Observações & %d & & %d & & %d & \\",
2594     nrow(dados_sfa_painel), nrow(dados_sfa_painel),
2595     ↪ nrow(dados_sfa_painel)),
2596   sprintf("\\hspace{1em}Prestadores & %d & & %d & & %d & \\",
2597     n_distinct(dados_sfa_painel$cod_prestador),
2598     n_distinct(dados_sfa_painel$cod_prestador),
2599     n_distinct(dados_sfa_painel$cod_prestador)),
2600   sprintf("\\hspace{1em}Anos & %d & & %d & & %d & \\",
2601     n_distinct(dados_sfa_painel$ano),
2602     n_distinct(dados_sfa_painel$ano),
2603     n_distinct(dados_sfa_painel$ano)),
2604   sprintf("\\hspace{1em}Log-Verossimilhança & %.2f & & %.2f & & %.2f & \\",
2605     as.numeric(logLik(modelo_completo)),
2606     as.numeric(logLik(modelo_sem_controles)),
2607     as.numeric(logLik(modelo_parcimonioso))),
2608   sprintf("\\hspace{1em}AIC & %.2f & & %.2f & & %.2f & \\",
2609     AIC(modelo_completo), AIC(modelo_sem_controles),
2610     ↪ AIC(modelo_parcimonioso)),
2611   sprintf("\\hspace{1em}BIC & %.2f & & %.2f & & %.2f & \\",
2612     BIC(modelo_completo), BIC(modelo_sem_controles),
2613     ↪ BIC(modelo_parcimonioso)),
2614   sprintf("\\hspace{1em}Eficiência Técnica Média & %.4f & & %.4f & & %.4f &
2615     ↪ \\",
2616     mean(dados_sfa_painel$ef_completo, na.rm=TRUE),
2617     mean(dados_sfa_painel$ef_sem_controles, na.rm=TRUE),
2618     mean(dados_sfa_painel$ef_parcimonioso, na.rm=TRUE)))
2619
2620 # Fechar tabela
2621 latex_3esp <- c(latex_3esp,
2622   "\\bottomrule",
2623   "\\end{tabular}",
2624   "\\vspace{0.3cm}",
2625   "\\begin{minipage}{\\textwidth}",
2626   "\\footnotesize",
2627   "\\textit{Notas:} Erros padrão entre parênteses. * p $<$ 0,05; ** p $<$ 0,01;
2628     ↪ *** p $<$ 0,001. ",
2629   "Variável dependente: $\\ln(CV / P_E)$ . CV = FN015. ",
2630   "Preços normalizados pelo preço da energia ($P_E$) como numerário. ",
2631   "Todas as variáveis explicativas contínuas foram centralizadas na média
2632     ↪ amostral. ",

```

```

2621         "Especificação (1): modelo completo com controles estruturais e dummies
↪ regionais. ",
2622         "Especificação (2): sem controles estruturais, mantém dummies regionais. ",
2623         "Especificação (3): sem controles estruturais e sem dummies regionais. ",
2624         "As variáveis institucionais são mantidas em todas as especificações. ",
2625         "Fonte: elaboração própria.",
2626         "\\end{minipage}",
2627         "\\end{table}")
2628
2629 writeLines(latex_3esp, file.path(dir_outputs, "tabela_comparacao_especificacoes.tex"))
2630 cat(" Tabela comparativa (3 especificações todas as variáveis) salva\\n")

```

B.3 Diferença em diferenças escalonado

Código B.3 – Diferença em diferenças escalonado

```

1  # =====
2  # PIPELINE COMPLETO  CONCESSIONES DE SANEAMENTO
3  # Artigo: "Participação privada no mercado de água reduz tarifas?
4  #       Evidências de uma reforma legal brasileira"
5  #
6  # ÍNDICE (Ctrl+Shift+O):
7  #   0. CONFIGURAO INICIAL
8  #   1. EXTRAO DE EQUIVALÊNCIAS SINISA-SNIS
9  #   2. CARREGAR E CONSOLIDAR DADOS SNIS POR MUNICÍPIO
10 #   3. CALCULAR INDICADORES SNIS (nomes IN)
11 #   4. IDENTIFICAR MUNICÍPIOS COM MLTIPLS PRESTADORES
12 #   5. LIMPEZA DE OUTLIERS
13 #   6. VALIDAO COM SNIS ORIGINAL (quando disponível)
14 #   7. CARREGAR SINISA 2023-2024 E MONTAR PAINEL 2017-2024
15 #   8. DICIONRIO E VARIVEL DE TRATAMENTO (Abcon + CORSAN)
16 #   9. BASE ANO-BASE 2017 (SNIS rawdata)
17 #  10. COVARIADAS (PIB, SICONFI, PNSB)
18 #  11. MONTAR BASE PSM COM TRATAMENTO
19 #  12. COMPARAO DE ESTRATGIAS DE PAREAMENTO (Completo / Rigoroso / Flexível)
20 #  13. PAREAMENTO PSM  ESTRATGIA FLEXÍVEL (ratios 1, 2 e 5)
21 #  14. TABELAS DE BALANO (LaTeX)
22 #  15. MAPA DA AMOSTRA PAREADA
23 #  16. MONTAR PAINIS DID (ratio 1, 2 e 5)
24 #  17. GRFICOS DE TENDÊNCIA (IAG0001, IFA1001, IFA2008)
25 #  18. ESTATÍSTICAS DESCRITIVAS POR GRUPO
26 #  19. ESTIMAO CALLAWAY & SANT'ANNA
27 #  20. TABELAS INDIVIDUAIS (IAG0001, IFA1001, IFA2008)
28 #  21. TABELAS AGREGADAS POR GRUPO (IAG, IFA)
29 #  22. ROBUSTEZ DID  RATIO 1, 2 E 5
30 #  23. TABELAS LATEX GERAIS (equivalências, validação, descritivas)
31 # =====
32
33
34 # =====
35 # 0. CONFIGURAO INICIAL ----
36 # =====
37
38 rm(list = ls())
39 gc()
40
41 options(scipen = 999)
42
43 library(tidyverse)
44 library(readxl)
45 library(readODS)
46 library(MatchIt)

```

```

47 library(did)
48 library(kableExtra)
49 library(sf)
50 library(geobr)
51 library(data.table)
52
53 setwd("C:/Users/Pedro Ricelly/Google Drive/Doutorado/Tese/Artigo 2")
54
55 dir.create("outputs",          showWarnings = FALSE)
56 dir.create("outputs/did",      showWarnings = FALSE, recursive = TRUE)
57 dir.create("outputs/did/plots", showWarnings = FALSE, recursive = TRUE)
58 dir.create("data",            showWarnings = FALSE)
59
60 cat("\n", rep("=", 80), "\n", sep = "")
61 cat("PIPELINE COMPLETO  CONCESSÕES DE SANEAMENTO\n")
62 cat(rep("=", 80), "\n\n", sep = "")
63
64
65 # =====
66 # 1. EXTRAÇÃO DE EQUIVALÊNCIAS SINISA-SNIS ----
67 # =====
68
69 cat(rep("=", 80), "\n", sep = "")
70 cat("PARTE 1: EXTRAINDO EQUIVALÊNCIAS SINISA-SNIS\n")
71 cat(rep("=", 80), "\n\n", sep = "")
72
73 extrair_equivalencias <- function(arquivo) {
74
75   cat("Processando:", basename(arquivo), "\n")
76   todas_planilhas <- excel_sheets(arquivo)
77   planilhas_pares <- todas_planilhas[seq(2, length(todas_planilhas), by = 2)]
78   resultados <- list()
79
80   for (planilha in planilhas_pares) {
81     df <- read_excel(arquivo, sheet = planilha)
82     col_codigo <- names(df)[str_detect(names(df), regex("C.DIGO", ignore_case =
83   ↪ TRUE))][1]
84     col_snis <- names(df)[str_detect(names(df), regex("INDICADOR.*SNIS", ignore_case =
85   ↪ TRUE))][1]
86
87     if (!is.na(col_codigo) && !is.na(col_snis)) {
88       for (i in seq_len(nrow(df))) {
89         codigo_sinisa <- df[[col_codigo]][i]
90         snis_completo <- df[[col_snis]][i]
91         if (!is.na(snis_completo) && str_detect(snis_completo, "IN0")) {
92           codigo_snis <- str_extract(snis_completo, "IN0\\d+[A-Z]*")
93           if (!is.na(codigo_snis)) {
94             col_nome <- names(df)[str_detect(names(df), regex("NOME.*INDICADOR",
95   ↪ ignore_case = TRUE))][1]
96             col_grupo <- names(df)[str_detect(names(df), regex("GRUPO",
97   ↪ ignore_case = TRUE))][1]
98             col_unidade <- names(df)[str_detect(names(df), regex("UNIDADE",
99   ↪ ignore_case = TRUE))][1]
100             resultados[[length(resultados) + 1]] <- data.frame(
101               codigo_sinisa = codigo_sinisa,
102               codigo_snis = codigo_snis,
103               descricao_completa = snis_completo,
104               nome_indicador = ifelse(!is.na(col_nome), df[[col_nome]][i], NA),
105               grupo = ifelse(!is.na(col_grupo), df[[col_grupo]][i], NA),
106               unidade = ifelse(!is.na(col_unidade), df[[col_unidade]][i], NA),
107               stringsAsFactors = FALSE
108             )
109           }
110         }
111       }
112     }
113   }
114 }

```

```

109
110 df_final <- bind_rows(resultados) %>%
111   mutate(grupo = str_replace_all(grupo, "\\n", " ")) %>%
112   distinct(codigo_sinisa, codigo_snis, .keep_all = TRUE) %>%
113   arrange(codigo_snis)
114
115   cat(" Equivalências encontradas:", nrow(df_final), "\n\n")
116   return(df_final)
117 }
118
119 equivalencias_agua <- extrair_equivalencias(
120   "data/INDICADORES_SINISA_ABASTECIMENTO_DE_AGUA_2024_v2.xlsx"
121 )
122 write_csv(equivalencias_agua, "outputs/equivalencias_sinisa_snis_2024.csv")
123
124
125 # =====
126 # 2. CARREGAR E CONSOLIDAR DADOS SNIS POR MUNICÍPIO ----
127 # =====
128
129 cat(rep("=", 80), "\n", sep = "")
130 cat("PARTE 2: CARREGANDO E CONSOLIDANDO DADOS SNIS\n")
131 cat(rep("=", 80), "\n\n", sep = "")
132
133 rawdata <- read_csv2(
134   "data/Desagregado-20241117134839.csv",
135   encoding = "UTF-16LE",
136   fileEncoding = "UTF-16LE",
137   stringsAsFactors = FALSE
138 )
139 rawdata <- as_tibble(rawdata) %>% slice(-n())
140 cat("Linhas:", nrow(rawdata), "| Colunas:", ncol(rawdata), "\n\n")
141
142 consolidar_por_municipio <- function(dados_raw) {
143
144   dados_raw <- as_tibble(dados_raw)
145   nomes_cols <- names(dados_raw)
146
147   col_codigo_mun <- nomes_cols[str_detect(nomes_cols, regex("C.digo.*Munic.pio",
148     ↳ ignore_case = TRUE))] [1]
149   col_ano <- nomes_cols[str_detect(nomes_cols, regex("Ano.*Refer.ncia",
150     ↳ ignore_case = TRUE))] [1]
151   col_municipio <- nomes_cols[str_detect(nomes_cols, regex("^Munic.pio$",
152     ↳ ignore_case = TRUE))] [1]
153   col_uf <- nomes_cols[str_detect(nomes_cols, regex("^Estado$",
154     ↳ ignore_case = TRUE))] [1]
155   col_sigla_prestador <- nomes_cols[str_detect(nomes_cols, regex("Sigla.*Prestador",
156     ↳ ignore_case = TRUE))] [1]
157   col_abrangencia <- nomes_cols[str_detect(nomes_cols, regex("^Abrang.ncia$",
158     ↳ ignore_case = TRUE))] [1]
159   col_pop_tot <- nomes_cols[str_detect(nomes_cols, regex("POP_TOT",
160     ↳ ignore_case = TRUE))] [1]
161   col_pop_urb <- nomes_cols[str_detect(nomes_cols, regex("POP_URB",
162     ↳ ignore_case = TRUE))] [1]
163
164   vars_para_manter <- names(dados_raw)[str_detect(names(dados_raw), "^ (AG|ES|FN|QD|BL) ")]
165
166   dados <- dados_raw %>%
167     select(
168       CodigoMunicipio = !!col_codigo_mun,
169       Ano = !!col_ano,
170       Municipio = !!col_municipio,
171       UF = !!col_uf,
172       SigladoPrestador = !!col_sigla_prestador,
173       Abrangencia = !!col_abrangencia,
174       population = !!col_pop_tot,
175       population_urb = !!col_pop_urb,

```

```

168   all_of(vars_para_manter)
169 )
170
171 vars_numericas <- names(dados)[str_detect(names(dados), "(AG|ES|FN|QD|BL)\\d{3}")]
172 dados <- dados %>%
173   mutate(across(all_of(vars_numericas),
174     ~suppressWarnings(as.numeric(as.character(.))))))
175 vars_numericas_validas <- vars_numericas[sapply(dados[vars_numericas], is.numeric)]
176
177 dados_consolidados <- dados %>%
178   group_by(CodigoMunicipio, Ano) %>%
179   summarise(
180     Municipio      = first(Municipio),
181     UF             = first(UF),
182     SiglaPrestador = paste(unique(SigladoPrestador), collapse = ", "),
183     Abrangencia    = first(Abrangencia),
184     Prestadores    = paste(unique(SigladoPrestador), collapse = ", "),
185     NumPrestadores = n_distinct(SigladoPrestador),
186     population     = ifelse(all(is.na(population)), NA_real_, max(population, na.rm
187   ↪ = TRUE)),
187     population_urb = ifelse(all(is.na(population_urb)), NA_real_, max(population_urb, na.rm
188   ↪ = TRUE)),
188     across(all_of(vars_numericas_validas), ~sum(., na.rm = TRUE)),
189     .groups = "drop"
190   )
191
192 # Médias móveis e variáveis auxiliares
193 encontrar_var <- function(codigo, nomes) {
194   r <- nomes[str_detect(nomes, paste0("^", codigo, "\\."))][1]
195   if (is.na(r)) r <- nomes[str_detect(nomes, paste0("^", codigo, "-"))][1]
196   r
197 }
198 vn <- names(dados_consolidados)
199
200 FN026_nome <- encontrar_var("FN026", vn); AG003_nome <- encontrar_var("AG003", vn)
201 AG002_nome <- encontrar_var("AG002", vn); AG004_nome <- encontrar_var("AG004", vn)
202 AG014_nome <- encontrar_var("AG014", vn); AG005_nome <- encontrar_var("AG005", vn)
203 AG021_nome <- encontrar_var("AG021", vn); AG001_nome <- encontrar_var("AG001", vn)
204 AG013_nome <- encontrar_var("AG013", vn); ES003_nome <- encontrar_var("ES003", vn)
205 FN002_nome <- encontrar_var("FN002", vn); FN003_nome <- encontrar_var("FN003", vn)
206 FN007_nome <- encontrar_var("FN007", vn); FN038_nome <- encontrar_var("FN038", vn)
207
208 dados_consolidados <- dados_consolidados %>%
209   group_by(CodigoMunicipio) %>%
210   arrange(Ano) %>%
211   mutate(
212     FN026_media = if (!is.na(FN026_nome)) (.data[[FN026_nome]] + lag(.data[[FN026_nome]]))
213   ↪ / 2 else NA_real_,
214     AG003_media = if (!is.na(AG003_nome)) (.data[[AG003_nome]] + lag(.data[[AG003_nome]]))
215   ↪ / 2 else NA_real_,
216     AG002_media = if (!is.na(AG002_nome)) (.data[[AG002_nome]] + lag(.data[[AG002_nome]]))
217   ↪ / 2 else NA_real_,
218     AG004_media = if (!is.na(AG004_nome)) (.data[[AG004_nome]] + lag(.data[[AG004_nome]]))
219   ↪ / 2 else NA_real_,
220     AG014_media = if (!is.na(AG014_nome)) (.data[[AG014_nome]] + lag(.data[[AG014_nome]]))
221   ↪ / 2 else NA_real_,
222     AG005_media = if (!is.na(AG005_nome)) (.data[[AG005_nome]] + lag(.data[[AG005_nome]]))
223   ↪ / 2 else NA_real_,
224     AG021_media = if (!is.na(AG021_nome)) (.data[[AG021_nome]] + lag(.data[[AG021_nome]]))
225   ↪ / 2 else NA_real_,
226     AG001_media = if (!is.na(AG001_nome)) (.data[[AG001_nome]] + lag(.data[[AG001_nome]]))
227   ↪ / 2 else NA_real_,
228     AG013_media = if (!is.na(AG013_nome)) (.data[[AG013_nome]] + lag(.data[[AG013_nome]]))
229   ↪ / 2 else NA_real_,
230     ES003_media = if (!is.na(ES003_nome)) (.data[[ES003_nome]] + lag(.data[[ES003_nome]]))
231   ↪ / 2 else NA_real_,

```

```

222     FN001 = if (!is.na(FN002_nome) & !is.na(FN003_nome) & !is.na(FN007_nome) &
↪ !is.na(FN038_nome))
223       .data[[FN002_nome]] + .data[[FN003_nome]] + .data[[FN007_nome]] + .data[[FN038_nome]]
224     else NA_real_,
225     GE06a = population_urb,
226     GE12a = population
227   ) %>%
228   ungroup()
229
230   cat("  Municípios:", n_distinct(dados_consolidados$CodigoMunicipio),
231     "| Obs.:", nrow(dados_consolidados), "\n\n")
232   return(dados_consolidados)
233 }
234
235 dados_municipio <- consolidar_por_municipio(rawdata)
236
237
238 # =====
239 # 3. CALCULAR INDICADORES SNIS ----
240 # =====
241
242 cat(rep("=", 80), "\n", sep = "")
243 cat("PARTE 3: CALCULANDO INDICADORES SNIS\n")
244 cat(rep("=", 80), "\n\n", sep = "")
245
246 calcular_indicador <- function(codigo, dados) {
247
248   get_var <- function(cod, df) {
249     nm <- names(df)[str_detect(names(df), paste0("^", cod, "\\."))][1]
250     if (is.na(nm)) nm <- names(df)[str_detect(names(df), paste0("^", cod, "-"))][1]
251     if (is.na(nm)) return(0)
252     df[[nm]]
253   }
254
255   AG001 <- get_var("AG001", dados); AG002 <- get_var("AG002", dados)
256   AG003 <- get_var("AG003", dados); AG004 <- get_var("AG004", dados)
257   AG005 <- get_var("AG005", dados); AG006 <- get_var("AG006", dados)
258   AG008 <- get_var("AG008", dados); AG010 <- get_var("AG010", dados)
259   AG011 <- get_var("AG011", dados); AG012 <- get_var("AG012", dados)
260   AG013 <- get_var("AG013", dados); AG014 <- get_var("AG014", dados)
261   AG017 <- get_var("AG017", dados); AG018 <- get_var("AG018", dados)
262   AG019 <- get_var("AG019", dados); AG021 <- get_var("AG021", dados)
263   AG024 <- get_var("AG024", dados); AG026 <- get_var("AG026", dados)
264   AG028 <- get_var("AG028", dados)
265   ES003 <- get_var("ES003", dados); ES007 <- get_var("ES007", dados)
266   ES028 <- get_var("ES028", dados)
267   FN001 <- dados$FN001
268   FN002 <- get_var("FN002", dados); FN003 <- get_var("FN003", dados)
269   FN005 <- get_var("FN005", dados); FN006 <- get_var("FN006", dados)
270   FN007 <- get_var("FN007", dados); FN008 <- get_var("FN008", dados)
271   FN010 <- get_var("FN010", dados); FN013 <- get_var("FN013", dados)
272   FN014 <- get_var("FN014", dados); FN015 <- get_var("FN015", dados)
273   FN016 <- get_var("FN016", dados); FN017 <- get_var("FN017", dados)
274   FN034 <- get_var("FN034", dados); FN038 <- get_var("FN038", dados)
275   BL001 <- get_var("BL001", dados)
276   QD002 <- get_var("QD002", dados); QD003 <- get_var("QD003", dados)
277   QD004 <- get_var("QD004", dados); QD015 <- get_var("QD015", dados)
278   QD021 <- get_var("QD021", dados); QD022 <- get_var("QD022", dados)
279   QD024 <- get_var("QD024", dados); QD025 <- get_var("QD025", dados)
280
281   FN026_media <- dados$FN026_media; AG001_media <- dados$AG001_media
282   AG002_media <- dados$AG002_media; AG003_media <- dados$AG003_media
283   AG004_media <- dados$AG004_media; AG005_media <- dados$AG005_media
284   AG013_media <- dados$AG013_media; AG014_media <- dados$AG014_media
285   AG021_media <- dados$AG021_media; ES003_media <- dados$ES003_media
286   population <- dados$population
287   population_urb <- dados$population_urb

```

```

288
289 resultado <- switch(codigo,
290                     "IN001" = AG003_media / AG002_media,
291                     "IN003" = (FN017 / (AG011 + ES007)) * (1/1000),
292                     "IN005" = FN002 / (AG011 - AG017 - AG019) * (1/1000),
293                     "IN008" = FN010 / FN026_media,
294                     "IN009" = (AG004_media / AG002_media) * 100,
295                     "IN010" = (AG008 / (AG006 + AG018 - AG019 - AG024)) * 100,
296                     "IN011" = ((AG012 - AG019) / (AG006 + AG018 - AG019)) * 100,
297                     "IN012" = (FN001 / FN017) * 100,
298                     "IN013" = ((AG006 + AG018 - AG011 - AG024) / (AG006 + AG018 - AG024)) *
↪ 100,
299                     "IN014" = (AG008 / AG014_media) * (1000/12),
300                     "IN017" = ((AG011 - AG019) / AG003_media) * (1000/12),
301                     "IN020" = (AG005_media / AG021_media) * 1000,
302                     "IN022" = ((AG010 - AG019) / AG001_media) * (1000000/365),
303                     "IN023" = (AG026 / population_urb) * 100,
304                     "IN025" = ((AG006 + AG018 - AG019) / AG003_media) * (1000/12),
305                     "IN026" = (FN015 / (AG011 + ES007)) * (1/1000),
306                     "IN027" = FN015 / (AG003_media + ES003_media),
307                     "IN028" = (AG011 / (AG006 + AG018 - AG024)) * 100,
308                     "IN029" = ((FN005 - FN006) / FN005) * 100,
309                     "IN030" = (FN015 / FN001) * 100,
310                     "IN031" = (FN010 / FN001) * 100,
311                     "IN032" = ((FN010 + FN014) / FN001) * 100,
312                     "IN033" = ((FN016 + FN034) / FN001) * 100,
313                     "IN035" = (FN010 / FN015) * 100,
314                     "IN036" = ((FN010 + FN014) / FN015) * 100,
315                     "IN037" = (FN013 / FN015) * 100,
316                     "IN042" = ((FN005 - FN001) / FN005) * 100,
317                     "IN043" = (AG013_media / AG003_media) * 100,
318                     "IN044" = (AG008 / (AG010 - AG019)) * 100,
319                     "IN049" = ((AG006 + AG018 - AG010 - AG024) / (AG006 + AG018 - AG024)) *
↪ 100,
320                     "IN050" = ((AG006 + AG018 - AG010 - AG024) / AG005_media) * (1000/365),
321                     "IN051" = ((AG006 + AG018 - AG010 - AG024) / AG002_media) *
↪ (1000000/365),
322                     "IN052" = (AG010 / (AG006 + AG018 - AG024)) * 100,
323                     "IN053" = ((AG010 - AG019) / AG003_media) * (1000/12),
324                     "IN054" = (FN008 / FN005) * 360,
325                     "IN055" = (AG001 / population) * 100,
326                     "IN058" = AG028 / (AG006 + AG018),
327                     "IN060" = FN013 / (AG028 + ES028) * (1/1000),
328                     "IN071" = QD004 / QD002,
329                     "IN072" = QD003 / QD002,
330                     "IN073" = QD015 / QD021,
331                     "IN074" = QD022 / QD021,
332                     "IN083" = QD025 / QD024,
333                     NA_real_
334 )
335
336 resultado[is.infinite(resultado) | is.nan(resultado)] <- NA
337 return(resultado)
338 }
339
340 calcular_todos_indicadores <- function(dados_consolidados) {
341   indicadores <- c(
342     "IN003", "IN005", "IN008", "IN012", "IN026", "IN027", "IN029", "IN030",
343     "IN031", "IN032", "IN033", "IN035", "IN036", "IN037", "IN042", "IN054", "IN060",
344     "IN023", "IN055",
345     "IN001", "IN009", "IN010", "IN011", "IN013", "IN014", "IN017", "IN020",
346     "IN022", "IN025", "IN028", "IN043", "IN044", "IN049", "IN050", "IN051",
347     "IN052", "IN053", "IN058",
348     "IN071", "IN072", "IN073", "IN074", "IN083"
349   )
350   cat("Calculando", length(indicadores), "indicadores...\n")
351   for (ind in indicadores) {

```

```

352   tryCatch({
353     dados_consolidados[[ind]] <- calcular_indicador(ind, dados_consolidados)
354     cat(sprintf(" OK %s: %d valores\n", ind, sum(!is.na(dados_consolidados[[ind]]))))
355   }, error = function(e) {
356     cat(sprintf(" ERRO %s: %s\n", ind, e$message))
357     dados_consolidados[[ind]] <- NA_real_
358   })
359 }
360 cat("\n")
361 return(dados_consolidados)
362 }
363
364 dados_com_indicadores <- calcular_todos_indicadores(dados_municipio)
365
366 # =====
367 # 4. MUNICÍPIOS COM MLTIPILOS PRESTADORES ----
368 # =====
369
370 cat(rep("=", 80), "\n", sep = "")
371 cat("PARTE 4: MUNICÍPIOS COM MLTIPILOS PRESTADORES\n")
372 cat(rep("=", 80), "\n\n", sep = "")
373
374 # NumPrestadores já foi calculado em consolidar_por_municipio (Parte 2).
375 # Aqui apenas extraímos os casos, salvamos o CSV e criamos a dummy.
376 municipios_multiplos <- dados_com_indicadores %>%
377   filter(NumPrestadores > 1) %>%
378   select(CodigoMunicipio, Municipio, Ano, NumPrestadores, Prestadores)
379
380 cod_multiplos <- unique(municipios_multiplos$CodigoMunicipio)
381 cat(" Municípios-ano com múltiplos prestadores:", nrow(municipios_multiplos), "\n")
382 cat(" Municípios únicos afetados:", length(cod_multiplos), "\n\n")
383
384 write_csv(municipios_multiplos, "outputs/municipios_multiplos_prestadores.csv")
385
386 dados_com_indicadores <- dados_com_indicadores %>%
387   mutate(multiplos_prestadores = if_else(CodigoMunicipio %in% cod_multiplos, 1L, 0L))
388
389 # =====
390 # 5. LIMPEZA DE OUTLIERS ----
391 # =====
392
393 cat(rep("=", 80), "\n", sep = "")
394 cat("PARTE 5: LIMPEZA DE OUTLIERS\n")
395 cat(rep("=", 80), "\n\n", sep = "")
396
397 aplicar_limites_outliers <- function(dados) {
398   dados %>% mutate(
399     across(
400       any_of(c("IN009", "IN010", "IN011", "IN013", "IN023", "IN028", "IN029",
401               "IN030", "IN031", "IN032", "IN033", "IN035", "IN036", "IN037",
402               "IN042", "IN043", "IN044", "IN049", "IN052", "IN055")),
403       ~if_else(. < 0 | . > 100, NA_real_, .)
404     ),
405     across(
406       any_of(c("IN001", "IN003", "IN005", "IN008", "IN012", "IN014", "IN017",
407               "IN020", "IN022", "IN025", "IN026", "IN027", "IN050", "IN051",
408               "IN053", "IN058", "IN060")),
409       ~if_else(. < 0, NA_real_, .)
410     ),
411     across(any_of("IN054"), ~if_else(. < 0 | . > 720, NA_real_, .)),
412     across(any_of(c("IN072", "IN074", "IN083")), ~if_else(. < 0 | . > 8760, NA_real_, .)),
413     across(any_of(c("IN071", "IN073")), ~if_else(. < 0, NA_real_, .))
414   )
415 }
416
417 }
418

```

```

419 n_antes <- sum(!is.na(select(dados_com_indicadores, starts_with("IN"))))
420 dados_com_indicadores <- aplicar_limites_outliers(dados_com_indicadores)
421 n_depois <- sum(!is.na(select(dados_com_indicadores, starts_with("IN"))))
422 cat(" Removidos:", n_antes - n_depois,
423     "(", round((n_antes - n_depois) / n_antes * 100, 2), "%)\n\n")
424
425
426 # =====
427 # 6. VALIDAO COM SNIS ORIGINAL (opcional) ----
428 # =====
429
430 cat(rep("=", 80), "\n", sep = "")
431 cat("PARTE 6: VALIDAO COM SNIS ORIGINAL\n")
432 cat(rep("=", 80), "\n\n", sep = "")
433
434 arquivo_snis_ind <- "data/Desagregado-20241019160853.csv"
435 comparacao <- NULL
436
437 indicadores_todos <- c(
438   "IN003", "IN005", "IN008", "IN012", "IN023", "IN026", "IN027", "IN029",
439   "IN030", "IN031", "IN032", "IN033", "IN035", "IN036", "IN037", "IN042",
440   "IN054", "IN055", "IN060",
441   "IN001", "IN009", "IN010", "IN011", "IN013", "IN014", "IN017", "IN020",
442   "IN022", "IN025", "IN028", "IN043", "IN044", "IN049", "IN050", "IN051",
443   "IN052", "IN053", "IN058",
444   "IN071", "IN072", "IN073", "IN074", "IN083"
445 )
446
447 if (file.exists(arquivo_snis_ind)) {
448   snis_original <- read.csv2(arquivo_snis_ind,
449                               encoding = "UTF-16LE",
450                               fileEncoding = "UTF-16LE",
451                               stringsAsFactors = FALSE) %>%
452     as_tibble() %>% slice(-n())
453
454   inds_snis <- sum(str_detect(names(snis_original), "^IN\\d{3}"))
455
456   if (inds_snis > 0) {
457     nomes_snis <- names(snis_original)
458     col_cod_snis <- nomes_snis[str_detect(nomes_snis, regex("C.digo.*Munic.pio", ignore_case
459     ↪ = TRUE))][1]
459     col_ano_snis <- nomes_snis[str_detect(nomes_snis, regex("Ano.*Refer.ncia", ignore_case
460     ↪ = TRUE))][1]
461
462     snis_um <- snis_original %>%
463       group_by(.data[[col_cod_snis]], .data[[col_ano_snis]]) %>%
464       filter(n() == 1) %>% ungroup() %>%
465       mutate(.chave = paste(as.character(.data[[col_cod_snis]]),
466                             as.character(.data[[col_ano_snis]]), sep = "_"))
467
468     calc_um <- dados_com_indicadores %>%
469       filter(NumPrestadores == 1) %>%
470       mutate(.chave = paste(as.character(CodigoMunicipio), as.character(Ano), sep = "_"))
471
472     chave_comum <- intersect(snis_um$.chave, calc_um$.chave)
473     snis_f <- snis_um %>% filter(.chave %in% chave_comum) %>% arrange(.chave)
474     calc_f <- calc_um %>% filter(.chave %in% chave_comum) %>% arrange(.chave)
475
476     resultados_val <- list()
477     for (ind in indicadores_todos) {
478       col_snis_ind <- names(snis_f)[str_detect(names(snis_f), paste0("^", ind, "[.-]"))][1]
479       if (is.na(col_snis_ind) || !(ind %in% names(calc_f))) next
480       vc <- calc_f[[ind]]
481       vs <- suppressWarnings(as.numeric(gsub(",", ".", as.character(snis_f[[col_snis_ind]]))))
482       ok <- !is.na(vc) & !is.na(vs)
483       if (sum(ok) < 10) next
484       vc <- vc[ok]; vs <- vs[ok]

```

```

484     corr      <- cor(vc, vs)
485     rmse      <- sqrt(mean((vc - vs)^2))
486     rmse_rel  <- if (abs(mean(vs)) > 1e-6) rmse / abs(mean(vs)) * 100 else NA_real_
487     bias_rel  <- mean((vc - vs) / (abs(vs) + 1e-6)) * 100
488     pval      <- tryCatch(t.test(vc, vs, paired = TRUE)$p.value, error = function(e)
↪ NA_real_)
489     valido    <- !is.na(corr) && corr >= 0.95 && !is.na(rmse_rel) && rmse_rel <= 20
490     resultados_val[[length(resultados_val) + 1]] <- data.frame(
491       Indicador = ind, N_pares = sum(ok),
492       Correlacao = round(corr, 4), RMSE_rel = round(rmse_rel, 2),
493       Bias_rel = round(bias_rel, 2), P_valor = round(pval, 4), Valido = valido
494     )
495   }
496   comparacao <- bind_rows(resultados_val) %>% arrange(desc(Correlacao))
497   write_csv(comparacao, "outputs/comparacao_calculado_vs_snis.csv")
498   cat("  Válidos:", sum(comparacao$Valido), "de", nrow(comparacao), "\n\n")
499 }
500 }
501
502 indicadores_validos <- if (!is.null(comparacao)) {
503   comparacao$Indicador[comparacao$Valido]
504 } else {
505   indicadores_todos
506 }
507
508
509 # =====
510 # 7. MONTAR PAINEL 2017-2024 COM NOMES SINISA ----
511 # =====
512
513 cat(rep("=", 80), "\n", sep = "")
514 cat("PARTE 7: MONTANDO PAINEL 2017-2024 (NOMES SINISA)\n")
515 cat(rep("=", 80), "\n\n", sep = "")
516
517 equivalencias <- read_csv("outputs/equivalencias_sinisa_snis_2024.csv", show_col_types =
↪ FALSE)
518 equivalencias_validas <- equivalencias %>% filter(codigo_snis %in% indicadores_validos)
519
520 # 7A. Preparar SNIS 2017-2022
521
522 dados_snis_base <- dados_com_indicadores %>% filter(Ano <= 2022)
523
524 # Corrigir códigos IBGE 6 -> 7 dígitos
525 codigos_6 <- sum(nchar(dados_snis_base$CodigoMunicipio) == 6, na.rm = TRUE)
526 if (codigos_6 > 0 && file.exists("data/br_bd_diretorios_brasil_municipio.csv")) {
527   dict_mun <- read_csv("data/br_bd_diretorios_brasil_municipio.csv", show_col_types = FALSE)
528   mapa_cod <- dict_mun %>%
529     select(Codigo6 = id_municipio_6, Codigo7 = id_municipio) %>%
530     mutate(across(everything(), as.character)) %>% filter(!is.na(Codigo6), !is.na(Codigo7))
531   dados_snis_base <- dados_snis_base %>%
532     left_join(mapa_cod, by = c("CodigoMunicipio" = "Codigo6")) %>%
533     mutate(CodigoMunicipio = coalesce(Codigo7, CodigoMunicipio)) %>% select(-Codigo7)
534 }
535
536 # Renomear IN -> SINISA
537 inds_presentes <- equivalencias_validas$codigo_snis[
538   equivalencias_validas$codigo_snis %in% names(dados_snis_base)
539 ]
540 rename_vec <- setNames(
541   inds_presentes,
542   equivalencias_validas$codigo_sinisa[equivalencias_validas$codigo_snis %in%
↪ names(dados_snis_base)]
543 )
544
545 dados_snis_renomeado <- dados_snis_base %>% rename(any_of(rename_vec))
546 inds_sinisa_renomeados <- intersect(names(rename_vec), names(dados_snis_renomeado))
547

```

```

548 colunas_base_snis <- intersect(
549   c("CodigoMunicipio", "Municipio", "UF", "Ano", "SiglaPrestador", "Abrangencia",
550     "Prestadores", "NumPrestadores", "population", "population_urb", "multiplos_prestadores"),
551   names(dados_snis_renomeado)
552 )
553 snis_para_painel <- dados_snis_renomeado %>%
554   select(all_of(colunas_base_snis), all_of(inds_sinisa_renomeados))
555
556 # 7B. Carregar SINISA 2023-2024
557
558 carregar_sinisa_ano <- function(arquivo, ano) {
559   cat("Carregando SINISA", ano, "...\\n")
560   planilhas <- excel_sheets(arquivo)
561   planilhas_dados <- planilhas[!str_detect(planilhas, regex("nota|metodol", ignore_case =
562     ↪ TRUE))]
563
564   detectar_skip <- function(arq, planilha) {
565     df_bruto <- read_excel(arq, sheet = planilha, col_names = FALSE, n_max = 15)
566     for (i in seq_len(nrow(df_bruto))) {
567       linha <- as.character(unlist(df_bruto[i, ]))
568       if (sum(str_detect(linha, "^I[A-Z]{2}\\d{4}$"), na.rm = TRUE) >= 3) return(i - 1)
569     }
570   }
571
572   dados_lista <- list()
573   for (planilha in planilhas_dados) {
574     skip <- tryCatch(detectar_skip(arquivo, planilha), error = function(e) 9L)
575     df <- tryCatch(read_excel(arquivo, sheet = planilha, skip = skip), error = function(e)
576       ↪ NULL)
577     if (!is.null(df) && nrow(df) > 0) {
578       cols_auto <- names(df)[str_detect(names(df), "\\\\.\\.\\.\\.\\d+$")]
579       if (length(cols_auto) > 0) df <- df %>% select(-all_of(cols_auto))
580       coll <- names(df)[1]
581       df <- df %>% filter(suppressWarnings(!is.na(as.numeric(as.character(.data[[coll]]))))))
582       cols_ind <- names(df)[str_detect(names(df), "^I[A-Z]{2}\\d{4}$")]
583       df <- df %>% mutate(across(all_of(cols_ind),
584         ↪ ~suppressWarnings(as.numeric(as.character(.))))))
585       dados_lista[[planilha]] <- df
586     }
587   }
588   if (length(dados_lista) == 0) return(NULL)
589
590   resultado <- dados_lista[[1]]
591   col_ibge <- names(resultado)[str_detect(names(resultado), regex("cod_ibge|ibge",
592     ↪ ignore_case = TRUE))]
593   if (length(dados_lista) > 1) {
594     for (i in 2:length(dados_lista)) {
595       df_i <- dados_lista[[i]]
596       col_ibge_i <- names(df_i)[str_detect(names(df_i), regex("cod_ibge|ibge", ignore_case =
597         ↪ TRUE))]
598       cols_novas <- setdiff(names(df_i), names(resultado))
599       if (length(cols_novas) > 0)
600         resultado <- resultado %>%
601           left_join(df_i %>% select(all_of(c(col_ibge_i, cols_novas))),
602             by = setNames(col_ibge_i, col_ibge))
603     }
604   }
605   resultado$Ano <- ano
606   cat(" Total:", nrow(resultado), "linhas,", ncol(resultado), "colunas\\n")
607   return(resultado)
608 }
609
610 sinisa_2023 <- carregar_sinisa_ano("data/SINISA_AGUA_Indicadores_Base
611   ↪ Municipal_2023_V2.xlsx", 2023)
612 sinisa_2024 <- carregar_sinisa_ano("data/SINISA_AGUA_Indicadores_Base Municipal_2024.xlsx",
613   ↪ 2024)

```

```

608 | sinisa_completo <- bind_rows(sinisa_2023, sinisa_2024)
609
610 # 7C. Preparar SINISA para o painel
611 # Colunas de cadastro do SINISA:
612 #   CAD0006 = Sigla do Prestador
613 #   CAD0005 = Nome do Prestador
614 #   CAD0003 = Abrangência
615
616 col_ibge_sinisa <- names(sinisa_completo)[str_detect(names(sinisa_completo),
617   ↪ regex("cod_ibge|ibge", ignore_case = TRUE))][1]
618 col_municipio_s <- names(sinisa_completo)[str_detect(names(sinisa_completo),
619   ↪ regex("^Município$|^Municipio$", ignore_case = TRUE))][1]
620 col_uf_s <- names(sinisa_completo)[str_detect(names(sinisa_completo), "UF$")][1]
621 col_sigla_s <- "CAD0006" # Sigla do Prestador
622 col_prestador_s <- "CAD0005" # Nome do Prestador
623 col_abrang_s <- "CAD0003" # Abrangência
624
625 cat(" SINISA  colunas de cadastro detectadas:\n")
626 cat(" IBGE:      ", col_ibge_sinisa, "\n")
627 cat(" Município: ", col_municipio_s, "\n")
628 cat(" UF:        ", col_uf_s, "\n")
629 cat(" Sigla prest.: ", col_sigla_s, " presente:", col_sigla_s %in%
630   ↪ names(sinisa_completo), "\n")
631 cat(" Prestador:   ", col_prestador_s, " presente:", col_prestador_s %in%
632   ↪ names(sinisa_completo), "\n")
633 cat(" Abrangência: ", col_abrang_s, " presente:", col_abrang_s %in%
634   ↪ names(sinisa_completo), "\n\n")
635
636 inds_sinisa_validos <- equivalencias_validas$codigo_sinisa
637 inds_sinisa_em_sinisa <- inds_sinisa_validos[inds_sinisa_validos %in% names(sinisa_completo)]
638
639 colunas_sinisa_base <- c(
640   CodigoMunicipio = col_ibge_sinisa,
641   Municipio       = col_municipio_s,
642   UF              = col_uf_s,
643   SiglaPrestador  = col_sigla_s,
644   Prestadores     = col_prestador_s,
645   Abrangencia     = col_abrang_s
646 )
647 colunas_sinisa_base <- colunas_sinisa_base[!is.na(colunas_sinisa_base) &
648   ↪ colunas_sinisa_base %in% names(sinisa_completo)]
649
650 sinisa_para_painel <- sinisa_completo %>%
651   rename(any_of(colunas_sinisa_base)) %>%
652   mutate(CodigoMunicipio = as.character(CodigoMunicipio), Ano = as.integer(Ano)) %>%
653   select(CodigoMunicipio, Municipio, UF, Ano,
654     ↪ any_of(c("SiglaPrestador", "Prestadores", "Abrangencia")),
655     ↪ all_of(inds_sinisa_em_sinisa))
656
657 # 7D. Unificar e limpar outliers no painel completo
658
659 snis_para_painel <- snis_para_painel %>%
660   mutate(CodigoMunicipio = as.character(CodigoMunicipio), Ano = as.integer(Ano))
661
662 painel_completo <- bind_rows(snis_para_painel, sinisa_para_painel)
663
664 # Limpeza de outliers no painel (renomear temporariamente para IN)
665 inds_painel <- names(painel_completo)[str_detect(names(painel_completo), "^I[A-Z]{2}\\d")]
666 mapa_sinisa_para_in <- setNames(
667   equivalencias_validas$codigo_sinisa,
668   equivalencias_validas$codigo_snis
669 )
670 mapa_sinisa_para_in <- mapa_sinisa_para_in[mapa_sinisa_para_in %in% names(painel_completo)]
671
672 painel_completo <- painel_completo %>%
673   rename(any_of(mapa_sinisa_para_in)) %>%
674   aplicar_limite_outliers() %>%

```

```

670   rename(any_of(setNames(names(mapa_sinisa_para_in), unname(mapa_sinisa_para_in))))
671
672 write_csv(painel_completo, "data/painel_completo_2017_2024.csv")
673 cat("Painel salvo: data/painel_completo_2017_2024.csv\n")
674 cat("  Linhas:", nrow(painel_completo), "| Municípios:",
      ↪ n_distinct(painel_completo$CodigoMunicipio), "\n\n")
675
676 # 7E. DEFLAÇÃO IPCA  indicadores monetários do painel
677 # Ano-base: dezembro/2024 (reais de hoje)
678 # Apenas indicadores com dimensão R$/unidade efetivamente estimados
679
680 cat("Deflacionando indicadores monetários pelo IPCA (base dez/2024)...\n")
681
682 library(deflateBR)
683
684 anos_unicos <- sort(unique(painel_completo$Ano))
685
686 ipca_fatores <- tibble(
687   Ano = anos_unicos,
688   data = as.Date(paste0(anos_unicos, "-12-01"))
689 ) %>%
690   rowwise() %>%
691   mutate(
692     fator = deflateBR::deflate(
693       nominal_values = 1,
694       nominal_dates = data,
695       real_date      = "12/2024",
696       index          = "ipca"
697     )
698   ) %>%
699   ungroup() %>%
700   select(Ano, fator)
701
702 cat("  Fatores IPCA (base dez/2024):\n")
703 print(ipca_fatores)
704 cat("\n")
705
706 # Indicadores com dimensão monetária R$/unidade
707 # Fonte: descrições da tabela de resultados
708 inds_monetarios <- c(
709   "IFA1001", # R$/m      receita operacional direta média
710   "IFA2002", # R$/m      despesa total média incluindo tributos
711   "IFA2004", # R$/m      despesa de exploração média
712   "IFA2005", # R$/eco.ano despesa de exploração por economia ativa
713   "IFA2006", # R$/emp.ano despesa com pessoal próprio
714   "IFA2007"  # R$/KWh     despesa de energia elétrica
715 )
716
717 # Indicadores adimensionais NO deflacionar:
718 # IFA1003, IFA1004 (%), IFA1005 (dias),
719 # IFA2001, IFA2008, IFA200G, IFA2010IFA2014 (%)
720
721 inds_deflacionar <- intersect(inds_monetarios, names(painel_completo))
722 cat("  Indicadores a deflacionar:", paste(inds_deflacionar, collapse = ", "), "\n")
723
724 if (length(inds_deflacionar) > 0) {
725   painel_completo <- painel_completo %>%
726     left_join(ipca_fatores, by = "Ano") %>%
727     mutate(across(all_of(inds_deflacionar), ~.x * fator)) %>%
728     select(-fator)
729   cat("  Deflação aplicada: OK\n\n")
730 }
731
732 # =====
733 # 8. DICIONÁRIO E VARIÁVEL DE TRATAMENTO ----
734 # =====
735

```

```

736 cat(rep("=", 80), "\n", sep = "")
737 cat("PARTE 8: DICIONARIO E VARIVEL DE TRATAMENTO\n")
738 cat(rep("=", 80), "\n\n", sep = "")
739
740 dicionario <- read_csv("data/br_bd_diretorios_brasil_municipio.csv", show_col_types = FALSE)
741   ↪ %>%
742   select(id_municipio, id_municipio_6, nome, capital_uf, sigla_uf, id_microrregiao)
743
744 # FONTE NICA: concessoes.ods planilha 2
745 # Colunas: município, estado, concessionária, holding, modalidade,
746 #         anodocontrato, prazo, população beneficiada
747 # "estado" contém o nome completo do estado (ex: "MINAS GERAIS"), não a sigla.
748 # O join com o dicionário IBGE é feito por nome_município + sigla_uf, após
749 # converter o nome do estado para sigla via tabela de conversão.
750
751 concessoes <- read_ods("data/concessoes.ods", sheet = 2)
752
753 col_mun_c <- names(concessoes)[str_detect(names(concessoes), regex("^munic",
754   ↪ ignore_case = TRUE))][1]
755 col_uf_c <- names(concessoes)[str_detect(names(concessoes), regex("^estado$|^uf$",
756   ↪ ignore_case = TRUE))][1]
757 col_concess <- names(concessoes)[str_detect(names(concessoes), regex("concession",
758   ↪ ignore_case = TRUE))][1]
759 col_ano_c <- names(concessoes)[str_detect(names(concessoes),
760   ↪ regex("anodocontrato|ano.*contrato", ignore_case = TRUE))][1]
761
762 cat(" Colunas detectadas:", col_mun_c, "|", col_uf_c, "|", col_concess, "|", col_ano_c,
763   ↪ "\n\n")
764
765 # Tabela de conversão nome completo do estado -> sigla UF
766 estados_sigla <- tibble(
767   nome_estado = c(
768     "ACRE", "ALAGOAS", "AMAPA", "AMAZONAS", "BAHIA", "CEARA", "DISTRITO FEDERAL",
769     "ESPIRITO SANTO", "GOIAS", "MARANHAO", "MATO GROSSO", "MATO GROSSO DO SUL",
770     "MINAS GERAIS", "PARA", "PARAIBA", "PARANA", "PERNAMBUCO", "PIAUI",
771     "RIO DE JANEIRO", "RIO GRANDE DO NORTE", "RIO GRANDE DO SUL", "RONDONIA",
772     "RORAIMA", "SANTA CATARINA", "SAO PAULO", "SERGIPE", "TOCANTINS"
773   ),
774   sigla_uf = c(
775     "AC", "AL", "AP", "AM", "BA", "CE", "DF", "ES", "GO", "MA", "MT", "MS",
776     "MG", "PA", "PB", "PR", "PE", "PI", "RJ", "RN", "RS", "RO", "RR", "SC", "SP", "SE", "TO"
777   )
778 )
779
780 # Normalizar nome do estado na planilha (remover acentos, maiúsculas)
781 # stringi::stri_trans_general com "Latin-ASCII" é robusto para qualquer acento
782 norm <- function(x) str_to_upper(str_trim(
783   stringi::stri_trans_general(as.character(x), "Latin-ASCII")
784 ))
785
786 concessoes_norm <- concessoes %>%
787   mutate(
788     nome_estado_norm = norm(.data[[col_uf_c]]),
789     nome_mun_norm = norm(.data[[col_mun_c]])
790   ) %>%
791   left_join(estados_sigla, by = c("nome_estado_norm" = "nome_estado"))
792
793 # Verificar estados não mapeados
794 estados_nao_mapeados <- concessoes_norm %>%
795   filter(is.na(sigla_uf)) %>%
796   distinct(nome_estado_norm = .data[[col_uf_c]])
797 if (nrow(estados_nao_mapeados) > 0) {
798   cat(" AVISO estados não mapeados para sigla:\n")
799   print(estados_nao_mapeados)
800   cat("\n")
801 }
802

```

```

797 # Dicionário normalizado para join por nome + sigla
798 dicionario_join <- dicionario %>%
799   mutate(nome_join = norm(nome))
800
801 # Municípios individuais (não-CORSAN)
802 # Filtros do desenho do estudo:
803 # 1. modalidade == "concessão plena" (excluir PPP, arrendamento, etc.)
804 # 2. anodocontrato >= 2020 (após o NMLSB)
805 # 3. excluir CORSAN (tratada separadamente)
806
807 col_modal_c <- names(concessoes_norm)[str_detect(
808   names(concessoes_norm), regex("modalidade", ignore_case = TRUE)
809 )][1]
810 cat(" Coluna modalidade:", col_modal_c, "\n")
811
812 # Verificar modalidades disponíveis
813 cat(" Modalidades disponíveis:\n")
814 concessoes_norm %>% count(.data[[col_modal_c]], sort = TRUE) %>% print()
815 cat("\n")
816
817 concessoes_ind <- concessoes_norm %>%
818   filter(
819     !str_detect(.data[[col_concess]], regex("corsan", ignore_case = TRUE)),
820     !is.na(sigla_uf),
821     # Apenas concessão plena
822     str_detect(.data[[col_modal_c]], regex("plena", ignore_case = TRUE)),
823     # Apenas a partir de 2020
824     suppressWarnings(as.numeric(as.character(.data[[col_ano_c]]))) >= 2020
825   ) %>%
826   transmute(
827     nome_join      = nome_mun_norm,
828     sigla_uf       = sigla_uf,
829     anodocontrato  = suppressWarnings(as.numeric(as.character(.data[[col_ano_c]]))),
830     concessionaria = .data[[col_concess]],
831     modalidade     = .data[[col_modal_c]],
832     municipio_corsan = FALSE
833   ) %>%
834   # Correções manuais de grafia divergente entre planilha e dicionário IBGE
835   mutate(nome_join = case_when(
836     # Bloco regional PR: apenas Campo Largo entra (os demais 15 estão em outras linhas)
837     nome_join == "CAMPO LARGO + 15 MUNICIPIOS" & sigla_uf == "PR" ~ "CAMPO
838     ↪ LARGO",
839     # Apóstrofes: IBGE mantém D' na grafia oficial
840     nome_join == "TANQUE D ARCA" & sigla_uf == "AL" ~ "TANQUE
841     ↪ D'ARCA",
842     nome_join == "OLHO D AGUA DO CASADO" & sigla_uf == "AL" ~ "OLHO D'AGUA
843     ↪ DO CASADO",
844     nome_join == "OLHO D AGUA GRANDE" & sigla_uf == "AL" ~ "OLHO D'AGUA
845     ↪ GRANDE",
846     # Rio de Janeiro subdividido em blocos: todos = mesmo município
847     str_detect(nome_join, "^RIO DE JANEIRO") & sigla_uf == "RJ" ~ "RIO DE
848     ↪ JANEIRO",
849     # São Paulo (Zona Oeste): mesmo município
850     str_detect(nome_join, "^SAO PAULO") & sigla_uf == "SP" ~ "SAO PAULO",
851     # Eldorado dos Carajás: IBGE grafa "DO" no singular
852     nome_join == "ELDORADO DOS CARAJAS" & sigla_uf == "PA" ~ "ELDORADO DO
853     ↪ CARAJAS",
854     # Taboão: verificar nome exato no dicionário
855     nome_join == "TABOCAO" & sigla_uf == "TO" ~ "TABOCAO",
856     TRUE ~ nome_join
857   ))
858
859 concessoes_municipais <- concessoes_ind %>%
860   left_join(dicionario_join, by = c("nome_join", "sigla_uf")) %>%
861   # Fallback para casos que não bateram no join por nome:
862   # Taboão (TO) código IBGE 1720978 (6 dígitos)
863   mutate(id_municipio_6 = case_when(

```

```

858     nome_join == "TABOCAO" & sigla_uf == "TO" & is.na(id_municipio_6) ~ 172097L,
859     TRUE ~ as.integer(id_municipio_6)
860   )) %>%
861   filter(!is.na(id_municipio_6)) %>%
862   rename(CódigodoMunicípio = id_municipio_6)
863
864   n_nao_encontrados <- nrow(concessoes_ind) - nrow(concessoes_municipais)
865   cat("  Municípios individuais pareados:", nrow(concessoes_municipais), "\n")
866   if (n_nao_encontrados > 0) {
867     cat("  AVISO:", n_nao_encontrados, "município(s) não pareado(s) grafia divergente:\n")
868     anti_join(concessoes_ind, dicionario_join, by = c("nome_join", "sigla_uf")) %>%
869       select(nome_join, sigla_uf, concessionaria) %>%
870       print(n = Inf)
871   }
872   cat("\n")
873
874   # CORSAN: expandir para todos os municípios RS via SNIS
875   # Apenas se for concessão plena e anodocontrato >= 2020
876   linha_corsan <- concessoes %>%
877     filter(
878       str_detect(.data[[col_concess]], regex("corsan", ignore_case = TRUE)),
879       str_detect(.data[[col_modal_c]], regex("plena", ignore_case = TRUE))
880     )
881
882   ano_corsan <- if (nrow(linha_corsan) > 0) {
883     suppressWarnings(as.numeric(as.character(linha_corsan[[col_ano_c]][1])))
884   } else {
885     NA_real_
886   }
887
888   # Só incluir se >= 2020
889   if (!is.na(ano_corsan) && ano_corsan < 2020) {
890     cat("  CORSAN ano do contrato:", ano_corsan, "< 2020, excluída do grupo tratado\n\n")
891     ano_corsan <- NA_real_
892   } else {
893     cat("  CORSAN ano do contrato:", ano_corsan, "\n")
894   }
895
896   if (!is.na(ano_corsan)) {
897
898     nomes_raw_temp <- names(rawdata)
899     col_cod_r2 <- nomes_raw_temp[str_detect(nomes_raw_temp, regex("C.digo.*Munic.pio",
900       ↪ ignore_case = TRUE))][1]
901     col_uf_r2 <- nomes_raw_temp[str_detect(nomes_raw_temp, regex("^Estado$",
902       ↪ ignore_case = TRUE))][1]
903     col_sigla_r2 <- nomes_raw_temp[str_detect(nomes_raw_temp, regex("Sigla.*Prestador",
904       ↪ ignore_case = TRUE))][1]
905
906     municipios_corsan <- rawdata %>%
907       filter(
908         .data[[col_uf_r2]] == "RS",
909         str_detect(.data[[col_sigla_r2]], regex("corsan", ignore_case = TRUE))
910       ) %>%
911       distinct(.data[[col_cod_r2]]) %>%
912       rename(CódigodoMunicípio = 1) %>%
913       mutate(
914         CódigoMunicípio = as.numeric(CódigodoMunicípio),
915         anodocontrato = ano_corsan,
916         concessionaria = "CORSAN",
917         municipio_corsan = TRUE
918       ) %>%
919       left_join(dicionario, by = join_by(CódigodoMunicípio == id_municipio_6))
920
921     cat("  CORSAN:", nrow(municipios_corsan), "municípios RS identificados no SNIS\n\n")
922
923     if (nrow(municipios_corsan) == 0)
924       warning("Nenhum município CORSAN encontrado. Verifique a coluna '", col_sigla_r2, "'.")

```

```

922 } else {
923   municipios_corsan <- tibble()
924   cat("  CORSAN não encontrada  expansão ignorada.\n\n")
925 }
926
927 # Consolidar
928 abcon <- bind_rows(concessoes_municipais, municipios_corsan)
929
930 cat("  Total de entradas em abcon:", nrow(abcon), "\n")
931 cat("  Privatizados >= 2020:", sum(abcon$anodocontrato >= 2020, na.rm = TRUE), "\n")
932 cat("  Privatizados < 2020:", sum(abcon$anodocontrato < 2020, na.rm = TRUE), "\n\n")
933
934 anodocontrato_map <- setNames(abcon$anodocontrato, abcon$CódigoMunicípio)
935
936 # =====
937 # 8B. FIGURA: CONCESSÕES PLENAS POR ANO ----
938 # =====
939
940 cat(rep("=", 80), "\n", sep = "")
941 cat("PARTE 8B: FIGURA  CONCESSÕES PLENAS POR ANO\n")
942 cat(rep("=", 80), "\n\n", sep = "")
943
944 # Contar municípios com concessão plena por ano  TODAS as concessões plenas
945 # disponíveis em concessoes.ods, sem filtro de ano (motiva o artigo)
946 # Inclui CORSAN como 317 municípios no ano do contrato
947
948 # Concessões plenas individuais (não-CORSAN)  todos os anos
949 concessoes_plenas_todas <- concessoes %>%
950   filter(
951     str_detect(.data[[col_modal_c]], regex("plena", ignore_case = TRUE)),
952     !str_detect(.data[[col_concess]], regex("corsan", ignore_case = TRUE))
953   ) %>%
954   transmute(anodocontrato = suppressWarnings(as.numeric(as.character(.data[[col_ano_c]]))))
955   ↪ %>%
956   filter(!is.na(anodocontrato))
957
958 # Adicionar CORSAN como 317 municípios no ano do contrato (se for plena)
959 n_corsan_fig <- if (!is.na(ano_corsan)) nrow(municipios_corsan) else 0L
960
961 concessoes_por_ano <- concessoes_plenas_todas %>%
962   count(anodocontrato, name = "n_municipios") %>%
963   bind_rows(
964     if (n_corsan_fig > 0)
965       tibble(anodocontrato = ano_corsan, n_municipios = n_corsan_fig)
966     else
967       tibble()
968   ) %>%
969   group_by(anodocontrato) %>%
970   summarise(n_municipios = sum(n_municipios), .groups = "drop") %>%
971   complete(anodocontrato = seq(min(anodocontrato), max(anodocontrato)),
972     fill = list(n_municipios = 0L)) %>%
973   arrange(anodocontrato)
974
975 media_anual <- mean(concessoes_por_ano$n_municipios)
976
977 cat("  Total de municípios com concessão:", sum(concessoes_por_ano$n_municipios), "\n")
978 cat("  Média anual:", round(media_anual, 1), "concessões/ano\n")
979 cat("  Ano com mais concessões:",
980   ↪ concessoes_por_ano$anodocontrato[which.max(concessoes_por_ano$n_municipios)],
981   "(", max(concessoes_por_ano$n_municipios), "municípios)\n\n")
982
983 fig_concessoes <- ggplot(concessoes_por_ano,
984   aes(x = anodocontrato, y = n_municipios)) +
985
986   # rea preenchida sob a linha

```

```

987 geom_area(fill = "#1E6BA8", alpha = 0.25) +
988
989 # Linha principal
990 geom_line(color = "#1E6BA8", linewidth = 0.9) +
991
992 # Pontos
993 geom_point(color = "#1E6BA8", size = 2.2) +
994
995 # Linha de média tracejada
996 geom_hline(yintercept = media_anual,
997           linetype = "dashed", color = "gray50", linewidth = 0.7) +
998
999 # Label da média deslocado para a esquerda, acima da linha tracejada
1000 annotate("text",
1001         x = min(concessoes_por_ano$anodocontrato) + 8,
1002         y = media_anual + 8,
1003         label = paste0("Média: ", round(media_anual, 0), " concessões/ano"),
1004         hjust = 0, size = 3.2, color = "gray40") +
1005
1006 # Linha vertical em 2020 NMLSB
1007 geom_vline(xintercept = 2020,
1008           color = "gray30", linewidth = 0.7, linetype = "solid") +
1009
1010 annotate("text",
1011         x = 2020.2, y = max(concessoes_por_ano$n_municipios) * 0.95,
1012         label = "NMLSB", hjust = 0, size = 3.4, color = "gray30") +
1013
1014 # Escalas
1015 scale_x_continuous(
1016   breaks = seq(min(concessoes_por_ano$anodocontrato),
1017               max(concessoes_por_ano$anodocontrato), by = 2)
1018 ) +
1019 scale_y_continuous(
1020   breaks = seq(0, max(concessoes_por_ano$n_municipios) + 20, by = 50),
1021   expand = expansion(mult = c(0, 0.08))
1022 ) +
1023
1024 labs(x = NULL, y = NULL) +
1025 theme_minimal(base_size = 11) +
1026 theme(
1027   plot.background = element_rect(fill = "white", color = NA),
1028   panel.background = element_rect(fill = "white", color = NA),
1029   panel.grid = element_blank(),
1030   axis.line.x = element_line(color = "gray20", linewidth = 0.4),
1031   axis.line.y = element_line(color = "gray20", linewidth = 0.4),
1032   axis.text = element_text(color = "gray20", size = 10),
1033   axis.ticks = element_line(color = "gray20", linewidth = 0.3),
1034   plot.margin = margin(20, 25, 15, 15)
1035 )
1036
1037 ggsave("outputs/figura_concessoes_tempo.png",
1038       plot = fig_concessoes,
1039       width = 12, height = 6, dpi = 300, bg = "white")
1040
1041 cat(" Figura salva: outputs/figura_concessoes_tempo.png\n\n")
1042
1043
1044 # =====
1045 # 9. BASE ANO-BASE 2017 ----
1046 # =====
1047
1048 cat(rep("=", 80), "\n", sep = "")
1049 cat("PARTE 9: BASE ANO-BASE 2017\n")
1050 cat(rep("=", 80), "\n\n", sep = "")
1051
1052 nomes_raw <- names(rawdata)

```

```

1053 col_cod_r <- nomes_raw[str_detect(nomes_raw, regex("C.digo.*Munic.pio", ignore_case =
      ↳ TRUE))][1]
1054 col_ano_r <- nomes_raw[str_detect(nomes_raw, regex("Ano.*Refer.ncia", ignore_case =
      ↳ TRUE))][1]
1055 col_mun_r <- nomes_raw[str_detect(nomes_raw, regex("^Munic.pio$", ignore_case =
      ↳ TRUE))][1]
1056 col_uf_r <- nomes_raw[str_detect(nomes_raw, regex("^Estado$", ignore_case =
      ↳ TRUE))][1]
1057 col_sig_r <- nomes_raw[str_detect(nomes_raw, regex("Sigla.*Prestador", ignore_case =
      ↳ TRUE))][1]
1058 col_abr_r <- nomes_raw[str_detect(nomes_raw, regex("^Abrang.ncia$", ignore_case =
      ↳ TRUE))][1]
1059 col_pop_r <- nomes_raw[str_detect(nomes_raw, regex("POP.TOT", ignore_case =
      ↳ TRUE))][1]
1060
1061 data_base_psm <- rawdata %>%
1062   rename(CodigoMunicipio = !!col_cod_r, Ano = !!col_ano_r,
1063     Municipio = !!col_mun_r, UF = !!col_uf_r,
1064     SigladoPrestador = !!col_sig_r, Abrangencia = !!col_abr_r,
1065     population = !!col_pop_r) %>%
1066   mutate(population = suppressWarnings(as.numeric(as.character(population))),
1067     CodigoMunicipio = as.numeric(CodigoMunicipio)) %>%
1068   filter(Ano == 2017) %>%
1069   distinct(CodigoMunicipio, Ano, .keep_all = TRUE)
1070
1071 cat(" Municípios no painel SNIS (2017):", nrow(data_base_psm), "\n\n")
1072
1073
1074 # =====
1075 # 10. COVARIADAS (PIB, SICONFI, PNSB) ----
1076 # =====
1077
1078 cat(rep("=", 80), "\n", sep = "")
1079 cat("PARTE 10: COVARIADAS\n")
1080 cat(rep("=", 80), "\n\n", sep = "")
1081
1082 # PIB per capita 2017
1083 pib_raw <- read_excel("data/PIB dos Municípios - base de dados 2010-2021.xlsx")
1084 pib2017 <- pib_raw %>%
1085   mutate(PIB.1000 = .data[[names(pib_raw)[39]]], PIBpc.1 = .data[[names(pib_raw)[40]]]) %>%
1086   filter(Ano == 2017) %>%
1087   select(PIB.1000, PIBpc.1, Ano, `Código do Município`)
1088
1089 data_base_psm <- data_base_psm %>%
1090   left_join(dicionario, by = join_by(CodigoMunicipio == id_municipio_6)) %>%
1091   left_join(pib2017, by = join_by(id_municipio == `Código do Município`)) %>%
1092   mutate(PIBpercapita = PIBpc.1)
1093
1094 cat(" PIB per capita: OK\n")
1095
1096 # SICONFI 2017
1097 if (file.exists("data/siconfi2017.ods")) {
1098   siconfi <- read_ods("data/siconfi2017.ods") %>%
1099     rename(id_municipio = Cod.IBGE, passivonaocirculante = Valor)
1100   data_base_psm <- data_base_psm %>% left_join(siconfi, by = "id_municipio")
1101   cat(" SICONFI: OK\n")
1102 } else {
1103   data_base_psm$passivonaocirculante <- NA_real_
1104   cat(" SICONFI: arquivo não encontrado NA inserido\n")
1105 }
1106
1107 # Covariadas: Tabela_7499 e Tabela_7505 (IBGE/PNSB 2017)
1108 # Tabela_7499: ligações ativas, ligações micromedidas, economias ativas,
1109 #             volume consumido/dia, volume faturado/dia
1110 # Estrutura: linha 4 = "Município" | linha 5 = nomes das variáveis
1111 #             dados a partir da linha 6 | formato "Nome do Município (UF)"
1112 #

```

```

1113 # Tabela_7505: municípios com controle de perdas físicas (0/1 ou contagem)
1114 #   Estrutura: linha 5 = cabeçalho | dados a partir da linha 6
1115
1116 # Tabela_7499
1117 arq_7499 <- "data/Tabela_7499.xlsx"
1118 if (!file.exists(arq_7499)) arq_7499 <- "data/Tabela_7499.xlsx"
1119
1120 tab7499_raw <- read_excel(arq_7499, sheet = "Tabela",
1121                           col_names = FALSE, col_types = "text")
1122
1123 # Linha 5 contém os nomes das variáveis (índice 5 no tibble)
1124 nomes_7499 <- c("municipio_raw",
1125                as.character(unlist(tab7499_raw[5, 2:6])))
1126
1127 tab7499 <- tab7499_raw %>%
1128   slice(6:n()) %>%                                # dados a partir da linha 6
1129   set_names(nomes_7499) %>%
1130   filter(!is.na(municipio_raw),
1131          !str_detect(municipio_raw, regex("^tabela|^variável|^ano|^tipo|^município$",
1132                                           ignore_case = TRUE))) %>%
1133   mutate(
1134     nome_join = norm(str_trim(str_extract(municipio_raw, "^([+])")),
1135                      sigla_uf = str_trim(str_extract(municipio_raw, "(?<=\\()\\w+")),
1136     across(-c(municipio_raw, nome_join, sigla_uf),
1137            ~suppressWarnings(as.numeric(gsub("-", NA_character_, .))))
1138   ) %>%
1139   left_join(
1140     dicionario %>% mutate(nome_join = norm(nome)) %>%
1141     select(id_municipio, id_municipio_6, nome_join, sigla_uf),
1142     by = c("nome_join", "sigla_uf")
1143   )
1144
1145 # Renomear para nomes padronizados no pipeline
1146 col_lig_at <- names(tab7499)[str_detect(names(tab7499), regex("liga.*ativas.*abastecidas",
1147   ↳ ignore_case = TRUE)) &
1148   ↳ !str_detect(names(tab7499), regex("micromedidas",
1149   ↳ ignore_case = TRUE))][1]
1148 col_lig_mic <- names(tab7499)[str_detect(names(tab7499), regex("micromedidas",
1149   ↳ ignore_case = TRUE))][1]
1149 col_vol_fat <- names(tab7499)[str_detect(names(tab7499), regex("faturado",
1150   ↳ ignore_case = TRUE))][1]
1150 col_vol_con <- names(tab7499)[str_detect(names(tab7499), regex("consumid",
1151   ↳ ignore_case = TRUE))][1]
1151
1152 cat(" Tabela_7499 colunas mapeadas:\n")
1153 cat("   ligacoes_ativas:      ", col_lig_at,   "\n")
1154 cat("   ligacoes_micromedidas:", col_lig_mic,  "\n")
1155 cat("   volume_faturado_dia:   ", col_vol_fat,  "\n")
1156 cat("   volume_consumido_dia:  ", col_vol_con,  "\n")
1157 cat("   Municípios pareados:   ", sum(!is.na(tab7499$id_municipio_6)), "\n\n")
1158
1159 tab7499_final <- tab7499 %>%
1160   select(
1161     id_municipio,
1162     ligacoes_ativas      = any_of(col_lig_at),
1163     ligacoes_micromedidas = any_of(col_lig_mic),
1164     volume_faturado_dia   = any_of(col_vol_fat),
1165     volume_consumido_dia  = any_of(col_vol_con)
1166   ) %>%
1167   filter(!is.na(id_municipio))
1168
1169 # Tabela_7505
1170 arq_7505 <- "data/Tabela_7505.xlsx"
1171
1172 tab7505_raw <- read_excel(arq_7505, sheet = "Tabela",
1173                           col_names = FALSE, col_types = "text")
1174

```

```

1175 tab7505 <- tab7505_raw %>%
1176   slice(6:n()) %>%
1177   set_names(c("municipio_raw", "controle_perdas_raw")) %>%
1178   filter(!is.na(municipio_raw),
1179          !str_detect(municipio_raw, regex("^tabela|^variável|^ano|^tipo|^município$",
1180                                           ignore_case = TRUE))) %>%
1181   mutate(
1182     nome_join = norm(str_trim(str_extract(municipio_raw, "[^()]+")),
1183     sigla_uf = str_trim(str_extract(municipio_raw, "(?<=\\()\\w+")),
1184     # 1 = tem controle de perdas; 0 = não tem ("-")
1185     controle_perdas_agua = if_else(controle_perdas_raw == "-", 0L,
1186                                   suppressWarnings(as.integer(controle_perdas_raw)))
1187   ) %>%
1188   left_join(
1189     dicionario %>% mutate(nome_join = norm(nome)) %>%
1190     select(id_municipio, id_municipio_6, nome_join, sigla_uf),
1191     by = c("nome_join", "sigla_uf")
1192   ) %>%
1193   select(id_municipio, controle_perdas_agua) %>%
1194   filter(!is.na(id_municipio))
1195
1196 cat(" Tabela_7505 municípios pareados:", nrow(tab7505), "\n")
1197 cat(" Com controle de perdas:", sum(tab7505$controle_perdas_agua > 0, na.rm = TRUE), "\n\n")
1198
1199 # Juntar covariadas ao data_base_psm
1200 data_base_psm <- data_base_psm %>%
1201   left_join(tab7499_final, by = "id_municipio") %>%
1202   left_join(tab7505, by = "id_municipio")
1203
1204 # Garantir que todas as covariadas esperadas existam
1205 for (cv in c("ligacoes_ativas", "ligacoes_micromedidas", "controle_perdas_agua",
1206             "volume_faturado_dia", "volume_consumido_dia")) {
1207   if (!cv %in% names(data_base_psm)) data_base_psm[[cv]] <- NA_real_
1208 }
1209
1210 cat(" Covariadas IBGE adicionadas: OK\n\n")
1211
1212
1213 # =====
1214 # 11. MONTAR BASE PSM ----
1215 # =====
1216
1217 cat(rep("=", 80), "\n", sep = "")
1218 cat("PARTE 11: MONTANDO BASE PSM\n")
1219 cat(rep("=", 80), "\n\n", sep = "")
1220
1221 # Lista completa de municípios que já concederam de qualquer forma
1222 # Fonte: concessoes.ods planilha 2 todas as modalidades, todos os anos
1223 # Esses municípios NO podem ser controle, independentemente de modalidade ou ano.
1224
1225 todos_concedentes <- concessoes %>%
1226   # Normalizar e fazer join com dicionário para obter código IBGE 6 dígitos
1227   mutate(
1228     nome_mun_norm = norm(.data[[col_mun_c]]),
1229     nome_estado_norm = norm(.data[[col_uf_c]])
1230   ) %>%
1231   left_join(estados_sigla, by = c("nome_estado_norm" = "nome_estado")) %>%
1232   # Correções de grafia (as mesmas de concessoes_ind)
1233   mutate(nome_mun_norm = case_when(
1234     nome_mun_norm == "CAMPO LARGO + 15 MUNICIPIOS" & sigla_uf == "PR" ~ "CAMPO
1235     ↪ LARGO",
1236     nome_mun_norm == "TANQUE D ARCA" & sigla_uf == "AL" ~ "TANQUE
1237     ↪ D'ARCA",
1238     nome_mun_norm == "OLHO D AGUA DO CASADO" & sigla_uf == "AL" ~ "OLHO
1239     ↪ D'AGUA DO CASADO",
1240     nome_mun_norm == "OLHO D AGUA GRANDE" & sigla_uf == "AL" ~ "OLHO
1241     ↪ D'AGUA GRANDE",

```

```

1238   str_detect(nome_mun_norm, "^RIO DE JANEIRO")           & sigla_uf == "RJ" ~ "RIO DE
1239   ↪ JANEIRO",
1240   str_detect(nome_mun_norm, "^SAO PAULO")               & sigla_uf == "SP" ~ "SAO
1241   ↪ PAULO",
1242   nome_mun_norm == "ELDORADO DOS CARAJAS"               & sigla_uf == "PA" ~
1243   ↪ "ELDORADO DO CARAJAS",
1244   TRUE ~ nome_mun_norm
1245 )) %>%
1246 left_join(dicionario_join %>% select(id_municipio_6, nome_join, sigla_uf),
1247           by = c("nome_mun_norm" = "nome_join", "sigla_uf")) %>%
1248 # Fallback Tabocão
1249 mutate(id_municipio_6 = case_when(
1250   nome_mun_norm == "TABOCAO" & sigla_uf == "TO" & is.na(id_municipio_6) ~ 172097L,
1251   TRUE ~ as.integer(id_municipio_6)
1252 )) %>%
1253 filter(!is.na(id_municipio_6)) %>%
1254 pull(id_municipio_6) %>%
1255 unique()
1256
1257 # Adicionar municípios CORSAN (RS) também são concedentes
1258 if (exists("municipios_corsan") && nrow(municipios_corsan) > 0) {
1259   todos_concedentes <- unique(c(todos_concedentes,
1260                                 as.integer(municipios_corsan$CódigoMunicípio)))
1261 }
1262
1263 cat(" Total de municípios concedentes (qualquer modalidade/ano):",
1264     length(todos_concedentes), "\n\n")
1265
1266 # Abrangência local e microrregional como dummies
1267 data_psm <- data_base_psm %>%
1268 mutate(
1269   abrangencia_local = if_else(str_detect(Abrangencia, regex("local", ignore_case = TRUE)),
1270     ↪ 1L, 0L),
1271   abrangencia_micro = if_else(str_detect(Abrangencia, regex("microrregional", ignore_case =
1272     ↪ TRUE)), 1L, 0L)
1273 ) %>%
1274 mutate(
1275   # Desenho do estudo
1276   # Tratados (private = 1): concessão PLENA com anodocontrato >= 2020
1277   # Controles (private = 0): NUNCA concederam de nenhuma forma
1278   # Excluídos (private = 2): concederam em qualquer modalidade/ano,
1279   # mas não são tratados saem da amostra
1280   private = case_when(
1281     CodigoMunicípio %in% abcon$CódigoMunicípio ~ 1L,           # tratados
1282     CodigoMunicípio %in% todos_concedentes      ~ 2L,           # outros concedentes
1283     TRUE                                           ~ 0L           # nunca concederam
1284   ),
1285   first_treat = case_when(
1286     private == 1L ~ as.integer(anodocontrato_map[as.character(CodigoMunicípio)]),
1287     private == 0L ~ 0L,
1288     TRUE          ~ NA_integer_
1289   )
1290 ) %>%
1291 filter(
1292   private != 2L
1293   #!CodigoMunicípio %in% c(330455L, 355030L, 230420L, 521010L)
1294 ) %>%
1295 mutate(grupo = if_else(private == 1L, "Tratado (concessão 2020)", "Controle
1296   ↪ (never-treated)"))
1297
1298 cat(" Base total:", nrow(data_psm), "\n")
1299 cat(" Tratados:", sum(data_psm$private == 1), "\n")
1300 cat(" Controles:", sum(data_psm$private == 0), "\n\n")
1301
1302 # Diagnóstico de NAs
1303 covariadas_candidatas <- c("population", "PIBpercapita", "passivonaocirculante",
1304   ↪ "ligacoes_ativas", "ligacoes_micromedidas", "controle_perdas_agua",

```

```

1299         "volume_faturado_dia", "volume_consumido_dia",
1300         "abrangencia_local", "abrangencia_micro")
1301
1302 na_pct <- data_psm %>%
1303   summarise(across(all_of(intersect(covariadas_candidatas, names(data_psm))),
1304     ~mean(is.na(.)) * 100)) %>%
1305   pivot_longer(everything(), names_to = "variavel", values_to = "pct")
1306
1307 cat("NAs por covariada (%):\n"); print(na_pct, n = Inf); cat("\n")
1308
1309
1310 # =====
1311 # 12. COMPARAO DE ESTRATEGIAS DE PAREAMENTO ----
1312 # =====
1313
1314 cat(rep("=", 80), "\n", sep = "")
1315 cat("PARTE 12: COMPARAO DE ESTRATEGIAS\n")
1316 cat(rep("=", 80), "\n\n", sep = "")
1317
1318 # Trs estratégias de seleção de covariadas (como na tabela do artigo)
1319 estrategias <- list(
1320   Completo = c("population", "PIBpercapita", "passivonaocirculante",
1321     "ligacoes_ativas", "ligacoes_micromedidas", "controle_perdas_agua",
1322     "volume_faturado_dia", "volume_consumido_dia",
1323     "abrangencia_local", "abrangencia_micro"),
1324   Rigoroso = c("population", "PIBpercapita", "passivonaocirculante"),
1325   Flexivel = c("abrangencia_local", "PIBpercapita", "passivonaocirculante",
1326     "ligacoes_micromedidas", "abrangencia_micro", "volume_faturado_dia")
1327 )
1328
1329 resultados_estrategias <- list()
1330
1331 for (nome_est in names(estrategias)) {
1332   covs_est <- intersect(estrategias[[nome_est]], names(data_psm))
1333   # Remover covariadas com >50% NAs nesta estratégia
1334   na_est <- data_psm %>%
1335     summarise(across(all_of(covs_est), ~mean(is.na(.)))) %>%
1336     pivot_longer(everything(), names_to = "v", values_to = "p")
1337   covs_est <- na_est %>% filter(p <= 0.5) %>% pull(v)
1338
1339   dados_est <- data_psm %>% filter(if_all(all_of(covs_est), ~!is.na(.)))
1340   if (sum(dados_est$private == 1) < 2) next
1341
1342   formula_est <- as.formula(paste("private ~", paste(covs_est, collapse = " + ")))
1343   m_est <- tryCatch(
1344     suppressWarnings(matchit(formula_est, data = dados_est, method = "nearest", ratio = 1,
1345       ↪ replace = FALSE)),
1346     error = function(e) NULL
1347   )
1348   if (is.null(m_est)) next
1349
1350   sm_est <- summary(m_est, un = FALSE)
1351   smd_pos <- sm_est$sum.matched[, "Std. Mean Diff."]
1352   dp_match <- match.data(m_est)
1353
1354   resultados_estrategias[[nome_est]] <- list(
1355     matchit = m_est,
1356     pareados = dp_match,
1357     smd_medio = round(mean(abs(smd_pos), na.rm = TRUE), 4),
1358     smd_max = round(max(abs(smd_pos), na.rm = TRUE), 4),
1359     n_variaveis = length(covs_est),
1360     n_tratados = sum(dp_match$private == 1),
1361     n_controles = sum(dp_match$private == 0)
1362   )
1363 }
1364
1365 # Tabela de comparação (como no artigo)

```

```

1365 tab_comparacao <- map_dfr(names(resultados_estrategias), function(nm) {
1366   r <- resultados_estrategias[[nm]]
1367   tibble(
1368     Estratégia = nm,
1369     `N Var.` = r$n_variaveis,
1370     Tratados = r$n_tratados,
1371     Controles = r$n_controles,
1372     Total = r$n_tratados + r$n_controles,
1373     `SMD Médio` = r$smd_medio,
1374     `SMD Máx.` = r$smd_max
1375   )
1376 })
1377
1378 cat("COMPARAO DAS ESTRATEGIAS:\n")
1379 print(tab_comparacao)
1380 cat("\n")
1381
1382 # Tabela LaTeX
1383 latex_comp <- tab_comparacao %>%
1384   kbl(format = "latex", booktabs = TRUE,
1385     caption = "Comparação das 3 Estratégias de Seleção de Covariadas para PSM (com
1386       ↪ Abrangência)",
1387     label = "tab:comparacao_psm") %>%
1388   kable_styling()
1389 writeLines(latex_comp, "outputs/tabela_comparacao_psm.tex")
1390 cat(" Tabela salva: outputs/tabela_comparacao_psm.tex\n\n")
1391
1392 # =====
1393 # 13. PAREAMENTO PSM ESTRATEGIA FLEXÍVEL (ratios 1, 2 e 5) ----
1394 # =====
1395
1396 cat(rep("=", 80), "\n", sep = "")
1397 cat("PARTE 13: PSM FLEXÍVEL RATIOS 1:1, 1:2 E 1:5\n")
1398 cat(rep("=", 80), "\n\n", sep = "")
1399
1400 covariadas_flex <- intersect(estrategias$Flexivel, names(data_psm))
1401 na_flex <- data_psm %>%
1402   summarise(across(all_of(covariadas_flex), ~mean(is.na(.)))) %>%
1403   pivot_longer(everything(), names_to = "v", values_to = "p")
1404 covariadas_flex <- na_flex %>% filter(p <= 0.5) %>% pull(v)
1405 dados_flex <- data_psm %>% filter(if_all(all_of(covariadas_flex), ~!is.na(.)))
1406
1407 formula_flex <- as.formula(paste("private ~", paste(covariadas_flex, collapse = " + ")))
1408
1409 ratios_psm <- c(1, 2, 5)
1410 pareamentos <- list()
1411
1412 for (r in ratios_psm) {
1413   cat(sprintf(" Ratio 1:%d ...", r))
1414   m <- tryCatch(
1415     matchit(formula_flex, data = dados_flex, method = "nearest", ratio = r, replace = FALSE),
1416     error = function(e) { cat(" ERRO:", e$message, "\n"); NULL }
1417   )
1418   if (!is.null(m)) {
1419     dp <- match.data(m)
1420     sm <- summary(m, un = FALSE)
1421     smd <- sm$sum.matched[, "Std. Mean Diff."]
1422     pareamentos[[paste0("ratio", r)] <- list(matchit = m, pareados = dp,
1423       smd_medio = mean(abs(smd), na.rm = TRUE),
1424       smd_max = max(abs(smd), na.rm = TRUE))
1425     cat(sprintf(" OK | N=%d | SMD médio=%.4f | SMD máx=%.4f\n",
1426       nrow(dp), mean(abs(smd), na.rm = TRUE), max(abs(smd), na.rm = TRUE)))
1427     write_csv(dp, sprintf("data/amostra_pareada_psm_ratio%d.csv", r))
1428     saveRDS(m, sprintf("data/pareamento_psm_ratio%d.rds", r))
1429   }
1430 }

```

```

1431
1432 # Selecionar automaticamente o ratio com menor SMD médio
1433 # Critério de desempate: maior N (mais observações)
1434 ratio_smd <- map_dfr(names(pareamentos), function(nm) {
1435   tibble(ratio = nm,
1436           smd_medio = pareamentos[[nm]]$smd_medio,
1437           smd_max    = pareamentos[[nm]]$smd_max,
1438           n          = nrow(pareamentos[[nm]]$pareados))
1439 }) %>% arrange(smd_medio, desc(n))
1440
1441 cat("\nComparação de ratios:\n")
1442 print(ratio_smd)
1443
1444 ratio_escolhido <- ratio_smd$ratio[1]
1445 cat(sprintf("\nRatio escolhido automaticamente: %s (SMD médio = %.4f)\n\n",
1446           ratio_escolhido, ratio_smd$smd_medio[1]))
1447
1448 dados_pareados <- pareamentos[[ratio_escolhido]]$pareados
1449 cat(sprintf("Amostra padrão (%s): %d tratados + %d controles\n\n",
1450           ratio_escolhido,
1451           sum(dados_pareados$private == 1),
1452           sum(dados_pareados$private == 0)))
1453
1454
1455 # =====
1456 # 14. TABELAS DE BALANO ----
1457 # =====
1458
1459 cat(rep("=", 80), "\n", sep = "")
1460 cat("PARTE 14: TABELAS DE BALANO\n")
1461 cat(rep("=", 80), "\n\n", sep = "")
1462
1463 nomes_legiveis <- c(
1464   population      = "População total",
1465   PIBpercapita    = "PIB per capita",
1466   passivonaocirculante = "Passivo não circulante",
1467   ligacoes_ativas  = "Ligações ativas abastecidas",
1468   ligacoes_micromedidas = "Ligações ativas micromedidas",
1469   controle_perdas_agua = "Controle de perdas físicas",
1470   volume_faturado_dia = "Volume faturado por dia (m)",
1471   volume_consumido_dia = "Volume consumido por dia (m)",
1472   abrangencia_local  = "Abrangência: local",
1473   abrangencia_micro  = "Abrangência: microrregional"
1474 )
1475
1476 calcular_balance <- function(dados, var) {
1477   trat <- dados[[var]][dados$private == 1]; cont <- dados[[var]][dados$private == 0]
1478   trat <- trat[!is.na(trat)]; cont <- cont[!is.na(cont)]
1479   if (length(trat) < 2 || length(cont) < 2) return(NULL)
1480   t_res <- t.test(trat, cont)
1481   sd_pool <- sqrt((var(trat) + var(cont)) / 2)
1482   smd <- (mean(trat) - mean(cont)) / sd_pool
1483   data.frame(
1484     variavel      = nomes_legiveis[var],
1485     Tratados     = length(trat),
1486     Controle     = length(cont),
1487     t_stat       = round(as.numeric(t_res$statistic), 3),
1488     p_valor      = round(t_res$p.value, 4),
1489     Significância = case_when(
1490       t_res$p.value < 0.001 ~ "***", t_res$p.value < 0.01 ~ "**",
1491       t_res$p.value < 0.05 ~ "*",   t_res$p.value < 0.10 ~ "ns",
1492       TRUE ~ "ns"
1493     ),
1494     smd          = round(smd, 4)
1495   )
1496 }
1497

```

```

1498 balance_antes <- map_dfr(covariadas_flex, ~calcular_balance(dados_flex, .x)) %>%
      ↪ mutate(Amostra = "Completa")
1499 balance_depois <- map_dfr(covariadas_flex, ~calcular_balance(dados_pareados, .x)) %>%
      ↪ mutate(Amostra = "Pareada")
1500 balance_total <- bind_rows(balance_antes, balance_depois)
1501
1502 # Tabela LaTeX (formato do artigo)
1503 latex_balance <- balance_total %>%
1504   select(Variável = variavel, Tratados, Controle,
1505     `Teste t` = t_stat, `Valor-p` = p_valor,
1506     Significância, Amostra) %>%
1507   kbl(format = "latex", booktabs = TRUE, escape = FALSE, row.names = FALSE,
1508     caption = paste0("Diferença de Médias entre Grupos Tratamento e Controle para ",
1509       "Amostras Sem Pareamento e Com Pareamento (2017) Estratégia FLEXÍVEL.
1510       ↪ Brasil, 2025."),
1511     label = "tab:balance_psm") %>%
1512   kable_styling(font_size = 9, latex_options = "repeat_header") %>%
1513   footnote(general = c("Significância: *** p<0.001; ** p<0.01; * p<0.05.",
1514     "Fonte: Resultados da Pesquisa"))
1514
1515 writeLines(latex_balance, "outputs/tabela_balance_psm.tex")
1516 cat(" Tabela salva: outputs/tabela_balance_psm.tex\n\n")
1517
1518
1519 # =====
1520 # 15. MAPA DA AMOSTRA PAREADA ----
1521 # =====
1522
1523 cat(rep("=", 80), "\n", sep = "")
1524 cat("PARTE 15: MAPA DA AMOSTRA PAREADA\n")
1525 cat(rep("=", 80), "\n\n", sep = "")
1526
1527 mapa_municipios <- tryCatch(
1528   read_municipality(year = 2020, showProgress = FALSE),
1529   error = function(e) {
1530     unlink(rappdirs::user_cache_dir("geobr"), recursive = TRUE)
1531     read_municipality(year = 2020, showProgress = FALSE)
1532   }
1533 ) %>% mutate(code_muni = as.character(code_muni))
1534
1535 mapa_amostra <- dados_pareados %>%
1536   select(id_municipio, private) %>%
1537   mutate(
1538     id_municipio = as.character(id_municipio),
1539     categoria = if_else(private == 1,
1540       "Tratado (concessão 2020)",
1541       "Controle (never-treated)")
1542   )
1543
1544 mapa_dados <- mapa_municipios %>%
1545   left_join(mapa_amostra, by = c("code_muni" = "id_municipio")) %>%
1546   mutate(na_amostra = !is.na(categoria))
1547
1548 smd_pos <- summary(pareamentos[[ratio_escolhido]]$matchit, un = FALSE)$sum.matched[, "Std.
      ↪ Mean Diff."]
1549
1550 mapa_plot <- ggplot() +
1551   # Municípios fora da amostra fundo cinza sem legenda
1552   geom_sf(data = filter(mapa_dados, !na_amostra),
1553     fill = "#E8E8E8", color = "white", linewidth = 0.03) +
1554   # Municípios da amostra tratados e controles
1555   geom_sf(data = filter(mapa_dados, na_amostra),
1556     aes(fill = categoria), color = "white", linewidth = 0.05) +
1557   scale_fill_manual(
1558     values = c("Tratado (concessão 2020)" = "#D7191C",
1559       "Controle (never-treated)" = "#2C7BB6"),
1560     name = NULL

```

```

1561 ) +
1562 theme_void() +
1563 theme(
1564   legend.position = "bottom",
1565   legend.text      = element_text(size = 11),
1566   legend.key.size  = unit(0.6, "cm"),
1567   plot.title       = element_text(hjust = 0.5, face = "bold", size = 13),
1568   plot.subtitle    = element_text(hjust = 0.5, size = 10, color = "gray40"),
1569   plot.caption     = element_text(hjust = 0, size = 8, color = "gray50"),
1570   plot.background  = element_rect(fill = "white", color = NA),
1571   plot.margin      = margin(10, 10, 10, 10)
1572 ) +
1573 labs(
1574   title      = "Amostra Pareada (PSM) Concessões de Saneamento",
1575   subtitle   = sprintf("N = %d | %d tratados + %d controles | SMD médio = %.4f",
1576                         nrow(dados_pareados),
1577                         sum(dados_pareados$private == 1),
1578                         sum(dados_pareados$private == 0),
1579                         mean(abs(smd_pos), na.rm = TRUE)),
1580   caption    = "Fonte: Elaboração própria com base em concessoes.ods, SNIS (2017) e IBGE
1581               ↪ (2017).",
1582   guides(fill = guide_legend(nrow = 1))
1583 )
1584 ggsave("outputs/mapa_amostra_pareada_nova.png",
1585        plot = mapa_plot, width = 10, height = 12, dpi = 300, bg = "white")
1586 cat(" Mapa salvo: outputs/mapa_amostra_pareada.png\n\n")
1587
1588 # =====
1589 # 16. MONTAR PAINIS DID (ratio 1, 2 e 5) ----
1590 # =====
1591
1592 cat(rep("=", 80), "\n", sep = "")
1593 cat("PARTE 16: MONTANDO PAINIS DID\n")
1594 cat(rep("=", 80), "\n\n", sep = "")
1595
1596 # Lógica de chave de ligação
1597 # A amostra PSM (data_psm / dados_pareados) foi construída a partir do rawdata
1598 # SNIS com Codigomunicipio de 6 dígitos (id_municipio_6 do IBGE).
1599 # O painel_completo tem Codigomunicipio de 7 dígitos (id_municipio do IBGE).
1600 # A chave cannica para o join é o código de 6 dígitos convertemos o painel
1601 # para 6 dígitos na hora do join, e depois restauramos o de 7 dígitos.
1602
1603 # Mapa 7 -> 6 dígitos para uso no join
1604 mapa_7_para_6 <- dicionario %>%
1605   select(id_municipio, id_municipio_6) %>%
1606   mutate(across(everything(), as.character))
1607
1608 # Verificar qual coluna de código IBGE a amostra PSM usa
1609 cod_psm_ex <- head(as.character(dados_pareados$CodigoMunicipio), 3)
1610 cat(" Exemplos de CodigoMunicipio na amostra PSM:", paste(cod_psm_ex, collapse = ", "), "\n")
1611 cat(" Exemplos de CodigoMunicipio no painel:      ",
1612     paste(head(as.character(painel_completo$CodigoMunicipio), 3), collapse = ", "), "\n\n")
1613
1614 montar_painel_did <- function(amostra_pareada, painel_indicadores) {
1615
1616   # Chave da amostra: CodigoMunicipio de 6 dígitos
1617   chave_psm <- amostra_pareada %>%
1618     transmute(
1619       cod6      = as.character(CodigoMunicipio),
1620       private   = private,
1621       first_treat = first_treat
1622     ) %>%
1623     distinct(cod6, .keep_all = TRUE)
1624
1625   # Painel: converter CodigoMunicipio de 7 para 6 dígitos para o join

```

```

1627 painel_com_cod6 <- painel_indicadores %>%
1628   mutate(CodigoMunicipio7 = as.character(CodigoMunicipio)) %>%
1629   left_join(mapa_7_para_6, by = c("CodigoMunicipio7" = "id_municipio")) %>%
1630   mutate(
1631     # Se já era 6 dígitos, id_municipio_6 será NA usar o original
1632     cod6 = coalesce(id_municipio_6, CodigoMunicipio7)
1633   )
1634
1635 # Join por cod6
1636 painel_out <- painel_com_cod6 %>%
1637   filter(cod6 %in% chave_psm$cod6) %>%
1638   left_join(chave_psm, by = "cod6") %>%
1639   filter(!is.na(private)) %>%
1640   mutate(
1641     first_treat = as.integer(coalesce(first_treat, 0L)),
1642     # Restaurar CodigoMunicipio original (7 dígitos)
1643     CodigoMunicipio = CodigoMunicipio7
1644   ) %>%
1645   select(-CodigoMunicipio7, -cod6, -id_municipio_6)
1646
1647   return(painel_out)
1648 }
1649
1650 for (r in ratios_psm) {
1651   chave <- paste0("ratio", r)
1652   if (!chave %in% names(pareamentos)) next
1653
1654   p_did <- montar_painel_did(pareamentos[[chave]]$pareados, painel_completo)
1655   write_csv(p_did, sprintf("data/painel_did_ratio%d.csv", r))
1656   cat(sprintf(" Painel DID ratio %d: %d linhas, %d municípios (%d tratados + %d
1657     ↪ controles)\n",
1658     r, nrow(p_did), n_distinct(p_did$CodigoMunicipio),
1659     n_distinct(p_did$CodigoMunicipio[p_did$private == 1]),
1660     n_distinct(p_did$CodigoMunicipio[p_did$private == 0])))
1661 }
1662
1663 # Painel principal = ratio escolhido automaticamente
1664 painel_did <- read_csv(sprintf("data/painel_did_ratio%s.csv",
1665   str_extract(ratio_escolhido, "\\d+")),
1666   show_col_types = FALSE) %>%
1667   mutate(CodigoMunicipio = as.character(CodigoMunicipio),
1668     id_num = as.integer(as.factor(CodigoMunicipio)))
1669 cat("\nPainel DID principal:\n")
1670 cat(" Linhas:", nrow(painel_did), "\n")
1671 cat(" Tratados:", n_distinct(painel_did$CodigoMunicipio[painel_did$private == 1]), "\n")
1672 cat(" Controles:", n_distinct(painel_did$CodigoMunicipio[painel_did$private == 0]), "\n\n")
1673
1674 # =====
1675 # 17. GRFICOS DE TENDENCIA ----
1676 # =====
1677
1678 cat(rep("=", 80), "\n", sep = "")
1679 cat("PARTE 17: GRFICOS DE TENDENCIA\n")
1680 cat(rep("=", 80), "\n\n", sep = "")
1681
1682 vars_tendencia <- intersect(c("IAG0001", "IFA1001", "IFA2008"), names(painel_did))
1683
1684 labels_vars <- c(
1685   IAG0001 = "IAG0001 Atendimento da população total com rede de abastecimento de água (%)",
1686   IFA1001 = "IFA1001 Receita operacional direta média de usuários de água (R$/m)",
1687   IFA2008 = "IFA2008 Incidência da despesa de pessoal próprio nas despesas de exploração de á
1688     ↪ gua (%) "
1689 )
1690 cores_grupo <- c("Controle (never-treated)" = "#2C7BB6", "Tratado (concessão 2020)" =
1691   ↪ "#D7191C")

```

```

1691 for (var in vars_tendencia) {
1692   tendencia <- painel_did %>%
1693     filter(!is.na(.data[[var]])) %>%
1694     mutate(grupo = factor(private, levels = c(0, 1),
1695                            labels = c("Controle (never-treated)", "Tratado (concessão
1696                                     ↳ 2020)")) %>%
1697     group_by(grupo, Ano) %>%
1698     summarise(media = mean(.data[[var]], na.rm = TRUE),
1699               mediana = median(.data[[var]], na.rm = TRUE),
1700               .groups = "drop") %>%
1701     pivot_longer(c(media, mediana), names_to = "estatistica", values_to = "valor") %>%
1702     mutate(estat_label = if_else(estatistica == "media", "Média", "Mediana"))
1703
1704 p <- ggplot(tendencia, aes(x = Ano, y = valor, color = grupo, linetype = estat_label)) +
1705   geom_line(linewidth = 0.9) + geom_point(size = 2.5) +
1706   geom_vline(xintercept = 2019.5, linetype = "dotted", color = "gray50") +
1707   annotate("text", x = 2019.7, y = Inf, vjust = 1.5, label = "Início tratamento ",
1708            hjust = 0, size = 3, color = "gray40") +
1709   scale_color_manual(values = cores_grupo, name = "Grupo") +
1710   scale_linetype_manual(values = c("Média" = "solid", "Mediana" = "dashed"), name =
1711                           ↳ "Estatística") +
1712   scale_x_continuous(breaks = 2017:2024) +
1713   labs(title = paste0("Tendência ", labels_vars[var]),
1714        subtitle = "Comparação entre grupos tratado e controle (amostra PSM-FLEXÍVEL)",
1715        x = "Ano", y = var,
1716        caption = "Fonte: Elaboração própria. Linha pontilhada: início do tratamento
1717                  ↳ (2020).") +
1718   theme_minimal(base_size = 12) +
1719   theme(plot.title = element_text(face = "bold", hjust = 0.5),
1720         plot.subtitle = element_text(hjust = 0.5, color = "gray40", size = 9),
1721         legend.position = "bottom", panel.grid.minor = element_blank())
1722
1723 ggsave(paste0("outputs/did/plots/tendencia_", var, ".png"),
1724        plot = p, width = 10, height = 6, dpi = 300, bg = "white")
1725 cat(" Salvo: tendencia_", var, ".png\n")
1726 }
1727 cat("\n")
1728
1729 # =====
1730 # 18. ESTATÍSTICAS DESCRITIVAS POR GRUPO ----
1731 # =====
1732
1733 cat(rep("=", 80), "\n", sep = "")
1734 cat("PARTE 18: ESTATÍSTICAS DESCRITIVAS\n")
1735 cat(rep("=", 80), "\n\n", sep = "")
1736
1737 outcomes_todos <- names(painel_did)[str_detect(names(painel_did), "^I[A-Z0-9]")]
1738 miss_df <- painel_did %>%
1739   select(all_of(outcomes_todos)) %>%
1740   summarise(across(everything(), ~mean(is.na(.)) * 100)) %>%
1741   pivot_longer(everything(), names_to = "outcome", values_to = "pct_missing") %>%
1742   arrange(pct_missing)
1743
1744 outcomes_validos_did <- miss_df %>% filter(pct_missing < 70) %>% pull(outcome)
1745 cat(" Outcomes válidos (<70% missing):", length(outcomes_validos_did), "\n\n")
1746
1747 calc_stats_grp <- function(x) {
1748   x <- x[!is.na(x)]
1749   if (length(x) == 0) return(tibble(n=0L, media=NA, mediana=NA, dp=NA))
1750   tibble(n=length(x), media=mean(x), mediana=median(x), dp=sd(x))
1751 }
1752
1753 desc_stats <- map_dfr(outcomes_validos_did, function(out) {
1754   map_dfr(c(0, 1), function(grp) {
1755     calc_stats_grp(painel_did %>% filter(private == grp) %>% pull(out)) %>%
1756     mutate(outcome = out,

```

```

1755     grupo = if_else(grp == 1, "Tratado", "Controle"),
1756     .before = n)
1757   })
1758 }) %>% arrange(outcome, desc(grupo))
1759
1760 write_csv(desc_stats, "outputs/did/estatisticas_descritivas.csv")
1761
1762 # Dicionário com nome + unidade a partir do arquivo SINISA
1763
1764 extrair_dicionario_sinisa <- function(arquivo) {
1765   xl <- readxl::excel_sheets(arquivo)
1766   pares <- xl[seq(2, length(xl), by = 2)] # planilhas com dados
1767   linhas <- list()
1768
1769   for (pl in pares) {
1770     df <- read_excel(arquivo, sheet = pl)
1771     cols <- names(df)
1772
1773     col_cod <- cols[str_detect(cols, regex("C.DIGO", ignore_case = TRUE))][1]
1774     col_nome <- cols[str_detect(cols, regex("NOME.*INDICADOR", ignore_case = TRUE))][1]
1775     col_uni <- cols[str_detect(cols, regex("UNIDADE", ignore_case = TRUE))][1]
1776
1777     if (!is.na(col_cod) && !is.na(col_nome) && !is.na(col_uni)) {
1778       sub <- df %>%
1779         select(codigo = !!col_cod,
1780              nome = !!col_nome,
1781              unidade = !!col_uni) %>%
1782         filter(!is.na(codigo), str_detect(codigo, "^I[A-Z]{2}\\d"))
1783       linhas[[pl]] <- sub
1784     }
1785   }
1786
1787   bind_rows(linhas) %>%
1788     mutate(
1789       nome = str_trim(str_replace_all(nome, "\\s+", " ")),
1790       unidade = str_trim(str_replace_all(unidade, "\\s+", " "))
1791     ) %>%
1792     distinct(codigo, .keep_all = TRUE)
1793 }
1794
1795 dicionario_indicadores <- extrair_dicionario_sinisa(
1796   "data/INDICADORES_SINISA_ABASTECIMENTO_DE_AGUA_2024_v2.xlsx"
1797 )
1798
1799 # Montar legenda: "CDIGO: Nome (unidade)" para cada indicador na tabela
1800
1801 legenda_texto <- dicionario_indicadores %>%
1802   filter(codigo %in% outcomes_validos_did) %>%
1803   arrange(codigo) %>%
1804   mutate(entrada = paste0(codigo, ": ", nome, " (", unidade, ")")) %>%
1805   pull(entrada) %>%
1806   paste(collapse = "; ")
1807
1808 # Indicadores sem correspondência no dicionário (ex: códigos com 'G')
1809 sem_dic <- setdiff(
1810   outcomes_validos_did,
1811   dicionario_indicadores$codigo
1812 )
1813 if (length(sem_dic) > 0)
1814   cat(" Aviso indicadores sem descrição no dicionário:", paste(sem_dic, collapse = ", "),
1815       "\n")
1816
1817 # Tabela LaTeX wide
1818
1819 desc_wide <- desc_stats %>%
1820   mutate(across(c(media, mediana, dp), ~round(.x, 3))) %>%
1821   pivot_wider(

```

```

1821     names_from = grupo,
1822     values_from = c(n, media, mediana, dp),
1823     names_glue = "{grupo}_{.value}"
1824 ) %>%
1825 select(outcome,
1826         Controle_n, Controle_media, Controle_mediana, Controle_dp,
1827         Tratado_n, Tratado_media, Tratado_mediana, Tratado_dp)
1828
1829 latex_desc <- desc_wide %>%
1830 kbl(
1831     format      = "latex",
1832     booktabs    = TRUE,
1833     col.names   = c("Indicador",
1834                     "N", "Média", "Mediana", "DP",
1835                     "N", "Média", "Mediana", "DP"),
1836     align       = c("l", rep("r", 8)),
1837     caption     = paste0(
1838         "Estatísticas Descritivas dos Indicadores por Grupo",
1839         " --- Amostra PSM-FLEXÍVEL, Brasil, 2017--2024."
1840     ),
1841     label       = "tab:desc_stats",
1842     escape      = FALSE,
1843     row.names   = FALSE
1844 ) %>%
1845 kable_styling(latex_options = c("hold_position", "scale_down")) %>%
1846 add_header_above(c(
1847     " "                        = 1,
1848     "Controle (never-treated)" = 4,
1849     "Tratado (concessão 2020)" = 4
1850 )) %>%
1851 footnote(
1852     general = c(
1853         "DP = desvio padrão.",
1854         paste0("Descrição e unidade dos indicadores --- ", legenda_texto)
1855     ),
1856     general_title = "",
1857     footnote_as_chunk = FALSE,
1858     escape = FALSE,
1859     threeparttable = TRUE # mantém a nota dentro da largura da tabela
1860 )
1861
1862 writeLines(latex_desc, "outputs/did/tabela_descritivas.tex")
1863 cat(" Tabela salva: outputs/did/tabela_descritivas.tex\n\n")
1864
1865 # =====
1866 # 19. ESTIMAO CALLAWAY & SANT'ANNA ----
1867 # =====
1868
1869 cat(rep("=", 80), "\n", sep = "")
1870 cat("PARTE 19: ESTIMAO CALLAWAY & SANT'ANNA\n")
1871 cat(rep("=", 80), "\n\n", sep = "")
1872
1873
1874 estimar_cs <- function(dados, outcome_var) {
1875     dados_out <- dados %>%
1876         filter(!is.na(.data[[outcome_var]])) %>%
1877         mutate(id_cs = as.integer(as.factor(CodigoMunicipio)))
1878     n_trat <- n_distinct(dados_out$CodigoMunicipio[dados_out$first_treat > 0])
1879     n_cont <- n_distinct(dados_out$CodigoMunicipio[dados_out$first_treat == 0])
1880     if (n_trat < 2 || n_cont < 2) { cat(" SKIP", outcome_var, "\n"); return(NULL) }
1881     tryCatch(
1882         att_gt(yname = outcome_var, tname = "Ano", idname = "id_cs",
1883               gname = "first_treat", data = dados_out,
1884               control_group = "nevertreated", est_method = "dr",
1885               clustervars = "id_cs", anticipation = 0,
1886               allow_unbalanced_panel = TRUE, print_details = FALSE),
1887         error = function(e) { cat(" ERRO:", outcome_var, ":", conditionMessage(e), "\n"); NULL }

```

```

1888   )
1889 }
1890
1891 resultados_cs <- list()
1892 resultados_agr <- list()
1893
1894 for (out in outcomes_validos_did) {
1895   cat("  Estimando:", out, "...")
1896   res <- estimar_cs(painel_did, out)
1897   if (!is.null(res)) {
1898     resultados_cs[[out]] <- res
1899     resultados_agr[[out]] <- list(
1900       simple = tryCatch(aggte(res, type = "simple"), error = function(e) NULL),
1901       group = tryCatch(aggte(res, type = "group"), error = function(e) NULL),
1902       dynamic = tryCatch(aggte(res, type = "dynamic"), error = function(e) NULL)
1903     )
1904     cat(" \n")
1905   }
1906 }
1907 cat("\nConcluído:", length(resultados_cs), "/", length(outcomes_validos_did), "\n\n")
1908
1909 saveRDS(resultados_cs, "data/resultados_cs_att_gt.rds")
1910 saveRDS(resultados_agr, "data/resultados_cs_agregados.rds")
1911
1912
1913 # =====
1914 # 20. TABELAS INDIVIDUAIS + EVENT-STUDY (IAG0001, IFA1001, IFA2008) ----
1915 # =====
1916
1917 cat(rep("=", 80), "\n", sep = "")
1918 cat("PARTE 20: TABELAS INDIVIDUAIS\n")
1919 cat(rep("=", 80), "\n\n", sep = "")
1920
1921 tabela_por_grupo <- function(out) {
1922   res_grp <- resultados_agr[[out]]$group
1923   if (is.null(res_grp)) return(NULL)
1924   tibble(Cohort = res_grp$egt, ATT = res_grp$att.egt, SE = res_grp$se.egt,
1925     CI_inf = res_grp$att.egt - 1.96 * res_grp$se.egt,
1926     CI_sup = res_grp$att.egt + 1.96 * res_grp$se.egt,
1927     p_valor = 2 * pnorm(-abs(res_grp$att.egt / res_grp$se.egt)),
1928     Sig = case_when(p_valor < 0.001 ~ "***", p_valor < 0.01 ~ "**",
1929       p_valor < 0.05 ~ "*", p_valor < 0.10 ~ ".", TRUE ~ ""))
1930 }
1931
1932 att_simples <- function(out) {
1933   s <- resultados_agr[[out]]$simple
1934   if (is.null(s)) return(NULL)
1935   pval <- 2 * pnorm(-abs(s$overall.att / s$overall.se))
1936   tibble(ATT = s$overall.att, SE = s$overall.se,
1937     CI_inf = s$overall.att - 1.96 * s$overall.se,
1938     CI_sup = s$overall.att + 1.96 * s$overall.se, p_valor = pval,
1939     Sig = case_when(pval < 0.001 ~ "***", pval < 0.01 ~ "**",
1940       pval < 0.05 ~ "*", pval < 0.10 ~ ".", TRUE ~ ""))
1941 }
1942
1943 vars_principais <- intersect(c("IAG0001", "IFA1001", "IFA2008"), names(resultados_agr))
1944
1945 for (out in vars_principais) {
1946   cat("----", out, "---\n")
1947   att_s <- att_simples(out)
1948   tbl_grp <- tabela_por_grupo(out)
1949
1950   if (!is.null(tbl_grp) && !is.null(att_s)) {
1951     tbl_latex <- bind_rows(
1952       tbl_grp %>%
1953         mutate(Cohort = as.character(Cohort),
1954           ATT_fmt = paste0(format(round(ATT, 4), nsmall = 4), Sig),

```

```

1955     SE_fmt = format(round(SE, 4), nsmall = 4),
1956     p_fmt = format(round(p_valor, 4), nsmall = 4),
1957     IC_fmt = sprintf("[% .4f; % .4f]", CI_inf, CI_sup)) %>%
1958     select(Cohort, ATT_fmt, SE_fmt, p_fmt, IC_fmt),
1959     tibble(
1960       Cohort = "\\textbf{Overall}",
1961       ATT_fmt = paste0("\\textbf{", format(round(att_s$ATT, 4), nsmall = 4), att_s$Sig,
1962 ↪     "}") ,
1963       SE_fmt = paste0("\\textbf{", format(round(att_s$SE, 4), nsmall = 4), "}") ,
1964       p_fmt = paste0("\\textbf{", format(round(att_s$p_valor, 4), nsmall = 4), "}") ,
1965       IC_fmt = paste0("\\textbf{[", sprintf("% .4f; % .4f", att_s$CI_inf, att_s$CI_sup),
1966 ↪     "]}")
1967     )
1968   )
1969   latex_out <- tbl_latex %>%
1970     kbl(format = "latex", booktabs = TRUE,
1971         col.names = c("Cohort", "ATT", "Erro Padrão", "Valor-p", "IC 95\\%"),
1972         align = c("l", "r", "r", "r", "c"),
1973         caption = paste0("Efeito do Tratamento por Cohort ", out,
1974             ". Estimador Callaway \\& Sant'Anna (2021). Brasil, 2017--2024."),
1975         label = paste0("tab:cs_", tolower(out)),
1976         escape = FALSE, row.names = FALSE) %>%
1977     kable_styling(latex_options = "hold_position") %>%
1978     footnote(general = paste0(
1979         "Erros-padrão clusterizados por município. Método doubly-robust (DR). ",
1980         "Controles: never-treated. Significância: *** p<0.001; ** p<0.01; * p<0.05; . p<0.10."
1981     ), general_title = "", footnote_as_chunk = TRUE, escape = FALSE)
1982
1983     writeLines(latex_out, paste0("outputs/did/tabela_", out, ".tex"))
1984     cat(" Tabela salva: outputs/did/tabela_", out, ".tex\n")
1985   }
1986
1987   # Event-study
1988   att_dyn <- resultados_agr[[out]]$dynamic
1989   if (!is.null(att_dyn)) {
1990     es_df <- tibble(e = att_dyn$egt, att = att_dyn$att.egt, se = att_dyn$se.egt,
1991         ci_lower = att_dyn$att.egt - 1.96 * att_dyn$se.egt,
1992         ci_upper = att_dyn$att.egt + 1.96 * att_dyn$se.egt)
1993
1994     p_es <- ggplot(es_df, aes(x = e, y = att)) +
1995       geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "#D7191C", alpha = 0.15) +
1996       geom_line(color = "#D7191C", linewidth = 0.9) +
1997       geom_point(color = "#D7191C", size = 2.5) +
1998       geom_hline(yintercept = 0, linetype = "dashed", color = "gray40") +
1999       geom_vline(xintercept = -0.5, linetype = "dotted", color = "gray60") +
2000       scale_x_continuous(breaks = sort(unique(es_df$e))) +
2001       labs(title = paste0("Event-Study ", labels_vars[out]),
2002           subtitle = "Callaway & Sant'Anna (2021) | Doubly-Robust | Never-Treated como
2003 ↪     controle",
2004           x = "Períodos relativos ao tratamento", y = "ATT estimado",
2005           caption = "IC 95%. Erros-padrão clusterizados por município.") +
2006       theme_minimal(base_size = 12) +
2007       theme(plot.title = element_text(face = "bold", hjust = 0.5),
2008           plot.subtitle = element_text(hjust = 0.5, color = "gray40", size = 9),
2009           panel.grid.minor = element_blank())
2010
2011     ggsave(paste0("outputs/did/plots/event_study_", out, ".png"),
2012         plot = p_es, width = 10, height = 6, dpi = 300, bg = "white")
2013     cat(" Event-study salvo\n")
2014   }
2015   cat("\n")
2016
2017   # Média descritiva pré-tratamento dos tratados (âncora interpretativa)
2018   # Período pré: 2017-2019 (antes da primeira coorte possível = 2020)

```

```

2019 # Usado apenas como referência de contexto no texto, não como denominador formal
2020
2021 media_pre_tratados <- painel_did %>%
2022   filter(
2023     first_treat > 0,      # apenas tratados
2024     Ano <= 2019,         # período pré universal
2025     !is.na(Ano)
2026   ) %>%
2027   summarise(across(
2028     all_of(outcomes_validos_did),
2029     ~mean(., na.rm = TRUE)
2030   )) %>%
2031   pivot_longer(everything(),
2032     names_to = "outcome",
2033     values_to = "media_pre_2017_2019")
2034
2035 write_csv(media_pre_tratados,
2036   "outputs/did/media_pre_tratados_2017_2019.csv")
2037
2038 cat("Médias pré-tratamento (2017-2019) grupo tratado:\n")
2039 media_pre_tratados %>%
2040   filter(outcome %in% vars_principais) %>%
2041   mutate(media_pre_2017_2019 = round(media_pre_2017_2019, 3)) %>%
2042   print()
2043 cat("\n")
2044
2045 # =====
2046 # 21. TABELAS AGREGADAS POR GRUPO (IAG, IFA) ----
2047 # =====
2048
2049 cat(rep("=", 80), "\n", sep = "")
2050 cat("PARTE 21: TABELAS AGREGADAS POR GRUPO\n")
2051 cat(rep("=", 80), "\n\n", sep = "")
2052
2053 tabela_geral <- map_dfr(names(resultados Agr), function(out) {
2054   s <- resultados_Agr[[out]]$simple
2055   if (is.null(s)) return(NULL)
2056   pval <- 2 * pnorm(-abs(s$overall.att / s$overall.se))
2057   tibble(outcome = out, grupo = str_extract(out, "[A-Z]{2,3}"),
2058     ATT = s$overall.att, SE = s$overall.se,
2059     CI_inf = s$overall.att - 1.96 * s$overall.se,
2060     CI_sup = s$overall.att + 1.96 * s$overall.se,
2061     p_valor = pval,
2062     Sig = case_when(pval < 0.001 ~ "***", pval < 0.01 ~ "**",
2063       pval < 0.05 ~ "*", pval < 0.10 ~ ".", TRUE ~ ""))
2064 })
2065
2066 write_csv(tabela_geral, "outputs/did/tabela_att_todos.csv")
2067
2068 for (grp in unique(tabela_geral$grupo)) {
2069   tbl_grp <- tabela_geral %>% filter(grupo == grp) %>% arrange(outcome) %>%
2070     mutate(ATT_fmt = paste0(format(round(ATT, 4), nsmall = 4), Sig),
2071       SE_fmt = format(round(SE, 4), nsmall = 4),
2072       p_fmt = format(round(p_valor, 4), nsmall = 4),
2073       IC_fmt = sprintf("[%.4f; %.4f]", CI_inf, CI_sup)) %>%
2074     select(outcome, ATT_fmt, SE_fmt, p_fmt, IC_fmt)
2075
2076   latex_grp <- tbl_grp %>%
2077     kbl(format = "latex", booktabs = TRUE,
2078       col.names = c("Indicador", "ATT", "Erro Padrão", "Valor-p", "IC 95\\%"),
2079       align = c("l", "r", "r", "r", "c"),
2080       caption = paste0("ATT Agregado Grupo ", grp,
2081         ". Estimador Callaway \\& Sant'Anna (2021). Brasil, 2017--2024."),
2082       label = paste0("tab:cs_grupo_", tolower(grp)),
2083       escape = FALSE, row.names = FALSE) %>%
2084     kable_styling(latex_options = "hold_position") %>%
2085     footnote(general = paste0(

```

```

2086     "ATT overall (agregado simples). Erros-padrão clusterizados por município. ",
2087     "Método doubly-robust (DR). Controles: never-treated. ",
2088     "Significância: *** p<0.001; ** p<0.01; * p<0.05; . p<0.10."
2089   ), general_title = "", footnote_as_chunk = TRUE, escape = FALSE)
2090
2091   writeLines(latex_grp, paste0("outputs/did/tabela_grupo_", grp, ".tex"))
2092   cat(" Salvo: tabela_grupo_", grp, ".tex\n")
2093 }
2094 cat("\n")
2095
2096
2097 # =====
2098 # 22. ROBUSTEZ DID RATIO 1, 2 E 5 ----
2099 # =====
2100
2101 cat(rep("=", 80), "\n", sep = "")
2102 cat("PARTE 22: ROBUSTEZ DID RATIOS 1, 2 E 5\n")
2103 cat(rep("=", 80), "\n\n", sep = "")
2104
2105 extrair_att_overall <- function(cs_obj) {
2106   if (is.null(cs_obj)) return(NULL)
2107   s <- tryCatch(aggte(cs_obj, type = "simple"), error = function(e) NULL)
2108   if (is.null(s)) return(NULL)
2109   pval <- 2 * pnorm(-abs(s$overall.att / s$overall.se))
2110   tibble(ATT = s$overall.att, SE = s$overall.se, p_valor = pval,
2111          Sig = case_when(pval < 0.001 ~ "***", pval < 0.01 ~ "**",
2112                         pval < 0.05 ~ "*", pval < 0.10 ~ ".", TRUE ~ ""))
2113 }
2114
2115 resultados_robustez <- list()
2116
2117 for (r in ratios_psm) {
2118   arquivo_painel <- sprintf("data/painel_did_ratio%d.csv", r)
2119   if (!file.exists(arquivo_painel)) { cat(" Painel ratio", r, "não encontrado\n"); next }
2120
2121   painel_r <- read_csv(arquivo_painel, show_col_types = FALSE) %>%
2122     mutate(CodigoMunicipio = as.character(CodigoMunicipio))
2123
2124   outcomes_r <- names(painel_r)[str_detect(names(painel_r), "^I[A-Z0-9]")]
2125   outcomes_validos_r <- painel_r %>%
2126     select(all_of(outcomes_r)) %>%
2127     summarise(across(everything(), ~mean(is.na(.)) * 100)) %>%
2128     pivot_longer(everything(), names_to = "outcome", values_to = "pct_missing") %>%
2129     filter(pct_missing < 70) %>% pull(outcome)
2130
2131   cat(sprintf(" Ratio 1:%d %d outcomes válidos\n", r, length(outcomes_validos_r)))
2132
2133   atts_ratio <- list()
2134   for (out in outcomes_validos_r) {
2135     cs_r <- estimar_cs(painel_r, out)
2136     att_r <- extrair_att_overall(cs_r)
2137     if (!is.null(att_r)) atts_ratio[[out]] <- att_r %>% mutate(outcome = out, ratio = r)
2138   }
2139   resultados_robustez[[paste0("ratio_", r)]] <- bind_rows(atts_ratio)
2140   cat(sprintf(" Estimações concluídas: %d\n", length(atts_ratio)))
2141 }
2142
2143 tabela_rob_raw <- bind_rows(resultados_robustez)
2144 write_csv(tabela_rob_raw, "outputs/did/robustez_att_todos.csv")
2145
2146 # Tabela wide
2147 tab_rob_wide <- tabela_rob_raw %>%
2148   mutate(ATT_fmt = paste0(format(round(ATT, 4), nsmall = 4), Sig)) %>%
2149   select(outcome, ratio, ATT_fmt) %>%
2150   pivot_wider(names_from = ratio, values_from = ATT_fmt, names_prefix = "ratio_") %>%
2151   mutate(grupo = str_extract(outcome, "[A-Z]{2,3}") %>%
2152     arrange(grupo, outcome) %>%

```

```

2153   mutate(across(starts_with("ratio_"), ~replace_na(.x, "")))
2154
2155   colunas_ratio <- intersect(paste0("ratio_", ratios_psm), names(tab_rob_wide))
2156
2157   grupos_rob <- tab_rob_wide %>% group_by(grupo) %>% summarise(n = n(), .groups = "drop")
2158
2159   latex_rob <- tab_rob_wide %>%
2160     select(-grupo) %>%
2161     kbl(format = "latex", booktabs = TRUE,
2162         col.names = c("Indicador", paste0("ATT (1:", ratios_psm[seq_along(colunas_ratio)],
2163         ↪ ")")),
2164         align = c("l", rep("r", length(colunas_ratio))),
2165         caption = paste0(
2166           "Análise de Robustez ATT Overall por Tamanho de Amostra PSM. ",
2167           "Estimador Callaway \\& Sant'Anna (2021). ",
2168           "Amostra PSM-FLEXÍVEL com Ratios 1:1, 1:2 e 1:5. Brasil, 2017--2024."
2169         ),
2170         label = "tab:robustez_ratio", escape = FALSE, row.names = FALSE) %>%
2171     kable_styling(latex_options = c("hold_position", "scale_down"))
2172
2173   # pack_rows por grupo
2174   inicio <- 1
2175   for (i in seq_len(nrow(grupos_rob))) {
2176     latex_rob <- latex_rob %>%
2177       pack_rows(grupos_rob$grupo[i], inicio, inicio + grupos_rob$n[i] - 1,
2178         bold = TRUE, latex_gap_space = "0.3em")
2179     inicio <- inicio + grupos_rob$n[i]
2180   }
2181
2182   latex_rob <- latex_rob %>%
2183     footnote(general = paste0(
2184       "ATT overall (agregado simples). Erros-padrão clusterizados por município. ",
2185       "Método doubly-robust (DR). Controles: never-treated. ",
2186       "Significância: *** p<0.001; ** p<0.01; * p<0.05; . p<0.10. ",
2187       "indica que a estimação não convergiu."
2188     ), general_title = "", footnote_as_chunk = TRUE, escape = FALSE)
2189
2190   writeLines(latex_rob, "outputs/did/tabela_robustez_ratio.tex")
2191   cat(" Tabela salva: outputs/did/tabela_robustez_ratio.tex\n\n")
2192
2193   # =====
2194   # 23. TABELAS LATEX GERAIS ----
2195   # =====
2196
2197   cat(rep("=", 80), "\n", sep = "")
2198   cat("PARTE 23: TABELAS LATEX GERAIS\n")
2199   cat(rep("=", 80), "\n\n", sep = "")
2200
2201   # Tabela de equivalências SINISA-SNIS
2202   latex_eq <- equivalencias_agua %>%
2203     select(`Código SINISA` = codigo_sinisa, `Código SNIS` = codigo_snis,
2204       `Nome do Indicador` = nome_indicador, Grupo = grupo, Unidade = unidade) %>%
2205     mutate(`Nome do Indicador` = str_trunc(`Nome do Indicador`, 45)) %>%
2206     kbl(format = "latex", booktabs = TRUE, longtable = TRUE,
2207         caption = "Harmonização dos Indicadores de Saneamento: Equivalências entre SINISA
2208         ↪ (2023-2024) e SNIS (2017-2022). Brasil, 2026.",
2209         label = "tab:equivalencias") %>%
2210     kable_styling(font_size = 10, latex_options = "repeat_header")
2211   writeLines(latex_eq, "outputs/tabela_equivalencias_sinisa_snis.tex")
2212   cat(" outputs/tabela_equivalencias_sinisa_snis.tex\n")
2213
2214   # Tabela de validação (se disponível)
2215   if (!is.null(comparacao)) {
2216     latex_val <- comparacao %>%
2217       select(`Código` = Indicador, `N pares` = N_pares,
2218         `Correlação` = Correlacao, `RMSE (%)` = RMSE_rel,

```

```

2218     `Bias (%)` = Bias_rel, `p-valor` = P_valor) %>%
2219     mutate(`Válido?` = if_else(comparacao$Valido, "Sim", "Não")) %>%
2220     kbl(format = "latex", booktabs = TRUE, longtable = TRUE,
2221         caption = "Validação da Metodologia: Correlação e RMSE entre Indicadores Calculados e
↪ SNIS Original. Brasil, 2026.",
2222         label = "tab:validacao") %>%
2223     kable_styling(font_size = 10, latex_options = "repeat_header")
2224     writeLines(latex_val, "outputs/tabela_validacao.tex")
2225     cat("  outputs/tabela_validacao.tex\n")
2226   }
2227
2228   # Tabela de estatísticas descritivas do painel final
2229   dados_finais <- painel_completo
2230   indicadores_col <- names(dados_finais)[str_detect(names(dados_finais), "^I[A-Z]{2}\\d")]
2231
2232   if (length(indicadores_col) > 0) {
2233     mapa_nomes <- equivalencias_agua %>%
2234       select(codigo = codigo_sinisa, nome = nome_indicador) %>%
2235       mutate(nome = str_trunc(nome, 35))
2236
2237     estat_painel <- dados_finais %>%
2238       select(all_of(indicadores_col)) %>%
2239       pivot_longer(everything(), names_to = "Código", values_to = "Valor") %>%
2240       group_by(Código) %>%
2241       summarise(N = sum(!is.na(Valor)), Média = mean(Valor, na.rm = TRUE),
2242                 `Desvio Padrão` = sd(Valor, na.rm = TRUE),
2243                 Mínimo = min(Valor, na.rm = TRUE), Máximo = max(Valor, na.rm = TRUE),
2244                 .groups = "drop") %>%
2245       left_join(mapa_nomes, by = c("Código" = "codigo")) %>%
2246       select(Código, Nome = nome, everything()) %>%
2247       mutate(across(where(is.numeric) & !N, ~round(., 2)))
2248
2249     latex_est <- estat_painel %>%
2250     kbl(format = "latex", booktabs = TRUE, longtable = TRUE,
2251         caption = "Estatísticas Descritivas do Painel Municipal de Saneamento (2017-2024).
↪ Brasil, 2026.",
2252         label = "tab:descritivas") %>%
2253     kable_styling(font_size = 10, latex_options = "repeat_header") %>%
2254     footnote(general = c(
2255         "Fonte: Elaboração própria com base em SNIS (2017-2022) e SINISA (2023-2024).",
2256         paste0("Painel: ", format(nrow(dados_finais), big.mark = "."), " obs.", " ",
2257             format(n_distinct(dados_finais$CodigoMunicipio), big.mark = "."),
2258             " municípios, 2017-2024.")
2259     ), threeparttable = TRUE)
2260
2261     writeLines(latex_est, "outputs/tabela_estatisticas_painel_final.tex")
2262     cat("  outputs/tabela_estatisticas_painel_final.tex\n\n")
2263   }
2264
2265
2266   # =====
2267   # RESUMO FINAL ----
2268   # =====
2269
2270   cat("\n", rep("=", 80), "\n", sep = "")
2271   cat("RESUMO FINAL\n")
2272   cat(rep("=", 80), "\n\n", sep = "")
2273
2274   cat("DADOS:\n")
2275   cat("  Municípios no painel 2017-2024:", n_distinct(painel_completo$CodigoMunicipio), "\n")
2276   cat("  Observações:", nrow(painel_completo), "\n")
2277   cat("  Anos:", paste(sort(unique(painel_completo$Ano)), collapse = ", "), "\n\n")
2278
2279   cat("PSM (Estratégia Flexível,", ratio_escolhido, "):\n")
2280   cat("  Tratados:", sum(dados_pareados$private == 1), "\n")
2281   cat("  Controles:", sum(dados_pareados$private == 0), "\n")
2282   cat("  SMD médio:", round(pareamentos[[ratio_escolhido]]$smd_medio, 4), "\n\n")

```

```

2283
2284 cat("DiD (Callaway & Sant'Anna):\n")
2285 cat("  Outcomes estimados:", length(resultados_cs), "\n\n")
2286
2287 cat("ARQUIVOS GERADOS:\n")
2288 arquivos <- c(
2289   "data/painel_completo_2017_2024.csv",
2290   "data/amostra_pareada_psm_ratio1.csv",
2291   "data/amostra_pareada_psm_ratio2.csv",
2292   "data/amostra_pareada_psm_ratio5.csv",
2293   "data/painel_did_ratio1.csv",
2294   "data/painel_did_ratio2.csv",
2295   "data/painel_did_ratio5.csv",
2296   "data/resultados_cs_att_gt.rds",
2297   "outputs/equivalencias_sinisa_snis_2024.csv",
2298   "outputs/municipios_multiplos_prestadores.csv",
2299   "outputs/tabela_comparacao_psm.tex",
2300   "outputs/tabela_balance_psm.tex",
2301   "outputs/mapa_amostra_pareada.png",
2302   "outputs/tabela_equivalencias_sinisa_snis.tex",
2303   "outputs/tabela_estatisticas_painel_final.tex",
2304   "outputs/did/tabela_descritivas.tex",
2305   "outputs/did/tabela_att_todos.csv",
2306   "outputs/did/tabela_IAG0001.tex",
2307   "outputs/did/tabela_IFA1001.tex",
2308   "outputs/did/tabela_IFA2008.tex",
2309   "outputs/did/tabela_grupo_IAG.tex",
2310   "outputs/did/tabela_grupo_IFA.tex",
2311   "outputs/did/tabela_robustez_ratio.tex",
2312   "outputs/did/robustez_att_todos.csv",
2313   "outputs/did/plots/tendencia_IAG0001.png",
2314   "outputs/did/plots/tendencia_IFA1001.png",
2315   "outputs/did/plots/tendencia_IFA2008.png",
2316   "outputs/did/plots/event_study_IAG0001.png",
2317   "outputs/did/plots/event_study_IFA1001.png",
2318   "outputs/did/plots/event_study_IFA2008.png"
2319 )
2320 cat(paste0("  ", arquivos, collapse = "\n"), "\n")
2321
2322 cat("\n", rep("=", 80), "\n", sep = "")
2323 cat("PIPELINE CONCLUÍDO!\n")
2324 cat(rep("=", 80), "\n\n", sep = "")
2325
2326 # =====
2327 # 24. ANLISE DE MORBIDADE HOSPITALAR ----
2328 # =====
2329
2330 cat(rep("=", 80), "\n", sep = "")
2331 cat("PARTE 22B: ANLISE DE MORBIDADE HOSPITALAR\n")
2332 cat(rep("=", 80), "\n\n", sep = "")
2333
2334 # Arquivos esperados em data/:
2335 # morbidademenorque1.csv   internações por doenças de saneamento, < 1 ano
2336 # morbidadela4.csv        idem, 1-4 anos
2337 # morbidadetotal.csv      idem, todas as idades
2338 # Formato: código IBGE (6 dígitos) + colunas por ano (2017, 2018, ..., 2024)
2339 # Anos 2020-2022 excluídos por distorção da pandemia COVID-19
2340
2341 dir.create("outputs/did", showWarnings = FALSE, recursive = TRUE)
2342
2343 arquivos_morb <- c(
2344   morbidademenorque1 = "data/morbidademenorque1.csv",
2345   morbidadela4       = "data/morbidadela4.csv",
2346   morbidadetotal     = "data/morbidadetotal.csv"
2347 )
2348 arquivos_existentes <- arquivos_morb[file.exists(arquivos_morb)]
2349

```

```

2350 if (length(arquivos_existentes) == 0) {
2351
2352   cat("AVISO: Nenhum arquivo de morbidade encontrado em data/\n")
2353   cat("  Esperados: morbidademenorquel.csv, morbidadela4.csv, morbidadetotal.csv\n\n")
2354   ANALISE_MORBIDADE <- FALSE
2355
2356 } else {
2357
2358   cat("Arquivos encontrados:", length(arquivos_existentes), "\n\n")
2359
2360   # Função: carregar e pivotar arquivo de morbidade
2361   carregar_morbidade <- function(arquivo, nome_var) {
2362     dados <- tryCatch(
2363       read_delim(arquivo, delim = ";",
2364                 locale = locale(encoding = "WINDOWS-1252"),
2365                 show_col_types = FALSE),
2366       error = function(e)
2367         read_delim(arquivo, delim = ";",
2368                   locale = locale(encoding = "UTF-8"),
2369                   show_col_types = FALSE)
2370     )
2371     cols_ano <- names(dados)[grepl("^20\\d{2}$", names(dados))]
2372     if (length(cols_ano) == 0) { cat("  AVISO: sem colunas de ano em", arquivo, "\n");
2373       ↪ return(NULL) }
2374
2375     dados %>%
2376       pivot_longer(all_of(cols_ano), names_to = "ano", values_to = nome_var) %>%
2377       mutate(
2378         ano = as.integer(ano),
2379         !!nome_var := suppressWarnings(as.numeric(as.character(.data[[nome_var]])))
2380       ) %>%
2381       rename_with(~"codigo_municipio",
2382                 any_of(c("id_municipio_6", "codigo_ibge", "cod_ibge", "CodMunicipio"))) %>%
2383       mutate(codigo_municipio = as.integer(codigo_municipio))
2384     }
2385
2386     lista_morb <- imap(arquivos_existentes, carregar_morbidade)
2387     lista_morb <- compact(lista_morb)
2388
2389     morbidade_completo <- reduce(lista_morb, full_join,
2390                                by = c("codigo_municipio", "ano")) %>%
2391     filter(!is.na(codigo_municipio))
2392
2393     cat("Dados de morbidade combinados:\n")
2394     cat("  Obs.:", nrow(morbidade_completo),
2395         "| Municípios:", n_distinct(morbidade_completo$codigo_municipio),
2396         "| Anos:", paste(sort(unique(morbidade_completo$ano)), collapse = ", "), "\n\n")
2397
2398     vars_morbidade <- intersect(names(arquivos_existentes), names(morbidade_completo))
2399
2400     # Função: preparar painel de morbidade para DID
2401     # Join por código de 6 dígitos; exclui anos de pandemia (2020-2022)
2402     preparar_morb_did <- function(dados_morb, amostra_psm) {
2403
2404       # Padronizar para 6 dígitos em ambos os lados
2405       cod6_psm <- amostra_psm %>%
2406         mutate(cod6 = as.integer(CodigoMunicipio)) %>%
2407         select(cod6, first_treat, private) %>%
2408         distinct(cod6, .keep_all = TRUE)
2409
2410       dados_morb %>%
2411         filter(ano %in% c(2017:2019, 2023:2024)) %>% # excluir pandemia
2412         filter(codigo_municipio %in% cod6_psm$cod6) %>%
2413         left_join(cod6_psm, by = c("codigo_municipio" = "cod6")) %>%
2414         filter(!is.na(private)) %>%
2415         mutate(
2416           first_treat = as.integer(coalesce(first_treat, 0L)),

```

```

2416     id_num      = as.integer(as.factor(codigo_municipio))
2417   )
2418 }
2419
2420 # Função: estimar C&S para variável de morbidade
2421 estimar_cs_morb <- function(dados, yvar) {
2422   dados_ok <- dados %>% filter(!is.na(.data[[yvar]]))
2423   n_trat    <- n_distinct(dados_ok$codigo_municipio[dados_ok$private == 1])
2424   n_cont    <- n_distinct(dados_ok$codigo_municipio[dados_ok$private == 0])
2425   if (n_trat < 2 || n_cont < 2) return(NULL)
2426   tryCatch(
2427     att_gt(yname = yvar, tname = "ano", idname = "id_num",
2428           gname = "first_treat", data = dados_ok,
2429           control_group = "nevertreated", est_method = "dr",
2430           clustervars = "id_num", anticipation = 0,
2431           allow_unbalanced_panel = TRUE, print_details = FALSE),
2432     error = function(e) NULL
2433   )
2434 }
2435
2436 extrair_att_morb <- function(cs_obj) {
2437   if (is.null(cs_obj)) return(NULL)
2438   s <- tryCatch(aggtc(cs_obj, type = "simple"), error = function(e) NULL)
2439   g <- tryCatch(aggtc(cs_obj, type = "group"), error = function(e) NULL)
2440   if (is.null(s)) return(NULL)
2441   pval <- 2 * pnorm(-abs(s$overall.att / s$overall.se))
2442   list(
2443     ATT      = s$overall.att,
2444     SE       = s$overall.se,
2445     IC_inf   = s$overall.att - 1.96 * s$overall.se,
2446     IC_sup   = s$overall.att + 1.96 * s$overall.se,
2447     p_valor  = pval,
2448     Sig      = case_when(pval < 0.001 ~ "***", pval < 0.01 ~ "**",
2449                         pval < 0.05 ~ "*", pval < 0.10 ~ ".", TRUE ~ ""),
2450     grupo_obj = g
2451   )
2452 }
2453
2454 label_etario <- c(
2455   morbidademenorque1 = "< 1 ano",
2456   morbidadela4       = "1--4 anos",
2457   morbidadetotal     = "Total"
2458 )
2459
2460 # Estimar para cada ratio e variável
2461 resultados_morb <- list() # [ratio][var]
2462
2463 for (r in ratios_psm) {
2464   chave <- paste0("ratio", r)
2465   if (!chave %in% names(pareamentos)) next
2466   amostra_r <- pareamentos[[chave]]$pareados
2467   morb_r    <- preparar_morb_did(morbidade_completo, amostra_r)
2468   cat(sprintf(" Ratio 1:%d %d obs, %d municípios\n",
2469             r, nrow(morb_r), n_distinct(morb_r$codigo_municipio)))
2470
2471   for (var in vars_morbidade) {
2472     cat(sprintf(" %s (%s) ...", var, label_etario[var]))
2473     cs <- estimar_cs_morb(morb_r, var)
2474     att <- extrair_att_morb(cs)
2475     if (!is.null(att)) {
2476       resultados_morb[[chave]][[var]] <- att
2477       cat(sprintf(" ATT=%.4f %s\n", att$ATT, att$Sig))
2478     } else {
2479       cat(" sem resultado\n")
2480     }
2481   }
2482 }

```

```

2483
2484 saveRDS(resultados_morb, "data/resultados_did_morbidade.rds")
2485
2486 # Tabela principal: grupo etário coorte (amostra escolhida)
2487 cat("\nGerando tabela de morbidade por grupo etário...\n")
2488
2489 res_principal <- resultados_morb[[ratio_escolhido]]
2490
2491 tabela_morb <- map_dfr(vars_morbidade, function(var) {
2492   att <- res_principal[[var]]
2493   if (is.null(att)) return(NULL)
2494
2495   linha <- tibble(
2496     `Grupo etário` = label_etario[var],
2497     ATT             = round(att$ATT, 4),
2498     SE              = round(att$SE, 4),
2499     `IC 95%\`      = sprintf("[% .4f; % .4f]", att$IC_inf, att$IC_sup),
2500     `Valor-p`      = round(att$p_valor, 4),
2501     Sig             = att$Sig
2502   )
2503
2504   # Adicionar ATT por coorte
2505   if (!is.null(att$grupo_obj)) {
2506     g <- att$grupo_obj
2507     for (i in seq_along(g$egt)) {
2508       coorte <- g$egt[i]
2509       linha[[paste0("ATT_", coorte)]] <- round(g$att.egt[i], 4)
2510       linha[[paste0("SE_", coorte)]] <- round(g$se.egt[i], 4)
2511       pval_c <- 2 * pnorm(-abs(g$att.egt[i] / g$se.egt[i]))
2512       linha[[paste0("Sig_", coorte)]] <- case_when(
2513         pval_c < 0.001 ~ "***", pval_c < 0.01 ~ "**",
2514         pval_c < 0.05 ~ "*", pval_c < 0.10 ~ ".", TRUE ~ "")
2515     }
2516   }
2517   linha
2518 })
2519
2520 write_csv(tabela_morb, "outputs/did/tabela_morbidade.csv")
2521
2522 # LaTeX tabela por grupo etário
2523 latex_morb <- tabela_morb %>%
2524   mutate(ATT_fmt = paste0(format(ATT, nsmall = 4), Sig)) %>%
2525   select(`Grupo etário`, ATT_fmt, SE, `IC 95%\`, `Valor-p`) %>%
2526   kbl(format = "latex", booktabs = TRUE,
2527     col.names = c("Grupo etário", "ATT", "Erro Padrão", "IC 95%\`, "Valor-p"),
2528     align = c("l", "r", "r", "c", "r"),
2529     caption = paste0("Efeito da Concessão Plena sobre Morbidade Hospitalar por ",
2530       "Doenças Relacionadas ao Saneamento. ",
2531       "Estimador Callaway \& Sant'Anna (2021). Brasil, 2017--2024."),
2532     label = "tab:morbidade", escape = FALSE, row.names = FALSE) %>%
2533   kable_styling(latex_options = "hold_position") %>%
2534   footnote(general = paste0(
2535     "ATT overall (agregado simples). Erros-padrão clusterizados por município. ",
2536     "Método doubly-robust (DR). Controles: never-treated. ",
2537     "Anos 2020--2022 excluídos (pandemia COVID-19). ",
2538     "Significância: *** p<0.001; ** p<0.01; * p<0.05; . p<0.10."
2539   ), general_title = "", footnote_as_chunk = TRUE, escape = FALSE)
2540
2541 writeLines(latex_morb, "outputs/did/tabela_morbidade.tex")
2542 cat(" Tabela salva: outputs/did/tabela_morbidade.tex\n\n")
2543
2544 # Tabela de robustez: grupo etário ratio
2545 cat("Gerando tabela de robustez de morbidade...\n")
2546
2547 tabela_morb_rob <- map_dfr(vars_morbidade, function(var) {
2548   map_dfr(ratios_psm, function(r) {
2549     chave <- paste0("ratio", r)

```

```

2550   att   <- resultados_morb[[chave]][[var]]
2551   if (is.null(att)) return(tibble(var = var, ratio = r,
2552                                   ATT_fmt = "", Nobs = NA_integer_))
2553   tibble(
2554     var      = var,
2555     ratio    = r,
2556     ATT_fmt = paste0(format(round(att$ATT, 4), nsmall = 4), att$Sig)
2557   )
2558 })
2559 }) %>%
2560   mutate(`Grupo etário` = label_etario[var]) %>%
2561   pivot_wider(names_from = ratio, values_from = ATT_fmt,
2562               names_prefix = "ratio_", values_fill = "") %>%
2563   select(`Grupo etário`, starts_with("ratio_"))
2564
2565 latex_morb_rob <- tabela_morb_rob %>%
2566   kbl(format = "latex", booktabs = TRUE,
2567        col.names = c("Grupo etário",
2568                      paste0("ATT (1:", ratios_psm, ")")),
2569        align = c("l", rep("r", length(ratios_psm))),
2570        caption = paste0("Análise de Robustez Efeito sobre Morbidade Hospitalar ",
2571                          "por Tamanho de Amostra PSM. ",
2572                          "Callaway \\& Sant'Anna (2021). Brasil, 2017--2024."),
2573        label = "tab:morbidade_robustez", escape = FALSE, row.names = FALSE) %>%
2574   kable_styling(latex_options = "hold_position") %>%
2575   footnote(general = paste0(
2576     "ATT overall (agregado simples). Erros-padrão clusterizados por município. ",
2577     "Método doubly-robust (DR). Controles: never-treated. ",
2578     "Anos 2020--2022 excluídos (pandemia). ",
2579     "Significância: *** p<0.001; ** p<0.01; * p<0.05; . p<0.10. ",
2580     " indica estimação não convergida."
2581   ), general_title = "", footnote_as_chunk = TRUE, escape = FALSE)
2582
2583 writeLines(latex_morb_rob, "outputs/did/tabela_morbidade_robustez.tex")
2584 cat(" Tabela salva: outputs/did/tabela_morbidade_robustez.tex\n\n")
2585
2586 # Event-study plots por grupo etário
2587 cat("Gerando event-study plots de morbilidade...\n")
2588
2589 for (var in vars_morbidade) {
2590   amostra_r   <- pareamentos[[ratio_escolhido]]$pareados
2591   morb_princ  <- preparar_morb_did(morbidade_completo, amostra_r)
2592   cs_es       <- estimar_cs_morb(morb_princ, var)
2593   if (is.null(cs_es)) next
2594
2595   att_dyn <- tryCatch(aggte(cs_es, type = "dynamic"), error = function(e) NULL)
2596   if (is.null(att_dyn)) next
2597
2598   es_df <- tibble(
2599     e      = att_dyn$egt,
2600     att    = att_dyn$att.egt,
2601     se     = att_dyn$sse.egt,
2602     ci_lower = att_dyn$att.egt - 1.96 * att_dyn$sse.egt,
2603     ci_upper = att_dyn$att.egt + 1.96 * att_dyn$sse.egt
2604   )
2605
2606   p_es <- ggplot(es_df, aes(x = e, y = att)) +
2607     geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), fill = "#D7191C", alpha = 0.15) +
2608     geom_line(color = "#D7191C", linewidth = 0.9) +
2609     geom_point(color = "#D7191C", size = 2.5) +
2610     geom_hline(yintercept = 0, linetype = "dashed", color = "gray40") +
2611     geom_vline(xintercept = -0.5, linetype = "dotted", color = "gray60") +
2612     scale_x_continuous(breaks = sort(unique(es_df$e))) +
2613     labs(
2614       title = paste0("Event-Study Morbidade Hospitalar (", label_etario[var], ")"),
2615       subtitle = "Callaway & Sant'Anna (2021) | Doubly-Robust | Never-Treated",
2616       x = "Períodos relativos ao tratamento", y = "ATT estimado",

```

```
2617     caption = "IC 95%. Erros-padrão clusterizados. Anos 2020-2022 excluídos (pandemia)."
2618   ) +
2619   theme_minimal(base_size = 12) +
2620   theme(plot.title = element_text(face = "bold", hjust = 0.5),
2621         plot.subtitle = element_text(hjust = 0.5, color = "gray40", size = 9),
2622         panel.grid.minor = element_blank())
2623
2624   ggsave(paste0("outputs/did/plots/event_study_morb_", var, ".png"),
2625         plot = p_es, width = 10, height = 6, dpi = 300, bg = "white")
2626   cat("  Event-study salvo:", var, "\n")
2627 }
2628
2629 ANALISE_MORBIDADE <- TRUE
2630 cat("\n")
2631 }
```