

Uma Análise Espacial da Relação Entre Saneamento Básico e Educação nos Municípios Brasileiros

Luiza Lins Bellon¹  | Ana Carolina da Cruz Lima² 

¹ Bacharel em Ciências Econômicas. E-mail: luiza.bellon@gmail.com

² Universidade Federal do Rio de Janeiro (UFRJ). Professora adjunta do Instituto de Economia (IE/UFRJ). E-mail: ana.lima@ie.ufrj.br

RESUMO

Este trabalho utiliza dados de 2017 do Inep, SNIS, PNSB, IBGE e RAIS para investigar a associação entre a cobertura de esgoto municipal e indicadores educacionais – a taxa de distorção idade-série e a nota padronizada do Saeb – sob uma perspectiva espacial. A análise exploratória de dados espaciais identificou dependência espacial positiva para as variáveis de interesse e para os resíduos das regressões por MQO. Os critérios de informação e razão de verossimilhança indicaram o Modelo do Erro Espacial de Durbin (SDEM) para os modelos econométricos. No caso do Saeb, não foi identificada significância da variável contínua local ou defasada da cobertura de esgoto, embora a presença de esgotamento (mesmo que insuficiente) esteja associada a um aumento do desempenho no Saeb. Nas cidades onde há infraestrutura, a ampliação do serviço está relacionada a uma redução de 0.017% da distorção idade-série, e a diminuição alcança 0.043% quando se expande a cobertura de municípios vizinhos.

PALAVRAS-CHAVE

Saneamento Básico, Educação, Análise Espacial

A Spatial Analysis of the Relationship Between Basic Sanitation and Education in Brazilian Municipalities

ABSTRACT

This paper uses 2017 data from INEP, SNIS, PNSB, IBGE and RAIS to investigate the association between municipal sewage coverage and educational indicators in the Brazilian context – the age-grade distortion and the standardized SAEB score – from a spatial perspective. The exploratory spatial data analysis identified positive spatial dependence for the variables of interest in this paper and for the residuals of the OLS models. The information criterion and likelihood ratio pointed the selection of the Spatial Durbin Error Model (SDEM) for both regressions. In the SAEB model, no significance was found for the local or lagged coverage rate, although the presence of sewage infrastructure (either high or low) is associated with an increase in the SAEB's score. In cities where the infrastructure exists, the expansion of sewage coverage relates to a reduction of 0.017% of age-grade distortion. This decrease reaches 0.043% when the sewage coverage of neighboring cities is expanded.

KEYWORDS

Sanitation, Education, Spatial Analysis

CLASSIFICAÇÃO JEL

R10, C21, I15, I25

1. Introdução

No Brasil, o direito ao saneamento básico é assegurado pela Constituição Federal¹ e pela Lei nº 11.445/2007², que estabelece as diretrizes nacionais do setor. Contudo, segundo dados do Sistema Nacional de Informações de Saneamento (SNIS), cerca de 40 milhões de brasileiros não têm ligação à rede de abastecimento de água, e 94,1 milhões estão marginalizados da infraestrutura de esgotamento sanitário (SNIS, 2020).

A insuficiência da cobertura leva a problemas de primeira ordem sobre o bem-estar das famílias e a produtividade dos trabalhadores. O contato com água contaminada por bactérias, parasitas e insetos vetores, assim como a exposição ao esgoto a céu aberto, estão associados a doenças cujos efeitos imediatos incluem febre, hemorragias e anemia (Hutton, 2013), impondo a interrupção de atividades de estudo e trabalho e um aumento da morbidade hospitalar.

As crianças configuram um dos recortes da população mais afetados pela ausência de saneamento, dado que são mais suscetíveis à contaminação por doenças infecciosas (Ngure et al., 2014) e podem ter seu desenvolvimento prejudicado por períodos de baixa absorção de nutrientes, com efeitos de longo prazo sobre habilidades sociais e cognitivas. Ademais, a contaminação por doenças gastrointestinais pode reduzir o aproveitamento escolar ao impor o absenteísmo. De fato, há evidências de que a ampliação do saneamento é um preditor para uma elevação da frequência escolar (Dreibelbis et al., 2013) e redução da taxa de distorção idade-série (Agol et al., 2018).

No caso brasileiro, há estudos que relacionam o saneamento ao aumento da incidência de doenças gastrointestinais (Ventura e Lopes, 2017; Uhr et al., 2016) e à mortalidade infantil (Barufi et al., 2012; Silva e Esperidião, 2017), mas há escassa literatura sobre a sua associação com indicadores educacionais. Scriptore (2016) identifica que a ampliação do saneamento leva a uma maior frequência escolar e redução da distorção idade-série, e Ortiz-Correa et al. (2016) demonstram que a conexão residencial duradoura com água e esgoto contribui para elevar os anos médios de estudo das crianças do domicílio.

Tendo em vista os potenciais ganhos de bem-estar associados ao saneamento, assim como a heterogeneidade na provisão do serviço, o artigo busca estudar a relação entre saneamento básico e educação nos municípios brasileiros. Particularmente, é investigada a interação entre a cobertura de esgotamento municipal e dois indicadores educacionais: a taxa de distorção idade-série e a nota padronizada do Sistema de Avaliação da Educação Básica (Saeb).

A existência da autocorrelação espacial entre os indicadores educacionais e sanitários – isto é, a correlação entre o valor de uma observação em um local e o valor

¹BRASIL. [Constituição (1988)]. Constituição da República Federativa do Brasil de 1988. Brasília, DF. Disponível em http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

²BRASIL. Lei nº 11.445, de 05 de janeiro de 2007. Disponível em http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/lei/111445.htm

de observações de espaços vizinhos – requer o uso de ferramentas econométricas que adereçam esse fenômeno de maneira apropriada (Almeida, 2012), uma vez que a independência entre variáveis compõe uma hipótese comum para a consistência e ausência de viés em procedimentos de estimação.

Portanto, o artigo analisa os dados sob o prisma da econometria espacial. Será realizada uma análise exploratória de dados espaciais (AEDE) para identificar o padrão da autocorrelação espacial nas variáveis do Saeb, distorção idade-série e cobertura de esgoto. Em seguida, serão estimados modelos econométricos que incorporam componentes de defasagem espacial na variável dependente, nas variáveis explicativas e no termo de erro. Será escolhido aquele com o melhor ajuste dos dados.

O artigo está estruturado da seguinte forma. A Seção 2 apresenta o marco teórico utilizado para analisar os mecanismos de interação entre saneamento básico e educação, bem como as principais evidências empíricas sobre o tema. A Seção 3 descreve as características da base de dados e a metodologia utilizada. A Seção 4 fornece os resultados da análise exploratória de dados espaciais e dos modelos econométricos estimados. Em seguida são expostas as considerações finais do estudo.

2. Saneamento básico e educação

Os efeitos do saneamento básico em indicadores socioeconômicos são amplamente documentados, sobretudo no âmbito da saúde pública. O consumo de água não tratada e a exposição ao esgoto a céu aberto constituem fatores de risco para uma série de enfermidades, convencionalmente chamadas Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI), tais como cólera, febre tifoide, leptospirose, teníase, dengue etc. (Hutton, 2013). Os efeitos imediatos dessas doenças incluem diarreia, febre, dores abdominais, hemorragias e anemia, podendo levar a óbito em casos mais graves.

Em estudo recente sobre o descontingenciamento de crédito para projetos de saneamento entre 2007 e 2017, Sant’Anna e Rocha (2021) apontam reduções significativas nas internações hospitalares de crianças de até 05 (cinco) anos e idosos associadas à expansão da cobertura dos serviços, resultados similares aos de Teixeira et al. (2012) e Uhr et al. (2016). A associação entre falta de saneamento e mortalidade é evidenciada, na literatura internacional, por Hutton (2013); Banerjee e Duflo (2011) e Benova et al. (2014), e, no caso do Brasil, por Barufi et al. (2012) e Silva e Esperidião (2017). Quando não causam óbito, as DRSAI podem gerar sequelas, como atrasos cognitivos, má formação fetal e susceptibilidade a outras doenças (Kothari et al., 2019), além de compelir o afastamento das atividades de trabalho e estudo, prejudicando a produtividade e o desempenho escolar (Ngure et al., 2014).

Glewwe e Miguel (2007) apresentam uma fundamentação teórica sobre a relação entre o saneamento básico e indicadores educacionais ao propor um modelo para analisar o impacto da saúde e da nutrição sobre a educação. O modelo assume dois

períodos: o primeiro engloba a faixa etária de 0 (zero) a 6 (seis) anos, quando as crianças costumam ser matriculadas na escola, e o segundo vai de 6 (seis) até 11 (onze) anos de idade, fase em que as crianças passam pela educação primária. Dados os esforços educacionais dos pais, a infraestrutura escolar, as habilidades inatas e os anos de estudo, a capacidade de aprendizado da criança é afetada por seu estado de saúde. Nesse cenário, a utilidade familiar é maximizada tendo em vista a restrição orçamentária, a função de produção das habilidades acadêmicas e duas funções de produção para a saúde da criança, que dependem das preferências dos pais. Assim, o modelo proposto permite analisar como alterações na saúde das crianças estão relacionadas a mudanças comportamentais, inclusive, sobre indicadores educacionais (Glewwe e Miguel, 2007). Os autores analisam o potencial impacto da melhoria da infraestrutura de saneamento básico e evidenciam que a provisão do serviço tende a apresentar efeitos de curto e longo prazo sobre o rendimento escolar de crianças. Apesar das dificuldades envolvidas na mensuração desses efeitos, as evidências empíricas apontam uma relação positiva entre a saúde infantil e indicadores educacionais. Neste sentido, cabe apresentar uma revisão da literatura recente sobre o impacto do ambiente sanitário – especificamente, das condições de saneamento básico – na educação.

Um dos indicadores educacionais mais utilizados para essa investigação é frequência escolar, pois a falta de saneamento pode impor o afastamento das atividades de estudo. Dreibelbis et al. (2013) apontam que a ausência do serviço em residências e escolas é um preditor significativo para o absenteísmo escolar no Quênia, e que esse resultado pode ser decorrente tanto da incidência de DRSAI quanto do fato de que estudantes tendem a evitar o ambiente escolar se a infraestrutura de saneamento não for adequada. Similarmente, a instalação de latrinas em escolas indianas foi responsável pelo aumento da taxa de matrículas e redução da evasão estudantil (Adukia, 2017).

As evidências apresentadas estão em consonância com o conhecimento acumulado sobre determinantes do desempenho escolar, tipicamente associados a condições escolares, *background* familiar e estado de saúde do estudante (Scriptore, 2016). De maneira geral, a literatura internacional tem focado nas externalidades da infraestrutura de saneamento escolar em detrimento da domiciliar, enquanto na literatura nacional o inverso ocorre.

Ortiz-Correa et al. (2016) usam dados de 2010 da Agência Nacional de Águas (ANA) e do Censo Demográfico, fornecido pelo Instituto Brasileiro de Geografia e Estatística (IBGE), para investigar como o acesso aos serviços de água e esgoto afeta o nível de escolaridade, por meio de um modelo de variáveis instrumentais com efeitos fixos de município. Os autores concluem que o total de anos de estudo cresce de acordo com o tempo de conexão do indivíduo com a rede de água e esgoto, indicando que os serviços contribuem para efeitos de bem-estar cumulativos. O investimento em saneamento é, portanto, essencial para quebrar o ciclo vicioso de pobreza: a precariedade das redes de água e esgoto se traduzem em piores indicadores educacionais e, consequente-

mente, em menor acumulação de capital humano.

Scriptore (2016) utiliza dados do Censo Demográfico e do Censo Escolar de 2010 e identifica que a elevação de 1 p.p. no acesso a serviços de saneamento está associada a um aumento de 0,11 p.p. na frequência escolar e a reduções de 0,12 e 0,31 p.p. na taxa de evasão e distorção idade-série no ensino fundamental, respectivamente. A autora insere componentes da econometria espacial na análise dos efeitos do saneamento na educação, como a realização de uma análise exploratória de dados espaciais (AEDE) e a estimação por meio de modelos espaciais³. Adicionalmente, são identificados efeitos de transbordamento espacial para a incidência de DRSAI nos municípios, o que pode ser explicado pelo deslocamento de insetos vetores de um distrito para outro, municípios que despejam esgoto em cidades vizinhas, rios e correntes marítimas que atravessam localidades e pela própria definição de doenças infectocontagiosas. Em trabalho posterior, Scriptore e Azzoni (2020) apontam transbordamentos espaciais significativos na taxa de matrícula escolar como consequência da ampliação do saneamento básico, o que também pode ser explicado pela redução da incidência de DRSAI entre crianças e adolescentes. Dessa forma, a coordenação do investimento entre regiões próximas é fundamental para incrementar a eficácia de políticas públicas em saneamento.

Nessa perspectiva, o artigo contribui para o tema de duas maneiras. Primeiramente, pela utilização de uma base de dados composta por uma combinação de informações do Sistema Nacional de Informações sobre Saneamento (SNIS) e da Pesquisa Nacional de Saneamento Básico (PNSB) para o ano de 2017, o que permitirá identificar tendências recentes e compatíveis com as mudanças socioeconômicas enfrentadas pelo país em comparação aos estudos existentes, que utilizam dados do Censo Demográfico 2010. Ademais, além da taxa de distorção idade-série – usualmente empregada na literatura internacional sobre as externalidades do saneamento –, é analisada a relação da ausência do serviço com uma medida quantitativa da avaliação do ensino: a nota padronizada do Saeb.

3. Base de dados e metodologia

O objetivo do artigo é investigar a relação entre saneamento básico e educação nos municípios brasileiros sob uma perspectiva espacial. A análise divide-se em dois objetivos específicos: (i) examinar como os indicadores de cobertura de esgoto e de qualidade do ensino – particularmente, a nota padronizada do Saeb e a taxa de distorção idade-série – variam no território, e se tal distribuição pode ser resultado da dependência espacial entre as observações; e (ii) investigar a associação entre a am-

³da Cruz et al. (2021) e Costa et al. (2020) utilizaram AEDE para investigar a distribuição espacial do saneamento básico no território e sua relação com indicadores sanitários, mas carecem estudos que explorem a associação espacial entre saneamento e indicadores educacionais. Similarmente, Barufi et al. (2012) utilizam modelos espaciais com efeitos fixos de municípios para investigar os efeitos do saneamento básico para o capital humano, mas a análise se limita ao impacto sobre a mortalidade infantil.

pliação da cobertura do saneamento (variável dependente) e os índices educacionais (variáveis de interesse). A seguir são apresentadas as principais características da base de dados e da metodologia utilizada.

3.1 Base de dados

Todos os dados utilizados são georreferenciados e possuem desagregação municipal. Os dados sobre a cobertura de esgoto foram obtidos por intermédio do Sistema Nacional de Informações sobre Saneamento (SNIS). As informações sobre indicadores educacionais provêm do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Informações da Relação Anual de Informações Sociais (RAIS) e do Instituto Brasileiro de Geografia e Estatística (IBGE) são utilizadas para construir variáveis de controle no âmbito municipal.

Embora constitua a principal fonte oficial de dados atualizados sobre saneamento, o SNIS possui uma alta quantidade de dados faltantes, visto que a pesquisa acontece em caráter voluntário para as prestadoras de serviço e carece de instrumentos de auditoria. Além disso, a forma de exibição dos dados faltantes não permite diferenciar municípios nos quais não há oferta de serviço de localidades nas quais as informações existem, mas não foram divulgadas. Para contornar esse problema, as informações do SNIS foram combinadas com a Pesquisa Nacional de Saneamento Básico (PNSB), cuja última edição ocorreu em 2017. Entre outros objetivos, a PNSB identifica quais municípios dispõem de abastecimento de água e coleta de esgoto, de modo que para as cidades em que esses serviços são ausentes a cobertura foi computada como 0% na base de dados. Já nas cidades em que a PNSB reporta haver serviços, mas que o SNIS não dispõe de valores, as observações permaneceram como dados faltantes (NA). Dessa forma, foi possível reduzir o número de NAs da variável de 3.128 para 843. A fim de manter a consistência, optou-se pelo ano de 2017 para coleta de todas as informações do trabalho.

A importância desta etapa reside no fato que a análise econométrica espacial requer a identificação da dependência espacial entre observações de diferentes pontos geográficos e dos efeitos da interação entre municípios vizinhos. Assim, a falta de dados pode reduzir a precisão das estimativas e impossibilitar a estimação correta de transbordamentos espaciais, sendo mais grave para observações que compõem clusters no território (Arbia et al., 2016).

A manipulação dos dados faltantes resultou em grande quantidade de zeros para a variável da cobertura de esgoto: cerca de 42% das observações, gerando receios de que a elevada distorção na distribuição não permitisse que o efeito linear do modelo fosse devidamente acomodado pela variável independente da coleta de esgoto. A fim de adereçar esse problema, foi introduzida uma dummy igual a 1, caso a cobertura de saneamento municipal seja nula, e igual a 0, caso ela exista. Dessa forma, o coeficiente da variável contínua da cobertura passa a representar o efeito da ampliação do saneamento em cidades em que a infraestrutura é presente.

Na esfera dos dados educacionais, foram coletadas informações agregadas sobre os anos iniciais⁴ do Ensino Fundamental na rede pública relativas à preparação docente, infraestrutura escolar e qualidade do ensino – como as notas do Sistema de Avaliação da Educação Básica (Saeb), exame padronizado anual que avalia o desempenho dos alunos em português e matemática. Além disso, três indicadores do Inep foram modificados para facilitar a análise: adequação, esforço e regularidade docentes. A adequação docente representa a porcentagem de professores com formação superior na área que lecionam, enquanto o esforço docente diz respeito à parcela de docentes nos níveis 5 e 6 do indicador homônimo do Inep (docentes que têm mais de 300 alunos e atuam nos três turnos, em duas ou três escolas e em duas ou três etapas.). Para os anos iniciais do Ensino Fundamental, consideram-se os níveis 5 e 6 como indicativos de sobrecarga na profissão (INEP, 2014). O Indicador de Regularidade Docente (IRD) mensura a capacidade de criação de vínculos com alunos e o desenvolvimento de projetos escolares de longo prazo: quanto menor o tempo de permanência médio dos professores em uma escola, menor é esta capacidade. O indicador valoriza docentes que lecionam no mesmo local por anos consecutivos, por meio de uma escala de pontuação de 0 a 5. A variável será registrada como a porcentagem de docentes com IRD maior ou igual a 3, o que equivale aos níveis alto e médio-alto de regularidade.

A RAIS serviu como fonte de dados para indicadores municipais⁵ relacionados à parcela de pessoas ocupadas nos municípios e outras características socioeconômicas, calculadas como proporção do total de vínculos ativos em 2017. É o caso da porcentagem de analfabetos, população feminina, emprego no setor agropecuário e população com renda de até um salário-mínimo. A variável relativa à parcela de vínculos de pessoas negras e pardas foi calculada diretamente dos microdados da RAIS, visto que as informações de raça não constam no sistema Dardo.

Por fim, o PIB *per capita* (a valores de 2017) e o nível hierárquico urbano têm como fonte, respectivamente, as Contas Nacionais e o estudo sobre as Regiões de Influência das Cidades, ambos fornecidos pelo IBGE. O Quadro 1 fornece uma definição das variáveis utilizadas no artigo.

⁴A escolha pelos anos iniciais do Ensino Fundamental (5 a 11 anos), em detrimento dos anos finais ou do Ensino Médio, se deu em virtude da maior susceptibilidade das crianças a doenças relacionadas à falta de saneamento, como citado na seção 1.

⁵O uso da RAIS apresenta limitações, visto que os dados englobam apenas vínculos trabalhistas, e não a população total do município. Contudo, configura uma *proxy* para as informações municipais de 2017, uma vez que a última edição do Censo Demográfico ocorreu em 2010.

Quadro 1.: Descrição das variáveis utilizadas

| Variável | Descrição | Fonte |
|------------|--|-------------|
| perc_esg | Percentual da população do município com ligação à rede geral de coleta de esgoto. | SNIS e PNSB |
| saeb | Nota padronizada do Saeb. | Inep |
| dist | Distorção idade-série: percentual de alunos com idade superior à recomendada para a série frequentada. | |
| ad | Adequação docente: percentual de professores com ensino superior na área que lecionam. | |
| a_turma | Média de alunos por turma. | |
| ed | Esforço docente: percentual de professores que se enquadram na categoria de sobrecarga de trabalho, segundo o Inep. | |
| horas_aula | Média de horas de aula diárias. | |
| rd | Regularidade docente: percentual de professores com indicador alto e médio-alto. | |
| pib_pc | PIB per capita municipal (R\$), a valores de 2017. | |
| hierarq | Conjunto de variáveis do tipo <i>dummy</i> indicadoras do nível de hierarquia urbana do município, a saber: metrópole, capital regional, centro sub-regional, centro de zona e centro local (categoria de referência). | IBGE |
| perc_vinc | Percentual de vínculos trabalhistas ativos em 31/12/2017 na população total do município. | RAIS |
| F | Percentual de mulheres no total de vínculos ativos. | |
| agro | Percentual de trabalhadores do setor agropecuário no total de vínculos ativos. | |
| analf | Percentual de pessoas analfabetas no total de vínculos ativos. | |
| renda_sm | Percentual de pessoas com renda até 1 salário-mínimo no total de vínculos ativos. | |
| ppardos | Percentual de negros e pardos no total de vínculos ativos. | |
| dummy | <i>Dummy</i> igual a 1 para municípios sem cobertura de esgoto, e igual a 0 para locais onde o serviço existe. | |

Fonte: Elaboração própria a partir de SNIS, PNSB, IBGE, Inep e RAIS.

A Tabela 1 provê as estatísticas descritivas dos dados.

Tabela 1. Estatística descritiva

| Variável | N | Média | Desvio padrão | Min | Pctl. 25 | Pctl. 75 | Max |
|------------------------|------|----------|---------------|---------|----------|----------|-----------|
| perc_esg | 4727 | 28.86 | 36.27 | 0 | 0 | 60.44 | 100 |
| saeb | 5462 | 5.07 | 0.61 | 2.62 | 4.65 | 5.49 | 7.24 |
| dist | 5570 | 12.24 | 9.2 | 0 | 4.7 | 18.5 | 57.6 |
| ad | 5570 | 63.52 | 22.25 | 0 | 48.7 | 80.88 | 100 |
| a_turma | 5570 | 19.77 | 3.82 | 4.3 | 17.5 | 22.4 | 33.9 |
| horas_aula | 5570 | 4.46 | 0.74 | 3.6 | 4 | 4.5 | 9.9 |
| rd | 5570 | 49.47 | 25.16 | 0 | 30.7 | 66.7 | 100 |
| ed | 5570 | 4.9 | 5.33 | 0 | 0 | 7.2 | 50 |
| pib_pc | 5570 | 21990.76 | 20946.32 | 3285.04 | 9492.89 | 27051.62 | 344847.17 |
| hierarq | 5570 | | | | | | |
| ... Metrópole | 177 | 3% | | | | | |
| ... Capital Regional | 189 | 3% | | | | | |
| ... Centro Subregional | 164 | 3% | | | | | |
| ... Centro de Zona | 561 | 10% | | | | | |
| ... Centro Local | 4479 | 80% | | | | | |
| perc_vinc | 5570 | 13.99 | 9.58 | 0.16 | 7 | 18.21 | 128.61 |
| F | 5570 | 46.05 | 9.52 | 6.31 | 40.28 | 52.06 | 96.24 |
| agro | 5570 | 10.34 | 12.61 | 0 | 1.11 | 14.76 | 87.68 |
| analf | 5570 | 0.6 | 1.36 | 0 | 0.13 | 0.67 | 66.76 |
| renda_sm | 5570 | 9.03 | 8.08 | 0 | 3.81 | 11.53 | 85.31 |
| ppardos | 5570 | 21.91 | 17.37 | 0 | 7.36 | 33.33 | 94.41 |
| dummy | 5570 | | | | | | |
| ... 0 | 3206 | 58% | | | | | |
| ... 1 | 2364 | 42% | | | | | |

Fonte: Elaboração própria a partir de SNIS, PNSB, IBGE, Inep e RAIS.

A partir dessas variáveis serão estimados dois modelos. Um deles terá a nota do

Saeb municipal como variável dependente e o outro, a taxa de distorção idade-série (equações (1) e (2), respectivamente). Ambos possuem as mesmas variáveis explicativas: a cobertura de esgotamento, cuja relação com a educação configura a principal motivação do trabalho, e as variáveis de controle relativas à infraestrutura escolar e demografia dos municípios, agrupadas no vetor X_i .

$$saeb_i = \alpha + perc_esg_i + X_i\beta + \varepsilon_i \quad (1)$$

$$dist_i = \gamma + perc_esg_i + X_i\beta + v_i \quad (2)$$

A próxima seção apresenta a metodologia empregada para a análise exploratória de dados espaciais (AEDE) e para a estimação de modelos econométricos.

3.2 Metodologia

A autocorrelação espacial é definida como a correlação de uma variável em diferentes espaços geográficos, isto é, a dependência entre o valor de uma observação e os valores de observações vizinhas (LeSage e Pace, 2009). O primeiro passo para mensurar a dependência espacial consiste na imposição de um arranjo para a ocorrência das interações espaciais, por meio da especificação de uma matriz de pesos espaciais/vizinhança (W). Assumindo a existência de n objetos espaciais (cidades, estados, países etc.), W é uma matriz quadrada de dimensão $n \times n$ que define quais observações serão classificadas como vizinhas (Sarrias, 2020). Há diversas formas de especificar os pesos da matriz de vizinhança. Uma das mais comuns configura a contiguidade direta das observações: apenas locais imediatamente próximos são classificados como vizinhos. Em alusão às peças do jogo de xadrez, a convenção de contiguidade é do tipo *rook* (torre) se considerar apenas fronteiras com extensão diferente de zero ou é do tipo *queen* (rainha) se também incluir os vértices contíguos. O critério de distância geográfica também é comumente utilizado para a elaboração de matrizes de vizinhança. Sua intuição é que regiões mais próximas (k vizinhos mais próximos) sofrerão interações mais intensas que locais distantes geograficamente⁶.

O próximo passo reside na mensuração da autocorrelação espacial. A ferramenta mais comumente usada para este fim é o Índice de Moran, que assume a forma global, local ou bivariada. O Índice de Moran global mede o nível de autocorrelação geral nos dados e é descrito por:

$$I = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^n (x_i - \bar{x})^2 / n} = \frac{n \sum_i \sum_j w_{ij} z_i z_j}{S_0 \sum_{i=1}^n z_i^2} \quad (3)$$

Em que n é o número de regiões, z representa os valores da variável de interesse x , padronizada, e S_0 é igual a $\sum_i \sum_j w_{ij} z_i z_j$. Reescrevendo I na forma matricial e

⁶Uma discussão mais aprofundada sobre os tipos de matrizes de vizinhança foge do escopo do presente trabalho, mas pode ser encontrada em Almeida (2012)

padronizando W :

$$I = \frac{z'Wz}{z'z} \quad (4)$$

O Índice de Moran é padronizado, varia entre -1 e 1, e testa a hipótese nula de aleatoriedade espacial: a estatística estimada é comparada com a distribuição obtida em sucessivas permutações dos dados. Valores de I que excedem o valor esperado de $-[1/(n-1)]$ são positivos e – se significativos estatisticamente – indicam a autocorrelação espacial positiva. Isto é, altos (baixos) valores da variável de interesse tendem a estar circundados por altos (baixos) valores dessa variável em regiões vizinhas. Já nas situações em que I é menor que o valor esperado, a estimativa é negativa, e aponta para maior grau de dispersão da variável. Quanto mais próximo de 1 (em módulo), mais forte é a autocorrelação.

O Índice de Moran local (equação (5)) permite analisar as relações de dependência espacial para cada observação. É útil na medida em que porções do território podem apresentar padrões de autocorrelação que diferem da estatística global ou que sequer apresentam autocorrelação espacial.

$$I_i = z_i \sum_j w_{ij} z_j \quad (5)$$

Na qual o somatório dos pesos abrange apenas observações vizinhas de i .

O Índice de Moran pode ser utilizado, ainda, em sua forma bivariada, que quantifica o grau de associação linear (positiva ou negativa) entre o valor de uma variável em uma dada região e a média de outra variável em regiões vizinhas. O indicador bivariado global na forma matricial e normalizado na linha é dado como na fórmula (6) abaixo. Este índice também assume uma forma local, representada em (7):

$$I^{z_1 z_2} = \frac{z_1' W z_2}{z_1' z_1} \quad (6)$$

$$I_i^{z_1 z_2} = z_{1i} W z_{2i} \quad (7)$$

Nas quais z_1 e z_2 são duas variáveis padronizadas. No artigo, esse indicador é utilizado para avaliar a associação entre educação e saneamento básico.

Ressalta-se que o Índice de Moran também pode ser utilizado para a especificação adequada da matriz de vizinhança, segundo a abordagem de Baumont et al. (2004). Os autores sugerem selecionar, dentre uma série de matrizes de contiguidade e k vizinhos mais próximos, aquela que gera o maior valor do Índice de Moran estatisticamente significativo para os resíduos dos modelos de regressão por mínimos quadrados ordinários (MQO). Ademais, quando o I de Moran dos resíduos de um modelo MQO denota significância estatística, há indicação para estimação por meio de um modelo de regressão espacial, visto que os dados violam pressupostos básicos de aleatoriedade

da amostra e de normalidade do termo do erro, para testes de hipótese.

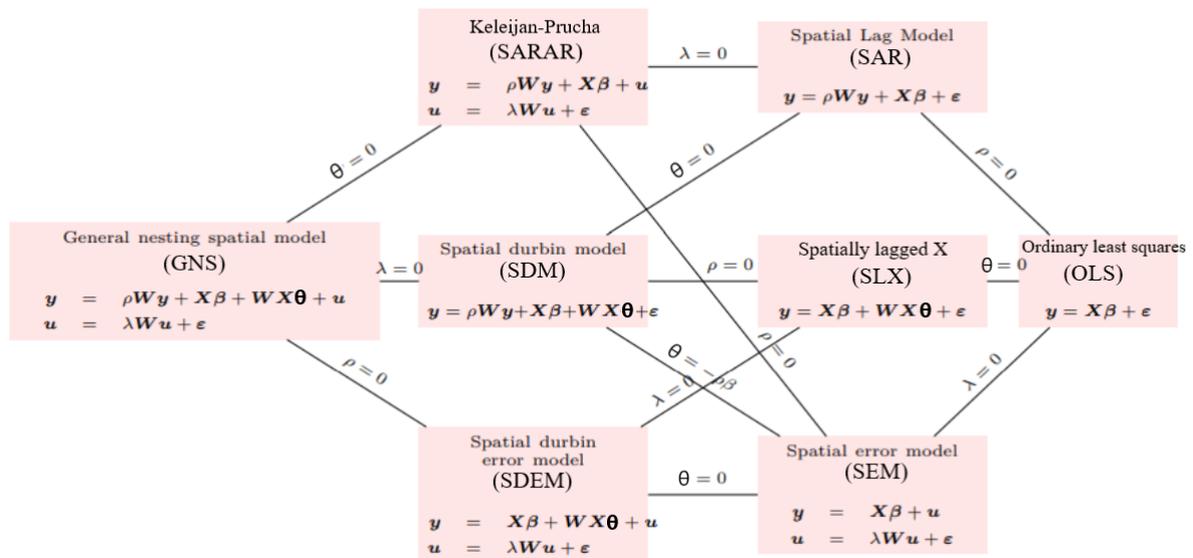
Os modelos de regressão espacial se diferenciam de acordo com o componente em que há defasagem espacial: pode estar presente na variável dependente, por meio da multiplicação de um fator autorregressivo espacial ρ e da matriz de ponderação espacial W (ρWy); nas variáveis explicativas ($WX\theta$); ou no termo do erro (λWu); ou ainda em uma combinação dessas possibilidades. ρ , θ e λ variam entre -1 e 1, sendo que sinais positivos indicam autocorrelação positiva e sinais negativos apontam dispersão dos dados. O modelo mais geral – o modelo de Manski, ou modelo de dependência espacial generalizada (GNS) – possui os três tipos de defasagem e é retratado em (8):

$$y = \rho Wy + X\beta + WX\theta + u, u = \lambda Wu + \varepsilon \quad (8)$$

O termo $X\beta$ indica como uma alteração na variável explicativa em uma dada região se associa a mudanças na variável dependente desta região, isolando-se o efeito das demais variáveis. Já $WX\theta$ parte do princípio de que as observações de X são espacialmente correlacionadas, de forma que a variável dependente de um local é afetada por distúrbios das variáveis explicativas de locais vizinhos. O termo λWu é o componente autorregressivo do erro u e permite identificar como fatores não observados de regiões próximas se relacionam com a variável de interesse y . Finalmente, ρWy caracteriza a associação da variável dependente em uma região com a média dos valores da variável dependente em regiões vizinhas. Tanto $WX\theta$ quanto λWu modelam efeitos *spillover* (de transbordamento) locais, isto é, y é afetado apenas por distúrbios em X ou u provenientes de observações vizinhas nos critérios determinados pela matriz de vizinhança. Já o termo ρWy caracteriza um efeito *feedback* global ou endógeno, visto que mudanças em y dariam início a uma reação em cadeia para (potencialmente) toda a amostra (LeSage, 2014).

O GNS é raramente usado para estudos empíricos, pois não é possível separar de maneira apropriada os efeitos resultantes de cada tipo de defasagem (Manski, 1993). A solução é a imposição de restrições nos parâmetros ρ , θ e λ , dando origem às principais especificações de modelos de regressão espacial *cross-section*, como retratado na Figura 1.

A escolha do modelo de regressão, assim como a seleção da matriz de pesos espaciais, ainda é um tema em discussão na literatura. O artigo opta por uma abordagem abrangente, porém consistente com as possibilidades de especificação definidas no arcabouço teórico de econometria espacial. Serão estimados os principais modelos de regressão espacial *cross-section* e a escolha se dará de acordo com aquele que obtiver o melhor ajuste de acordo com a estatística de *log-verossimilhança* e os critérios de informação Akaike (AIC) e bayesiano (BIC), definidos pelas equações (9) e (10), respectivamente. Em seguida, a necessidade de restrição do modelo será testada pela razão de verossimilhança. Por fim, aplica-se o Índice de Moran sobre os resíduos para

Figura 1. Modelos de regressão espacial

Fonte: Sarrias (2020), modificado.

verificar se a autocorrelação espacial foi devidamente capturada pelo modelo.

$$AIC = 2k - 2 \ln L' \quad (9)$$

$$BIC = k \ln n - 2 \ln L' \quad (10)$$

Em que k é a quantidade de parâmetros estimados, n é o tamanho da amostra e L' é o valor máximo da função de *log*-verossimilhança. Os critérios de informação permitem escolher o modelo que possui o melhor ajuste *relativo*. Ambas as estatísticas premiam a qualidade do ajuste (via função de verossimilhança) e penalizam a inclusão de parâmetros estimados, de modo que o modelo com melhor ajuste terá o menor valor para o AIC e BIC e o maior para L' .

Por fim, aplica-se o teste de razão de verossimilhança (11). A estatística assume que a *log*-verossimilhança do modelo não restrito é maior que a do restrito, e testa a significância da diferença entre ambas. Se for significativa, o modelo mais simples deve ser descartado; caso contrário, o modelo ilimitado não agrega valor, então não deve ser utilizado (Kopczewska, 2020).

$$RV = -2 \left(\ln \frac{L'_{\text{modelo restrito}}}{L'_{\text{modelo irrestrito}}} \right) \quad (11)$$

Estabelecidos os procedimentos metodológicos, a próxima seção expõe os resultados obtidos.

4. Resultados

Seria o espaço determinante para explicar as desigualdades regionais entre indicadores educacionais e de saneamento básico? Estaria o desempenho escolar em uma cidade condicionado não apenas ao contexto local, mas também à situação da educação e de saneamento em municípios vizinhos? A AEDE investigará a presença de autocorrelação no esgotamento, notas do Saeb e taxa de distorção idade-série, além da existência de *clusters* abaixo ou acima da média nacional para as variáveis de interesse. Em seguida, os modelos econométricos quantificam a associação entre a cobertura de saneamento e os indicadores educacionais, a fim de testar tanto a hipótese de que essa relação existe no âmbito local, como de que gera transbordamentos espaciais significativos.

4.1 Escolha da matriz de pesos espaciais e análise exploratória de dados espaciais

A escolha da matriz de pesos espaciais seguiu os critérios sugeridos por Baumont et al. (2004). Dessa forma, para os resíduos dos dois modelos estimados (SAEB e distorção idade-série), o Índice de Moran foi aplicado com base em 22 matrizes de vizinhança, a fim de selecionar aquela que melhor explica o padrão de defasagem espacial nos dados: contiguidade do tipo *queen*, *rook* e uma série de matrizes de k vizinhos mais próximos, nas quais k vai de 1 a 20. Para todas as matrizes, os índices de Moran globais são positivos e significativos, indicando a presença de dependência espacial positiva. Rejeita-se a hipótese nula de distribuição aleatória dos resíduos, a uma significância de 0,1%. Tendo em vista que as matrizes de contiguidades do tipo *rook* e *queen* apresentaram o maior valor do Índice de Moran para os modelos do Saeb e da distorção idade-série, respectivamente, essas serão utilizadas para a estimação dos resultados econométricos do trabalho. Os resíduos do modelo da distorção idade-série, com I de Moran igual a 0,481, aparentam possuir uma autocorrelação espacial mais forte (dados mais concentrados no território) que a nota padronizada do Saeb, cuja estatística é de 0,387.

Além da aplicação para escolha da matriz de pesos espaciais, o índice de Moran configura ferramenta útil para a análise da distribuição de variáveis específicas no território, bem como a identificação de *clusters* de valores abaixo ou acima da média amostral. Esta subseção explora a autocorrelação espacial nas três principais variáveis de interesse do trabalho: as notas do Saeb, distorção idade-série e a cobertura de esgotamento.

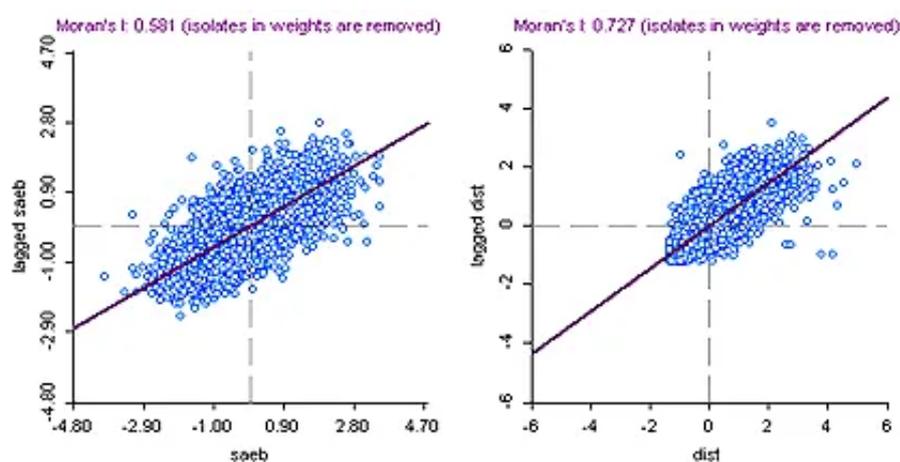
Os índices de Moran globais para os indicadores educacionais são positivos (0,581 para o Saeb e 0,727 para a distorção idade-série) e significativos estatisticamente, isto é, há indícios de que municípios com indicadores de qualidade do ensino acima da média são vizinhos de municípios com indicadores também favoráveis. Analogamente, o contrário – proximidade geográfica entre cidades com indicadores abaixo da média

– também ocorre.

O índice de Moran global fornece uma perspectiva geral da dependência espacial nos dados. Contudo, em todo o território, é possível encontrar conjuntos de observações que vão contra a tendência de autocorrelação positiva, ou nas quais a dependência espacial sequer está presente. A Figura 2 exibe os diagramas de dispersão de Moran para as variáveis educacionais.

O quadrante superior direito (Alto-Alto) exibe observações acima da média que estão cercadas por outras acima da média, enquanto o quadrante inferior esquerdo (Baixo-Baixo) mostra locais que formam *clusters* de indicadores abaixo da média. Ambos os casos refletem uma dependência espacial positiva. Já os quadrantes superior esquerdo e inferior direito indicam tendências de autocorrelação espacial negativa: cidades com valores abaixo do *benchmark* nacional e que se avizinham de outras acima do padrão (Baixo-Alto), ou vice-versa (Alto-Baixo).

Figura 2. Diagramas de dispersão de Moran – Saeb (a) e Distorção idade-série (b)



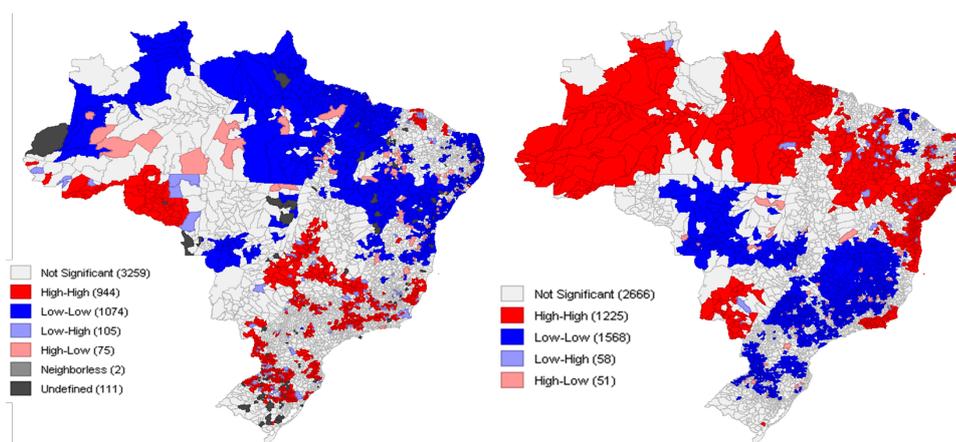
Fonte: Elaboração própria, a partir de dados do Inep e IBGE.

O índice de Moran local foi utilizado para facilitar a visualização dos agrupamentos no território e sua significância estatística. No caso das notas do Saeb, a autocorrelação espacial abrange 2.198 municípios, a uma significância de 5% (Figura 3, painel (a)). *Clusters* Alto-Alto podem ser observados principalmente em cidades do Centro-Sul, Ceará, Rondônia e Acre, embora entre os locais com notas acima da média também se identifiquem cidades destoantes do padrão (*clusters* Baixo-Alto). Já os *clusters* Baixo-Baixo estão presentes no Mato Grosso e em maior proporção no Norte e Nordeste, onde também são mais comuns os agrupamentos Alto-Baixo – uma das exceções é o município do Rio de Janeiro (RJ). Para a distorção idade-série (Figura 3, painel (b)), *clusters* Baixo-Baixo são observados principalmente em Mato Grosso, Goiás, São Paulo, Minas Gerais, Paraná, Santa Catarina e Ceará⁷, enquanto

⁷Os resultados positivos para o Ceará, em contraposição ao restante do Nordeste, podem ser reflexo das políticas educacionais implementadas no estado a partir dos anos 2000 (Verdiano, 2019).

agrupamentos com valores acima da média estão presentes em parte considerável do Nordeste e Norte, além de no Mato Grosso do Sul e em quase a totalidade do Rio de Janeiro. Agrupamentos do tipo Baixo-Alto ou Alto-Baixo são menos comuns, contabilizando 109 municípios. No total, 2.904 municípios apresentaram algum tipo de dependência espacial para a desfasagem idade-série.

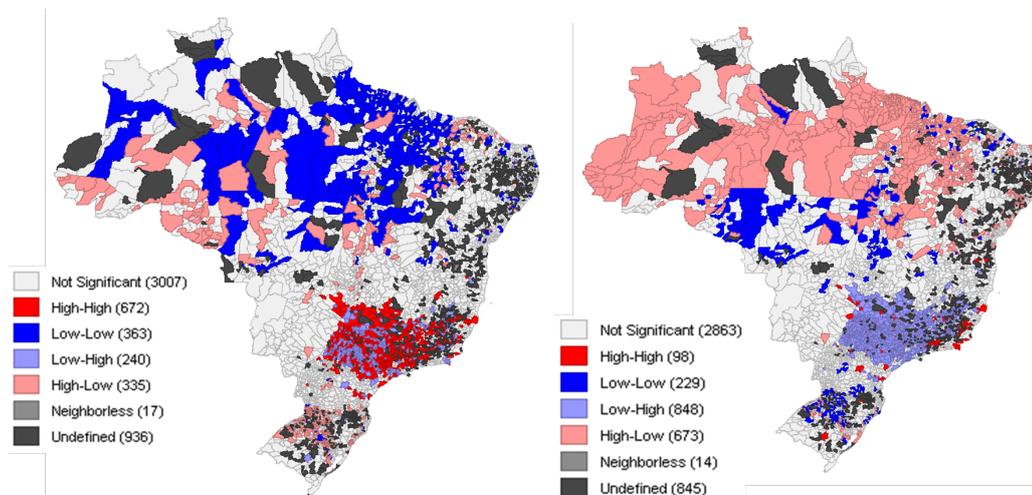
Figura 3. Índice de Moran local, notas do Saeb (a) e distorção idade-série (b) (Brasil, 2017)



Fonte: Elaboração própria, a partir de dados do Inep e IBGE.

Por fim, cabe analisar como ambas as variáveis se relacionam com a cobertura de esgoto no território. Em consonância com o esperado, o Índice de Moran bivariado global encontrou uma autocorrelação espacial positiva (0.188) entre o Saeb e a coleta de esgoto, sendo que a estatística local foi significativa em 1.610 municípios (Figura 4, painel (a)). Há agrupamentos Baixo-Baixo concentrados na região Norte e agrupamentos Alto-Alto principalmente no Sudeste, onde também estão a maior parte dos municípios do padrão Baixo-Alto. Já *clusters* Alto-Baixo são mais comuns no Norte, Mato Grosso e no Rio Grande do Sul. O painel (b) da Figura 4 demonstra a relação entre a distorção idade-série e a coleta de esgoto. Nesse caso, o Índice de Moran bivariado indica uma autocorrelação espacial negativa (-0.273) e a estatística local identificou significância em 1.848 cidades. Em outras palavras, a taxa de distorção idade-série em um município guarda uma associação negativa com a cobertura de esgoto em regiões vizinhas. O painel mostra agrupamentos Alto-Baixo em quase a totalidade da região Norte e parte da região Nordeste, e *clusters* Baixo-Alto no Sudeste. A grande quantidade de dados faltantes da cobertura de esgoto para o Nordeste pode ter prejudicado a identificação da dependência espacial bivariada tanto para o Saeb quanto para a distorção idade-série nesta região.

Figura 4. Figura 4: Índice de Moran local bivariado (Brasil, 2017). Saeb (X) e coleta de esgoto (Y) (a) e distorção idade-série (X) e coleta de esgoto (Y) (b)



Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS e PNSB.

A próxima seção se destina a identificar e estimar um modelo de regressão espacial que capture com eficácia os tipos de interações espaciais entre os indicadores municipais de saneamento e qualidade do ensino, além de outras variáveis de controle.

4.2 Modelos econométricos

4.2.1 Cobertura sanitária e notas do Saeb

Os testes LM-erro e LM-lag tradicionais foram executados para o modelo do Saeb e ambos rejeitaram a hipótese nula. Ou seja, o teste LM-erro indicou a presença da defasagem na variável dependente, enquanto o teste LM-lag favoreceu a existência de defasagem no termo do erro. A seguir, executaram-se as versões robustas desses testes e, novamente, ambos rejeitaram as hipóteses nulas de ausência de *lag* na variável dependente e no erro, respectivamente. O teste SARMA também apontou para a presença de ambos os tipos de defasagem no modelo. A Tabela 2 apresenta os resultados.

Tabela 2. Resultados dos testes multiplicadores de Lagrange – Saeb

| Teste | Estatística | p-valor |
|-----------------|-------------|---------|
| LM-erro | 1591,5 | <0.001 |
| LM-lag | 1178 | <0.001 |
| LM-erro robusto | 497,57 | <0.001 |
| LM-lag robusto | 84,073 | <0.001 |
| SARMA | 1675,5 | <0.001 |

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Foram estimados os principais modelos lineares de regressão espacial *cross-section* (SDM, SDEM, SLX, SEM, SAR e SARAR), além do modelo por mínimos quadrados ordinários (MQO). O modelo SDEM aparenta possuir o melhor ajuste entre os estimados

na medida em que apresentou os menores valores para os critérios de informação e o maior valor para a estatística de *log-verossimilhança*, e, portanto, foi o selecionado para avaliar a associação entre as notas do Saeb e a cobertura de saneamento básico. A Tabela 3 apresenta as estimativas obtidas:

Tabela 3. Resultados dos modelos para as notas do Saeb (Brasil, 2017) (continua)

| Variável | MGO | SLX | SDM | SEM | SDEM | SAR | SARAR |
|---------------------------|------------------------|-----------------------|-----------------------|-----------------------|------------------------------|------------------------|-----------------------|
| perc_esg | -0.0001 (0.0003) | -0.0007 (0.0004) | -0.0005 (0.0003) | 0.0001 (0.0003) | -0.0004 (0.0003) | -0.0004 (0.0003) | 0.0000 (0.0003) |
| ad | 0.0055*** (0.0004) | 0.0019*** (0.0005) | 0.0019*** (0.0004) | 0.0027*** (0.0004) | 0.0021*** (0.0004) | 0.0034*** (0.0003) | 0.0024*** (0.0004) |
| rd | 0.0005 (0.0003) | 0.0005 (0.0003) | 0.0006* (0.0002) | 0.0004 (0.0002) | 0.0006* (0.0002) | 0.0005* (0.0002) | 0.0004 (0.0002) |
| ed | 0.0058*** (0.0015) | 0.0023 (0.0015) | 0.0016 (0.0013) | 0.0027* (0.0012) | 0.0025 (0.0013) | 0.0035** (0.0013) | 0.0023 (0.0012) |
| a_turma | -0.0185*** (0.0023) | -0.0020 (0.0025) | 0.0046* (0.0022) | -0.0044* (0.0021) | -0.0040 (0.0022) | -0.0090*** (0.0020) | -0.0034 (0.0021) |
| horas_aula | 0.0108 (0.0110) | 0.0143 (0.0108) | 0.0316*** (0.0095) | 0.0144 (0.0093) | 0.0126 (0.0098) | 0.0163 (0.0096) | 0.0128 (0.0090) |
| pib_pc | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) |
| Metrópole | -0.01211** (0.0451) | -0.0364 (0.0802) | 0.0114 (0.0697) | -0.0074 (0.0581) | 0.0082 (0.0625) | -0.1066*** (0.0395) | 0.0157 (0.0599) |
| Capital regional | -0.0214 (0.0441) | 0.0603 (0.0485) | 0.0716 (0.0422) | 0.0995* (0.0411) | 0.0660 (0.041) | -0.0034 (0.0387) | 0.1033* (0.0405) |
| Centro subregional | 0.0322 (0.0459) | 0.03519 (0.0442) | 0.0056 (0.0384) | 0.0499 (0.0345) | 0.0493 (0.0415) | 0.0360 (0.0402) | 0.0493 (0.0333) |
| Centro de Zona | 0.07995** (0.0265) | 0.0567* (0.0260) | 0.0333 (0.0226) | 0.0455* (0.0205) | 0.0551* (0.0243) | 0.0599* (0.0232) | 0.0420* (0.0199) |

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

A análise dos *p*-valores das estimativas permite constatar que, *para as cidades com infraestrutura de saneamento*, não foi encontrada significância estatística entre a nota padronizada do Saeb e a ampliação da cobertura de esgoto, nem entre o Saeb local e o esgotamento de cidades vizinhas. Por outro lado, a existência de saneamento básico – em comparação ao cenário de ausência do serviço – está associada a um efeito positivo de 0.074 na nota do Saeb (erro padrão de 0.024), isolando-se os efeitos das demais variáveis. A estimativa do *lag* desta variável não foi significativa. Portanto, descarta-se a hipótese de transbordamentos espaciais das externalidades da cobertura de esgoto sobre este indicador educacional.

No âmbito das variáveis controle, a adequação docente, a regularidade docente e o percentual de vínculos ativos têm o sinal esperado positivo e significativo sobre a nota municipal do Saeb. O modelo também encontra uma associação positiva caso o município configure um Centro de Zona, em comparação a cidades classificadas como Centro Local. Já o aumento de vínculos no setor agropecuário e vínculos analfabetos guarda uma relação negativa com a nota da educação municipal, respeitando os sinais esperados⁸ para esses indicadores.

⁸Para um estudo mais aprofundado da relação entre emprego agrícola e nível de desenvolvimento humano, ver Köle et al. (2021).

Tabela 3. Resultados dos modelos para as notas do Saeb (Brasil, 2017) (continuação)

| Variável | MGO | SLX | SDM | SEM | SDEM | SAR | SARAR |
|-------------------------------|------------------------|------------------------|-------------------------|-----------------------|-------------------------------|------------------------|-----------------------|
| ppard | -0.0044*** (0.0005) | -0.0001 (0.0006) | 0.0007 (0.0005) | -0.0014** (0.0005) | -0.0005 (0.0005) | -0.0025*** (0.0004) | -0.0009 (0.0005) |
| perc_vinc | 0.0190*** (0.0012) | 0.0102*** (0.0012) | 0.0086*** (0.0010) | 0.0086*** (0.0010) | 0.0101*** (0.0010) | 0.0133*** (0.0010) | 0.0076*** (0.0010) |
| F | -0.0010 (0.0011) | 0.0001 (0.0011) | 0.0019* (0.0009) | -0.0014 (0.0009) | -0.0000 (0.0010) | 0.0000 (0.0009) | -0.0013 (0.0009) |
| agro | -0.0014 (0.0007) | -0.0025** (0.0008) | -0.0015* (0.0007) | -0.0015* (0.0007) | -0.0022** (0.0007) | -0.0200** (0.0006) | -0.0015* (0.0006) |
| analf | -0.0315*** (0.0060) | -0.0158** (0.0059) | -0.0112* (0.0051) | -0.0103* (0.0049) | -0.0125* (0.0053) | -0.0193*** (0.0053) | -0.0087 (0.0048) |
| renda_sm | -0.0018 (0.0011) | 0.0012 (0.0011) | 0.0014 (0.0009) | 0.0011 (0.0009) | 0.0006 (0.0010) | -0.0006 (0.0009) | -0.0008 (0.0009) |
| dummy0 | 0.1024*** (0.0266) | 0.0807** (0.0268) | 0.0703** (0.0233) | 0.0572* (0.0227) | 0.0739** (0.0240) | 0.0850*** (0.0233) | 0.0516* (0.0221) |
| lag.perc_esg | - | -0.0011 (0.0006) | 0.0004 (0.0005) | - | -0.0001 (0.0006) | - | - |
| lag.ad | - | 0.0040*** (0.0006) | -0.0002 (0.0006) | - | 0.0038*** (0.0007) | - | - |
| lag.rd | - | 0.0007 (0.0005) | -0.0006 (0.0004) | - | 0.0006 (0.0006) | - | - |
| lag.ed | - | 0.0033 (0.0027) | -0.0021 (0.0023) | - | 0.0044 (0.0030) | - | - |
| lag.a_turma | - | -0.0225*** (0.0037) | -0.0300*** (0.0032) | - | -0.0217*** (0.0042) | - | - |
| lag.horas_aula | - | -0.0141 (0.0182) | -0.0864*** (0.01601) | - | 0.0020 (0.0199) | - | - |
| lag.pib_pc | - | -0.0000 (0.0000) | -0.0000 (0.0000) | - | 0.0000 (0.0000) | - | - |
| lag.Metrópole | - | -0.2954** (0.0996) | -0.1838* (0.0866) | - | -0.2271* (0.0959) | - | - |
| lag.Capital regional | - | -0.3979*** (0.0738) | -0.2480*** (0.0642) | - | -0.2486** (0.0768) | - | - |
| lag.Centro subregional | - | -0.1328 (0.0758) | -0.1000 (0.0659) | - | -0.0284 (0.0812) | - | - |
| lag.Centro de Zona | - | 0.0809 (0.0464) | 0.0329 (0.0404) | - | 0.0557 (0.0503) | - | - |
| lag.ppard | - | -0.0031*** (0.0008) | -0.0030*** (0.0007) | - | -0.0041*** (0.0009) | - | - |
| lag.perc_vinc | - | 0.0273*** (0.0022) | 0.0089*** (0.0019) | - | 0.0185*** (0.0025) | - | - |
| lag.F | - | -0.0005 (0.0017) | -0.0134*** (0.0015) | - | 0.0001 (0.0018) | - | - |
| lag.agro | - | -0.0020 (0.0013) | -0.0045*** (0.0011) | - | -0.0010 (0.0014) | - | - |
| lag.analf | - | -0.0692*** (0.0121) | -0.0468*** (0.0105) | - | -0.0279* (0.0133) | - | - |
| lag.renda_sm | - | 0.0050* (0.0020) | 0.0022 (0.0018) | - | 0.0024 (0.0023) | - | - |
| lag.dummy0 | - | 0.1514*** (0.0430) | 0.0231 (0.0375) | - | 0.0711 (0.0498) | - | - |
| Rho | - | - | 0.4832*** | - | - | 0.4106*** | -0.1062*** |
| Lambda | - | - | - | 0.6488*** | 0.5675*** | - | 0.7194*** |
| AIC | 7301,072 | 6838,449 | 5833,89 | 5880,628 | 5664,974 | 6275,422 | 5860,203 |
| BIC | 7429,904 | 7083,23 | 6085,154 | 6015,902 | 5916,196 | 6410,696 | 6001,918 |
| log-verossimilhança | -3630,536 | -3381,224 | -2877,97 | -2919,31 | -2793,49 | -3116,71 | -2908,1 |
| Observações | 4636 | 4636 | 4636 | 4636 | 4636 | 4636 | 4636 |

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.
Significância: *** p < 0.001; ** p < 0.01, *p < 0.05. Erros padrão apresentados entre parênteses.

Transbordamentos espaciais *positivos* sobre o Saeb foram identificados para a adequação docente e proporção de vínculos ativos na população. Já efeitos *spillover* espaciais negativos ocorrem para a quantidade de alunos por turma, proporção de pessoas negras e pardas nos vínculos ativos e percentual de analfabetos. Isso significa que a ampliação desses fatores em cidades vizinhas está associada a uma redução local na nota do Saeb. Similarmente, transbordamentos espaciais negativos ocorrem quando um município se situa próximo a capitais regionais e metrópoles, em comparação à vizinhança com centros locais.

O coeficiente Lambda é estatisticamente significativo, corroborando a presença de dependência espacial no termo do erro. Isto é, o componente da variabilidade das notas do Saeb não explicada pelo modelo é positivamente correlacionado no espaço, o que pode ser resultado da omissão de variáveis explicativas relevantes.

Ao aplicar o índice de Moran sobre os resíduos do SDEM, obtém-se um *p*-valor não significativo (~ 0.99), o que indica que o modelo foi eficaz em capturar a dependência espacial existente nos dados. Com exceção do SARAR e do SDM, todos os outros modelos falharam neste teste. Ademais, o *p*-valor encontrado para o SDM foi consideravelmente mais próximo da significância estatística (0.07).

Por fim, os testes de razão de verossimilhança rejeitaram a hipótese nula de que a restrição do SDEM para modelos mais simples, tais como o SLX, SEM ou MQO, seria mais adequada.

4.2.2 Cobertura sanitária e taxa de distorção idade-série

Os mesmos procedimentos seguidos no modelo com a variável dependente do Saeb foram aplicados para o modelo com a distorção idade-série como variável de interesse. Tanto os testes multiplicadores de Lagrange tradicionais como os robustos, além do teste SARMA, rejeitaram as hipóteses nulas, o que indica a presença de defasagem no termo de erro e na variável dependente. Dessa forma, modelos que não possuem nenhum desses tipos de defasagem, como o SLX e o MQO, podem não ser os mais adequados. A Tabela 4 expõe os resultados.

Tabela 4. Resultados dos testes multiplicadores de LaGrange – Distorção idade-série

| Teste | Estatística | p-valor |
|-----------------|-------------|---------|
| LM-erro | 2516 | <0.001 |
| LM-lag | 3080,9 | <0.001 |
| LM-erro robusto | 10,977 | <0.001 |
| LM-lag robusto | 575,95 | <0.001 |
| SARMA | 3091,9 | <0.001 |

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Ao comparar os critérios de informação e a estatística de *log*-verossimilhança obti-

dos nos diferentes modelos espaciais e no MQO, novamente o modelo SDEM aparenta possuir o melhor ajuste, sendo selecionado para a análise das estimativas. Os resultados estão descritos na Tabela 5:

Tabela 5. Resultados dos modelos para a distorção idade-série (Brasil) (continua)

| Variável | MQO | SLX | SDM | SEM | SDEM | SAR | SARAR |
|---------------------------|-------------------------|------------------------|------------------------|------------------------|-------------------------------|------------------------|------------------------|
| perc_esg | -0.0688*** (0.0076) | -0.0208* (0.0083) | -0.0134* (0.0067) | -0.0121* (0.0061) | -0.0175** (0.0067) | 0.0174** (0.0053) | -0.0213*** (0.0045) |
| ad | -0.2327*** (0.0097) | -0.0960*** (0.0126) | -0.0957*** (0.0099) | -0.1073*** (0.0091) | -0.1023*** (0.0094) | -0.0886*** (0.0071) | -0.0692*** (0.0058) |
| rd | 0.0416*** (0.0062) | 0.0199** (0.0061) | 0.0078 (0.0052) | 0.0028 (0.0045) | 0.0105* (0.0051) | 0.0126** (0.0043) | 0.0156*** (0.0036) |
| ed | 0.3947*** (0.0334) | 0.1947*** (0.0061) | 0.1740*** (0.0279) | 0.0978*** (0.0240) | 0.1516*** (0.0273) | 0.1645*** (0.0232) | 0.1824*** (0.0203) |
| a_turma | 0.3187*** (0.0366) | 0.1059** (0.0392) | 0.1822*** (0.0324) | 0.0968** (0.0295) | 0.1356*** (0.0315) | 0.1329*** (0.0255) | 0.1114*** (0.0209) |
| horas_aula | -0.08591*** (0.2281) | -0.7374** (0.2293) | -0.7024*** (0.1984) | -0.5969*** (0.1735) | -0.5601** (0.1880) | -0.4588** (0.1584) | -0.3859** (0.1312) |
| pib_pc | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) | -0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) | 0.0000 (0.0000) |
| Metrópole | 2.3245** (0.8990) | 18.426 (-1.5188) | 3.1935*** (0.7333) | 14.436 (-10.748) | 17.300 (-10.663) | 10.169 (0.6243) | 1.8378*** (0.5530) |
| Capital regional | 4.9313*** (0.8734) | 15.020 (0.9258) | 2.7823*** (0.7128) | 0.5142 (0.6943) | 1.4582* (0.7285) | 1.4512* (0.6068) | 2.3154*** (0.5505) |
| Centro subregional | 2.6615** (0.9098) | 2.0786* (0.8400) | 0.7354 (0.7423) | 0.1677 (0.5631) | 13.521 (0.7390) | 0.4533 (0.6317) | 0.6671 (0.5666) |
| Centro de Zona | 0.4692 (0.5262) | 0.3364 (0.4939) | -.2193 (0.4320) | 0.0329 (0.3350) | 0.1794 (0.4353) | -0.2499 (0.3653) | 0.0642 (0.3336) |
| ppard | 0.0334** (0.0172) | -0.0103 (0.0130) | -0.0008 (0.0101) | 0.0086 (0.0096) | 0.0009 (0.0099) | -0.0026 (0.0074) | -0.0067 (0.0057) |
| perc_vinc | -0.2229*** (0.0241) | -0.0909*** (0.0237) | -0.1245*** (0.0201) | -0.0270 (0.0176) | -0.0705*** (0.0198) | -0.0654*** (0.0169) | -0.1102*** (0.0152) |

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Os resultados apontam significância estatística tanto para a variável contínua da cobertura de esgoto local, quanto para a defasagem espacial desse componente. Para as cidades em que a infraestrutura existe, a ampliação de 1% na cobertura de esgoto municipal estaria associada a uma redução de 0.017% na taxa de distorção idade-série (erro padrão igual a 0.006), enquanto o incremento percentual da cobertura em cidades vizinhas associa-se a uma diminuição de 0.043% do indicador (erro padrão de 0.013). Dessa forma, os efeitos relacionados aos transbordamentos espaciais do saneamento seriam mais fortes que a ampliação da cobertura local, isolando-se a influência das demais variáveis. Em contrapartida, o modelo não encontrou significância estatística para a *dummy* que representa a presença (ou ausência) do serviço no município ou em cidades próximas.

A ampliação da adequação docente, média de horas-aula diárias e percentual de vínculos empregatícios locais e de municípios vizinhos se relaciona com uma redução

Tabela 5. Resultados dos modelos para a distorção idade-série (Brasil) (continuação)

| Variável | MGO | SLX | SDM | SEM | SDEM | SAR | SARAR |
|-------------------------------|------------------------|------------------------|------------------------|---------------------|-------------------------------|---------------------|----------------------|
| F | -0.0211 (0.0225) | -0.0205 (0.0213) | -0.0344 (0.0186) | 0.0237 (0.0158) | 0.0128 (0.0176) | 0.0104 (0.0156) | -0.0268 (0.0139) |
| agro | -0.1361*** (0.0151) | -0.0059 (0.0159) | -0.0119 (0.0128) | 0.0111 (0.0118) | -0.0007 (0.0127) | -0.0141 (0.0105) | -0.0354 (0.0091) |
| analf | 0.5524*** (0.1213) | 0.0988 (0.1133) | 0.2172* (0.0996) | -0.0084 (0.0832) | 0.0823 (0.0968) | 0.1107 (0.0843) | 0.2155** (0.0765) |
| renda_sm | 0.0098 (0.0223) | -0.0427* (0.0217) | -0.0252 (0.0186) | 0.0025 (0.0158) | -0.0219 (0.0182) | -0.0193 (0.0155) | -0.0291* (0.0137) |
| dummy0 | -0.2412 (0.5290) | 0.1256 (0.5114) | 0.6481 (0.4411) | -0.4186 (0.3761) | -0.5674 (0.4382) | -0.1772 (0.3672) | 0.2301 (0.3114) |
| lag.perc_esg | - | -0.0794*** (0.0120) | -0.0514*** (0.0104) | - | -0.0433** (0.0137) | - | - |
| lag.ad | - | -0.1776*** (0.0173) | -0.0388** (0.0135) | - | -0.1248*** (0.0195) | - | - |
| lag.rd | - | 0.1181*** (0.0102) | 0.0496*** (0.0080) | - | 0.0326** (0.0121) | - | - |
| lag.ed | - | 0.7591*** (0.0584) | 0.2619*** (0.0445) | - | 0.2588*** (0.0688) | - | - |
| lag.a_turma | - | 0.3992*** (0.0584) | 0.0556 (0.0475) | - | 0.3052*** (0.0687) | - | - |
| lag.horas_aula | - | -0.6483 (0.3548) | 0.6091* (0.2989) | - | -0.2204 (0.4742) | - | - |
| lag.pib_pc | - | 0.0000 (0.0000) | 0.0000 (0.0000) | - | 0.0000 (0.0000) | - | - |
| lag.Metrópole | - | 0.8892 -18.815 | -3.5893** -12.179 | - | 12.096 -18.326 | - | - |
| lag.Capital regional | - | 9.6293*** -14.038 | 0.9540 -12.407 | - | 4.0777** -14.763 | - | - |
| lag.Centro subregional | - | 7.3809*** -14.580 | 11.316 -12.716 | - | 4.3065** -15.127 | - | - |
| lag.Centro de Zona | - | 11.471 (0.9064) | 11.147 (0.7399) | - | 0.6462 (0.9601) | - | - |
| lag.ppard | - | -0.0136 (0.0172) | 0.0018 (0.0138) | - | 0.0224 (0.0203) | - | - |
| lag.perc_vinc | - | -0.5088*** (0.0429) | -0.1661*** (0.0331) | - | -0.2576*** (0.0506) | - | - |
| lag.F | - | -0.0479 (0.0338) | -0.0449 (0.0314) | - | 0.0186 (0.0362) | - | - |
| lag.agro | - | -0.1907*** (0.0251) | -0.1407*** (0.0201) | - | -0.1155*** (0.0284) | - | - |
| lag.analf | - | 1.1090*** (0.2410) | 0.4519* (0.19997) | - | 0.5259* (0.2553) | - | - |
| lag.renda_sm | - | -0.1492*** (0.0409) | -0.0441 (0.0304) | - | -0.0557 (0.0451) | - | - |
| lag.dummy0 | - | 1.7760* (0.8422) | -0.2552 (0.6962) | - | -10.763 -10.005 | - | - |
| Rho | - | - | 0.4297*** | - | - | 0.7076*** | 0.7643*** |
| Lambda | - | - | - | 0.791*** | 0.7046*** | - | -0.7019*** |
| AIC | 35758,23 | 34885,98 | 34112,41 | 33049,83 | 32870 | 32951,69 | 33998,93 |
| BIC | 35887,45 | 35131,5 | 34364,39 | 33185,51 | 32869,84 | 33087,37 | 34141,07 |
| log-verossimilhança | -17859,1 | -17405 | -17017 | -16503,9 | -16395,9 | -16454,8 | -16977,5 |
| Observações | 4727 | 4727 | 4727 | 4727 | 4727 | 4727 | 4727 |

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Significância: *** p < 0.001; ** p < 0.01, * p < 0.05. Erros padrão apresentados entre parênteses.

da distorção idade-série, sendo que para a variável das horas-aula o coeficiente aponta uma diminuição expressiva, de 0.56%. Já a ampliação da média de alunos por turma e do indicador de esforço (sobrecarga) docente – tanto na esfera municipal quanto de cidades próximas – está associada a um aumento relativo de alunos com idade superior à recomendada para a série que frequentam.

A maior distorção serial também é observada para cidades classificadas como capitais regionais, em comparação aos centros locais. Já para municípios vizinhos de capitais regionais ou de centros sub-regionais, associa-se uma elevação da ordem de 4% deste indicador educacional, comparativamente à proximidade com centros locais. A ampliação dos vínculos de pessoas analfabetas está relacionada com um aumento da distorção idade-série, enquanto esta decresce com o aumento de vínculos no setor agropecuário. Em um resultado contraintuitivo, o modelo aponta um coeficiente positivo e significativo para a proporção de docentes com IRD alto e médio-alto.

A defasagem no termo de erro é corroborada pelo valor do Lambda, significativo estatisticamente. O teste do índice de Moran sobre os resíduos não rejeitou a hipótese nula de distribuição aleatória no espaço (p -valor = ~ 0.99), de forma que o modelo aparenta ter capturado com eficácia a autocorrelação espacial presente nos dados. Os únicos modelos que também passaram neste teste foram o SEM e o SAR. O teste de razão de verossimilhança rejeitou a necessidade de restringir o SDEM para um SEM, SLX ou MQO.

4.2.3 Discussão

Os resultados encontrados foram, de certa forma, heterogêneos. Nos dois modelos estimados, há associação entre a ampliação do serviço de coleta de esgoto e melhorias no rendimento escolar, embora no caso do Saeb esta interação ocorra via instalação de sistemas de saneamento, enquanto para a distorção idade-série a externalidade existe para a ampliação da infraestrutura.

A literatura internacional sobre a relação do saneamento com as notas de estudantes é escassa, o que dificulta traçar paralelos entre o modelo do Saeb e evidências empíricas. Contudo, tais resultados são consistentes com o modelo teórico de Glewwe e Miguel (2007), que argumentam que choques positivos no ambiente sanitário têm efeitos incertos (positivos ou negativos) sobre as habilidades acadêmicas das crianças, a depender de fatores como a função de utilidade dos pais, impacto marginal sobre os gastos com saúde, efeito renda e efeito preço. Já os resultados sobre a distorção idade-série têm correspondência com os trabalhos mencionados na seção 2, como Scriptorre (2016) e Agol et al. (2018), que reportam evidências para o Brasil e Zâmbia, respectivamente.

A seleção pelo modelo SDEM – que modela efeitos de transbordamento espacial nas variáveis explicativas (saneamento) e termo do erro (fatores não observáveis) –

pode encontrar respaldo no fato de que resíduos sólidos são levados por rios e correntes marítimas, afetando a população de cidades com sistema próprio de saneamento. Insetos vetores, como mosquitos e parasitas, também se deslocam além dos limites municipais, trazendo consigo doenças infecciosas. Esses são argumentos apresentados por Scriptori (2016). Ademais, a caracterização socioeconômica das cidades reforça esses transbordamentos ao segregar o espaço de acordo com o poder aquisitivo da população e acessibilidade à infraestrutura básica.

Os modelos estimados podem ter omitido variáveis relevantes para a determinação de indicadores educacionais, mas cujos dados inexitem ou não estavam disponíveis para o ano analisado. Uma evidência nesse sentido é o baixo valor para a estatística do R^2 ajustado nos modelos estimados por MQO (0,27 para o Saeb e 0,49 para a distorção). A seleção do SDEM também é respaldada por LeSage (2014), que argumenta que esta é a especificação mais comum para estudos empíricos na ciência regional.

No âmbito das variáveis de controle da distorção idade-série, Scriptori (2016) também encontra significância estatística e sinal positivo para o analfabetismo e média de alunos por turma, e negativo para a adequação docente. No entanto, obteve resultados significativos para a população negra e parda e renda per capita, o que não ocorreu neste artigo⁹.

5. Considerações finais

A associação entre saneamento básico e saúde pública, como a incidência de doenças infecciosas e morbidade hospitalar, é amplamente evidenciada na literatura. Uma face menos intuitiva, mas com fortes implicações para as políticas públicas, é sua relação com a qualidade da educação, tendo em vista as externalidades da ampliação do saneamento sobre o desenvolvimento cognitivo de crianças, aumento da frequência escolar e redução da desigualdade de gênero. Este trabalho utiliza dados do Inep, SNIS, PNSB, IBGE e RAIS para investigar a associação entre a cobertura de esgoto municipal e indicadores educacionais no contexto brasileiro, a saber: a taxa de distorção idade-série e a nota padronizada do Saeb. A análise foi realizada com dados de 2017, de modo a compatibilizar as informações de diferentes bases municipais.

A análise exploratória de dados espaciais, que visa identificar e quantificar a autocorrelação espacial no território, indicou a presença da dependência espacial positiva para as três variáveis de interesse e para os resíduos dos modelos de regressão por MQO. A distorção idade-série apresentou o maior valor do Índice de Moran global, da ordem de 0.727, seguido do Saeb (0.581). Em ambos os casos, pode-se observar *clusters* de indicadores educacionais favoráveis (altos para o Saeb, e baixos para a defasagem serial) no Centro-Sul, e resultados desfavoráveis nas regiões Norte, Nordeste e em grande porção do Mato Grosso.

⁹Cabe ressaltar que a autora utiliza dados do Censo Demográfico 2010, enquanto neste artigo os dados socioeconômicos provêm da RAIS, que inclui apenas informações de vínculos trabalhistas.

Todavia, mesmo nestes agrupamentos, identificam-se regiões que vão contra a tendência. O estado do Ceará, por exemplo, possui *clusters* de notas mais altas que a média nacional do Saeb e menores taxas de distorção idade-série; no estado do Rio de Janeiro, por outro lado, o alto patamar de distorção idade-série chama atenção. O Índice de Moran bivariado global denotou que municípios com notas municipais do Saeb acima da média nacional ou taxas de distorção abaixo dela tendem a ter como vizinhos municípios em que a cobertura de esgoto municipal se situa acima da média. Mais uma vez, o indicador bivariado local identifica agrupamentos favoráveis no âmbito das três variáveis de interesse em estados do Centro-Sul, e desfavoráveis no Norte, Nordeste e parte do Centro-Oeste, embora a falta de dados da cobertura de esgoto e Saeb possa ter prejudicado a estimação dos resultados locais.

Tendo em vista a presença de autocorrelação espacial nos resíduos, a estimação dos modelos econométricos ocorreu pela inserção de componentes de defasagem espacial (*lag*), de modo que foram avaliados os principais tipos de modelos de regressão espacial *cross-section*. A escolha final foi orientada pelos critérios de informação Akaike e BIC, valor da *log-verossimilhança* e teste de razão de verossimilhança, metodologia análoga à adotada por Scriptore (2016). Dessa forma, a especificação selecionada – tanto para o Saeb quanto para a distorção idade-série – foi o Modelo do Erro Espacial de Durbin (SDEM), que possui defasagem nas variáveis explicativas e no termo do erro, dando origem a efeitos de transbordamento espacial locais.

Se por um lado não foi encontrada associação significativa entre ampliação do saneamento existente e melhorias no Saeb, o modelo encontrou que a existência da infraestrutura, mesmo que insuficiente, se relaciona com um melhor rendimento das notas da Educação Básica. Já para a distorção idade-série, constata-se que, nas cidades onde a infraestrutura existe, a ampliação da cobertura de esgoto está relacionada a uma redução de cerca de 0.017% desse indicador, e a diminuição é ainda mais expressiva quando se eleva a cobertura de municípios vizinhos: da ordem de 0.043%, isolando-se o efeito das demais variáveis.

Em ambos os casos, o Índice de Moran denotou a eliminação da autocorrelação espacial dos resíduos, indicando que os modelos foram especificados adequadamente. Contudo, as estimativas encontradas podem ter sido enviesadas pela endogeneidade, isto é, a presença de autocorrelação das variáveis explicativas com o termo de erro. Esta, por sua vez, pode ser resultado da omissão de variáveis explicativas relevantes, visto que uma série de fatores socioeconômicos contribuem para a variabilidade dos indicadores educacionais. Ademais, os critérios de informação AIC, BIC e a *log-verossimilhança* dão suporte à especificação dentro de um conjunto de modelos estimados, mas não necessariamente refletem o modelo mais adequado dentro da miríade de modelos existentes.

Por fim, por resultarem de modelos de regressão linear, as estimativas encontradas demonstram apenas a associação entre saneamento e educação, e não uma relação causal entre os fatores. Futuros trabalhos sobre o tema devem considerar a estimação

com o uso de variáveis instrumentais, a fim de eliminar o viés da endogeneidade e aprofundar o conhecimento sobre as formas de interação entre o saneamento básico e o aproveitamento escolar. Não obstante, os resultados encontrados para a distorção idade-série indicam que iniciativas públicas eficazes para a ampliação da qualidade da educação devem ser coordenadas a nível regional, e não conduzidas de maneira isolada pelos municípios.

Referências

- Adukia, A. (2017). Sanitation and education. *American Economic Journal: Applied Economics*, 9(2):23–59. <https://doi.org/10.1257/app.20150083>.
- Agol, D., Harvey, P., e Maílló, J. (2018). Sanitation and water supply in schools and girls' educational progression in zambia. *Journal of Water Sanitation and Hygiene for Development*, 8(1):53–61. <https://doi.org/10.2166/washdev.2017.032>.
- Almeida, E. S. (2012). Econometria espacial aplicada. *Editora Alínea*, (1988).
- Arbia, G., Espa, G., e Giuliani, D. (2016). Dirty spatial econometrics. *Annals of Regional Science*, 56(1):177–189. <https://doi.org/10.1007/s00168-015-0726-5>.
- Banerjee, A. e Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. PublicAffairs, New York.
- Barufi, A. M., Haddad, E., e Paez, A. (2012). Infant mortality in brazil, 1980-2000: A spatial panel data analysis. *BMC Public Health*, 12(1). <https://doi.org/10.1186/1471-2458-12-181>.
- Baumont, C., Ertur, C., e Gallo, J. L. (2004). Spatial analysis of employment and population density: The case of the agglomeration of dijon 1999. *Geographical Analysis*, 36(2):146–176. <https://doi.org/10.1111/j.1538-4632.2004.tb01130.x>.
- Benova, L., Cumming, O., e Campbell, O. M. (2014). Systematic review and meta-analysis: Association between water and sanitation environment and maternal mortality. *Tropical Medicine and International Health*, 19(4):368–387. <https://doi.org/10.1111/tmi.12275>.
- Costa, D., Rego, A., de Souza, L., e Falcão, B. (2020). Associação entre taxa de mortalidade infantil e indicadores de habitação e saneamento: um guia para tomada de decisão em medicina preventiva e social. *Revista de Administração em Saúde*, 20(79):1–19. <https://doi.org/10.23973/ras.79>.
- da Cruz, F. P., de Oliveira, B. F., Accioly, E. M. F. B., e Soares, I. M. (2021). A regionalização dos serviços de saneamento básico e os desafios da universalização no brasil: Uma análise exploratória de dados espaciais para os anos de 2010 e 2018. In: *49ª Encontro Nacional de Economia (ANPEC)*.

- Dreibelbis, R., Greene, L. E., Freeman, M. C., Saboori, S., Chase, R. P., e Rheingans, R. (2013). Water, sanitation, and primary school attendance: A multi-level assessment of determinants of household-reported absence in kenya. *International Journal of Educational Development*, 33(5):457–465. <https://doi.org/10.1016/j.ijedudev.2012.07.002>.
- Glewwe, P. e Miguel, E. A. (2007). *The Impact of Child Health and Nutrition on Education in Less Developed Countries*, volume 4, Página 3561–3606. North Holland. [https://doi.org/10.1016/S1573-4471\(07\)04056-9](https://doi.org/10.1016/S1573-4471(07)04056-9).
- Hutton, G. (2013). Global costs and benefits of drinking-water supply and sanitation interventions to reach the mdg target and universal coverage. *Journal of Water and Health*, 11(1):01–12. <https://doi.org/10.2166/wh.2012.105>.
- INEP (2014). *Nota técnica nº 039 de 17 de dezembro de 2014. Indicador de Esforço Docente*. Brasília.
- Kopczewska, K. (2020). *Applied spatial statistics and econometrics: data analysis in R*. Routledge, Oxford, New York.
- Kothari, M. T., Coile, A., Huestis, A., Pullum, T., Garrett, D., e Engmann, C. (2019). Exploring associations between water, sanitation, and anemia through 47 nationally representative demographic and health surveys. *Annals of the New York Academy of Sciences*, 1450(1):249–267. <https://doi.org/10.1111/nyas.14109>.
- LeSage, J. e Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press, Boca Raton, Oxford, New York.
- LeSage, J. P. (2014). What regional scientists need to know about spatial econometrics. *Review of Regional Studies*, 44(1):13–32. <https://doi.org/10.52324/001c.8081>.
- Manski, C. F. (1993). Identification of endogenous social effects the reflection problem. *Review of Economic Studies*, 60(3):531–542. <https://doi.org/10.2307/2298123>.
- Ngure, F. M., Reid, B. M., Humphrey, J. H., Mbuya, M. N., Pelto, G., e Stoltzfus, R. J. (2014). Water, sanitation, and hygiene (wash), environmental enteropathy, nutrition, and early child development: Making the links. *Annals of the New York Academy of Sciences*, 1308(1):118–128. <https://doi.org/10.1111/nyas.12330>.
- Ortiz-Correa, J. S., Filho, M. R., e Dinar, A. (2016). Impact of access to water and sanitation services on educational attainment. *Water Resources and Economics*, 14:31,43. <https://doi.org/10.1016/j.wre.2015.11.002>.
- Sant’Anna, A. e Rocha, R. (2021). Saneamento e morbidade: evidências a partir de variações exógenas no tempo de execução dos investimentos. Texto para discussão, Instituto de Economia da UFRJ. Instituto de Economia da UFRJ. Texto para discussão 041.
- Sarrias, M. (2020). Notes on spatial econometrics.

- Scriptore, J. S. (2016). *Impactos do saneamento sobre saúde e educação: uma análise espacial*. Tese de Doutorado, Programa de Pós- Graduação do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo.
- Scriptore, J. S. e Azzoni, C. (2020). A spatial estimation of the impacts of sanitation on education. *Regional Science Policy and Practice*, 12(1):159–166. <https://doi.org/10.1111/rsp3.12136>.
- Silva, V. A. e Esperidião, F. (2017). Saneamento básico e seus impactos na mortalidade infantil e no desenvolvimento econômico da região nordeste. *Scientia Plena*, 13(10):01–17. <https://doi.org/10.14808/sci.plena.2017.109905>.
- Teixeira, J. C., Gomes, M. H. R., e Souza, J. A. D. (2012). Associação entre cobertura por serviços de saneamento e indicadores epidemiológicos nos países da América latina: Estudo com dados secundários. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health*, 32(6):419–425.
- Uhr, J. G. Z., Schmechel, M., e Uhr, D. D. A. P. (2016). Relação entre saneamento básico no Brasil e saúde da população sob a ótica das internações hospitalares por doenças de veiculação hídrica. *Revista de Administração, Contabilidade e Economia da Fundace*, 7(2):01–16. <https://doi.org/10.13059/racef.v7i2.104>.
- Ventura, G. K. A. e Lopes, F. W. A. (2017). Infraestrutura de saneamento básico e incidência de doenças associadas: uma análise comparativa entre Belo Horizonte e Ribeirão das Neves - Minas Gerais /sanitation infrastructure and associated diseases: a comparative analysis between Belo Horizonte (...). *Caderno de Geografia*, 27(51):788–805. <https://doi.org/10.5752/p.2318-2962.2017v27n51p788>.
- Verdiano, A. (2019). Um modelo explicativo do avanço na qualidade da educação do ensino fundamental no estado do Ceará. Dissertação de Mestrado, Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Modelagem e Métodos Quantitativos, Fortaleza.