

Diferença do desempenho educacional entre zonas rurais e urbanas: Uma análise quantílica incondicional

Italo Spinelli da Cruz¹
Fábio Rodrigues Moura²
Fernanda Esperidião³

RESUMO

O Brasil dispõe de um sólido sistema de avaliação e diagnóstico da educação básica e da qualidade do ensino que ajudam a compreender os diferenciais no desempenho do aluno nas escolas em áreas rurais e urbanas. As discussões recentes na literatura aplicada centram-se na perspectiva de que não a localização em si que afetaria as diferenças de resultados, mas a combinação de outros fatores como as características dos alunos, das famílias, dos professores, das escolas, dentre outras. Nesta perspectiva o objetivo deste trabalho é avaliar se a localização na zona rural é um fator significativo para explicar o desempenho inferior nesta área nos municípios brasileiros no período de 2013-2017. Para tanto, será avaliado o impacto distributivo da variável binária localização da escola no desempenho acadêmico mantida constante a distribuição das outras características, por meio da estimação de Regressões Quantílicas Incondicionais e o Tratamento de Efeito de Quantil tendo como base os microdados da Prova Brasil Saeb para o 5º ano do ensino fundamental. A análise descritiva indica a presença de distorções nas notas da prova de matemática, quando feita a distinção urbano-rural das escolas, mostrando sinais para todas as regiões que o desempenho do estudante da zona rural é inferior ao da zona urbana. Já análise dos resultados estimados pelos modelos de regressão do tratamento de efeitos de quantil para avaliação do diferencial de desempenho rural-urbano indicam de forma robusta a presença de uma penalidade aos alunos da escola rural mesmo com um controle das variáveis estabelecidas na literatura aplicada. Sendo assim, julga-se importante e plausível avaliar a evolução do *background* formativo do aluno, tanto para fins de investigações futuras, como para efeitos de avaliação e formulação de políticas públicas educacionais.

Palavras-chave: Desempenho educacional. Lacuna educacional rural-urbana. Tratamento de Efeito de Quantil.

ABSTRACT

Brazil has a solid system for diagnosis and assessment on quality of basic education which helps to understand the differences between student performance in schools located in rural and urban areas. Prior applied literature focuses on the fact that it is not the location itself that would affect the differences in results, but the combination of other factors such as the characteristics of students, families, teachers and schools. From this perspective, the aim of this study is to evaluate whether the location in rural areas is a significant factor to explain the gap in academic performance in Brazilian municipalities from 2013 to 2017. Drawing on microdata from Prova Brasil Saeb for students enrolled in the 5th grade of elementary school, the distributive impact of the binary variable school location on academic performance will be evaluated by estimating Unconditional Quantile Regressions and the Quantile Treatment Effect. Descriptive analysis suggests the presence of distortions on math test scores, when urban-rural distinction among

¹ Doutorando em Economia pelo Centro de Desenvolvimento e Planejamento Regional - Cedeplar/UFMG. E-mail: italospinelli@cedeplar.ufmg.br

² Prof.º Dr. do Departamento de Economia e do Programa Acadêmico de Pós-Graduação em Economia da Universidade Federal de Sergipe. E-mail: fabiomoura@gmail.com

³ Prof.ª Dr.ª do Departamento de Economia e do Programa Acadêmico de Pós-Graduação em Economia da Universidade Federal de Sergipe. Pesquisadora do Laboratório de Economia aplicada e Desenvolvimento Regional (LEADER – UFS). E-mail: nandaesper16@gmail.com.

schools is taken into consideration, which indicates that rural students' performance lag behind their urban peers. The analysis of the results estimated by the regression models of the treatment of quantile effects to evaluate the rural-urban gap in academic performance suggests the existence of a penalty for rural school students even when the variables established in the applied literature are controlled. Therefore, it is plausible to further investigate the evolution of a student's educational background, in order to evaluate and formulate public educational policies.

Keywords: Educational performance. Rural-urban educational gap. Quantile Treatment Effect.

Área: 17 – Desenvolvimento rural e local Código JEL: C21; I21; I24; I26

1 INTRODUÇÃO

Um dos ramos da economia da educação que tem motivado o desenvolvimento de novos estudos e pesquisas nos últimos anos e que ainda apresenta lacunas explicativas tem sido a análise dos fatores determinantes dos resultados do desempenho educacional dos estudantes. Com o avanço da aplicação de testes educacionais internacionais e nacionais padronizados e maior disponibilidade de informações estatísticas facilitou-se o processo de mensuração deste desempenho.

Neste sentido, o Brasil por meio do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep, dispõe do Sistema de Avaliação da Educação Básica - Saeb, que constitui um conjunto de avaliações externas e de larga escala, servindo como instrumento de diagnóstico da educação básica e da qualidade do ensino brasileiro. As médias de desempenho dos estudantes na avaliação do Saeb, bem como outros indicadores apurados no Censo Escolar, formam o Índice de Desenvolvimento da Educação Básica (Ideb), sendo um instrumento aliado para o desenvolvimento de políticas públicas educacionais.

Por sua vez, os resultados do Saeb ajudam a compreender os diferenciais no desempenho do aluno nas escolas em áreas rurais e urbanas. Avaliar os fatores que influenciam e explicam este diferencial ainda é um problema que merece investigação, em especial para o Brasil, país marcado por fortes disparidades urbanos-rurais em múltiplas esferas socioeconômicas e educacionais.

Apesar do avanço recente na literatura aplicada, poucos estudos têm examinado os impactos da localização rural ou urbano no desempenho final dos estudantes. As discussões centram-se na perspectiva de que não a localização em si que afetaria as diferenças de resultados, mas sim outros fatores como as características dos alunos, das famílias, das escolas, dentre outras. Neste sentido, a proposta deste estudo, consiste justamente no entendimento do quanto a variável localização continua a ser significativa quando o conjunto das outras variáveis situacionais estão sobre controle empírico.

Assim, questiona-se, dado um controle das múltiplas dimensões associadas ao processo do desempenho educacional admitidas na literatura: a lacuna de desempenho educacional entre a zona rural e a zona urbana pode ainda ser explicada de forma significativa pelo fator locacional?

Para tanto, assume-se como hipótese a ideia de que controladas as variáveis de características dos alunos, da família, dos professores, da infraestrutura da escola, o fator localização é significativo na determinação da lacuna de desempenho educacional entre as zonas urbanas e rurais. Portanto, espera-se que com o tratamento das regressões o efeito da localização da escola sobre o desempenho do aluno desapareça.

Sendo assim, o objetivo deste trabalho é avaliar se o fato de estar localizado na zona rural é um fator representativo para demonstrar desempenho inferior nesta área nos municípios brasileiros no período de 2013-2017. Para tanto, este trabalho avaliará o impacto distributivo

da variável binária localização da escola no desempenho acadêmico mantida constante a distribuição das outras características.

Nesta direção, além desta introdução artigo está dividido em cinco seções. Inicialmente é revisada a literatura que trata da perspectiva da função de produção da educação nos estudos econômicos, e dos estudos aplicados que avaliam os efeitos do diferencial rural-urbano do desempenho escolar. Na terceira seção é apresentada a estratégia empírica, o conjunto de variáveis e base de dados do modelo. Na quarta seção descreve-se as principais distorções nas notas da prova de matemática para as escolas urbanas-rurais, e os resultados estimados pelos modelos de regressão para avaliação do diferencial de desempenho. Por fim, na quinta seção são apresentadas as considerações finais. Seguido do arcabouço bibliográfico da literatura da economia da educação utilizado com base deste trabalho.

2 Revisão da literatura.

2.1 Função de produção da educação e avaliação do desempenho educacional.

Na literatura econômica a educação é vista como um importante fator determinante do bem-estar econômico, as discussões teóricas enfatizam o papel do capital humano no mercado de trabalho, na renda e no crescimento econômico. Para Hanushek (2020b) muito do que é desenvolvido na análise e nos estudos da Economia da Educação derivam de um modelo simples de produção, que relaciona insumos diversos, como recursos da escola, qualidade do professor, características da família dos estudantes, dentre outras, ao resultado do desempenho do aluno.

Todd e Wolpin (2003) destacam que os estudos empíricos sobre a FPE buscam compreender a combinação técnica dos insumos escolares e familiares na criação de desempenho cognitivo e de habilidades do indivíduo, maximizando seus resultados educacionais. A especificação preliminar da FPE pode ser feita como⁴:

$$Y_t = f(A_t, F_t, E_t) \quad (1)$$

onde Y_t corresponde ao rendimento escolar do aluno, que é uma função de três grupos diferenciados de fatores relacionados ao desempenho escolar: A_t representa as características individuais do estudante; F_t representa as características e *background* dos familiares, E_t representa as características escolares e dos professores.

A história da Função de Produção da Educação (FPE), como Hanushek (1987) descreve, está associada ao relatório *Equality of Educational Opportunity* elaborado por Coleman *et al* (1966) que avaliava a qualidade educacional por um conjunto de múltiplas variáveis como currículo oferecido, infraestrutura da escola, disponibilidade de livros, o uso de bibliotecas e laboratórios como práticas acadêmicas.

Para Hanushek (2020b) muito da atenção que a literatura dá às funções de produção da educação ocorre principalmente pelo reconhecimento de que as habilidades individuais têm significativos resultados no mercado de trabalho e na determinação dos salários. Por outro lado, aponta que um questionamento natural que deve ser feito é como desenvolver e aprimorar estas habilidades, e qual seria o papel da escola e da qualidade do processo formativo para potencializar o desenvolvimento destas competências.

Para Hanushek e Woessmann (2007) apenas a obtenção de graus de escolaridade e expansão do sucesso escolar não tem sido suficientes para assegurar melhoras nas condições

⁴KRUEGER (1999); LEE E BARRO (2001); HANUSHEK (1979); TODD E WOLPIN (2003).

socioeconômicas de regiões menos desenvolvidas, visto que múltiplos fatores influenciam no processo educacional, como por exemplo a relação familiar e com os colegas, aspectos de saúde e nutricionais que impactam nas competências cognitivas, dentre outros pontos.

Na literatura empírica há uma série de trabalhos dedicados a avaliar de forma específica os múltiplos fatores que afetam e diferenciam o desempenho escolar de crianças, que podem ser generalizados nos seguintes pontos: características familiares e socioeconômicas; características das escolas e dos professores; efeito das relações com os pares e ambiente escolar, dentre outras. O trabalho seminal de Hanushek (1971) é um dos pioneiros em analisar a eficiência do sistema escolar, avaliando se as características dos professores e da sala de aula eram fatores importantes para os ganhos de desempenho dos estudantes.

Boardman e Murnane (1979) propõem a construção de um modelo de dados em painel para promover melhorias nos instrumentos estatísticos de avaliação dos determinantes do sucesso educacional. Este trabalho ganha notoriedade por considerar em seu modelo que a educação é um processo acumulativo que implica que o sucesso dos estudantes será uma função de investimentos realizados em períodos anteriores.

Neste sentido, Chevalier e Lanot (2002) avaliam os efeitos relativos da situação financeira e das características das famílias no sucesso educacional para pessoas que vivem na Grã-Bretanha, levando em consideração variáveis como: educação dos pais, número de irmãos, local de moradia, renda familiar, dentre outras, associadas ao desempenho escolar, o estudo analisa uma distribuição de diferentes faixas etárias e em diferentes níveis de escolaridade para as duas pesquisas estabelecidas.

Häkkinen *et al.* (2003) analisam os efeitos das mudanças nos gastos escolares sobre as mudanças no desempenho dos alunos, para tanto utilizam uma amostra das notas dos exames dos alunos finlandeses do ensino médio nos anos de 1990-1998. Os resultados apontam que as mudanças nas despesas não tiveram efeitos significativos nas pontuações dos testes, os resultados apontavam ainda que o nível de educação dos pais eram tinham maior influência no desempenho dos alunos.

Rivkin, Hanushek e Kain (2005) avaliam o impacto das escolas e dos professores no desempenho escolar dos estudantes em amostra de alunos de escolas do Texas – EUA, para tanto estimam uma função de produção educacional com dados em painel. Os resultados apontam ainda que a redução do número de alunos em sala é menor do que os benefícios de elevar a qualidade do professor, destacando a eficácia do professor na determinação da qualidade da escola.

Martins e Veiga (2010) analisam se as desigualdades no nível educacional dos pais têm um papel relevante nas diferenças de desempenho dos estudantes no teste de matemática do PISA. Os resultados apontam que existe uma relação entre a desigualdade socioeconômica e o desempenho em matemática, sendo os grupos mais elevados os mais favorecidos, identificou-se ainda que as desigualdades são maiores na Alemanha, Grécia, Grã-Bretanha, Bélgica e Portugal, e menor na Suécia e Finlândia.

Woessmann (2016) tendo por objetivo avaliar a diferenças entre países nos resultados dos testes internacionais de desempenho educacional, avalia como as diferenças nos sistemas escolares, em especial suas estruturas institucionais, seriam responsáveis por uma parcela representativa destas diferenças de resultados no desempenho dos alunos.

Lavy (2020) realiza um experimento em escolas de Israel para investigar o efeito casual dos recursos instrucionais da escola, o número de horas que as crianças passam na escola semanalmente e a quantidade de tempo dedicado as disciplinas principais impactam no desempenho acadêmico e nos resultados não cognitivos de estudantes da quinta série em todas as escolas do país. As estimativas apontam ainda que um aumento do tempo de instrução de disciplinas específicas, levam a efeitos positivos e significativos nas pontuações de matemática, ciências e inglês.

2.2 Fatores determinantes no desempenho educacional: estudos aplicados ao caso brasileiro.

Barros *et al* (2001) investigam os determinantes do fraco desempenho educacional brasileiro considerando quatro tipos de fatores: impacto da disponibilidade e qualidade dos serviços educacionais; impacto da atratividade do mercado de trabalho local; disponibilidade de recursos (financeiros e não-financeiros) das famílias; e volume de recursos da comunidade do indivíduo. Os resultados da estratégia empírica econométrica adotada revelam que as características familiares se destacavam como de maior importância. Indicando ainda evidências de que mercados de trabalho atraentes estão relacionados à menor escolaridade.

Menezes-Filho (2007) avalia os determinantes do desempenho escolar dos estudantes brasileiros, para tanto analisa os dados do SAEB do ano 2003, considerando os alunos da 4ª e 8ª séries do ensino fundamental e da 3ª série do ensino médio nos testes de proficiência em Matemática. Os resultados estimados indicam que as variáveis que melhor explicam o desempenho escolar são as características familiares e do aluno, tais como nível educacional da mãe, raça (cor), atraso escolar e reprovação prévia, número de livros, presença de computador em casa e trabalho fora de casa.

Rodrigues *et al* (2011) buscam investigar o efeito das mudanças ocorridas nas características individuais, familiares e escolares dos alunos da 4ª série do ensino fundamental sobre as alterações ocorridas no desempenho escolar no período entre 1997 e 2005. Para tanto, consideram que as mudanças na proficiência escolar no período em análise estariam associadas a três fatores: mudanças na composição das características observáveis dos alunos e da escola; alterações nas taxas de retorno destas variáveis; e em mudanças na composição das características não observáveis dos alunos e escolas.

Franco e Menezes Filho (2017) analisam o impacto das características de alunos, professores, turmas, diretores e escolas sobre o desempenho em matemática de alunos da 4ª série, utilizando dados do SAEB 1997- 2005 e do Censo Escolar, controlando-se pelos efeitos não observáveis específicos de cada escola. Os resultados apontam que apenas algumas variáveis se mostram altamente significativas para explicar o desempenho escolar, independente do ano e da extensão do painel utilizado, sendo elas basicamente, as características dos alunos, o seu atraso escolar, e as condições socioeconômicas de suas famílias.

Raposo *et al* (2019) avaliam a disseminação dos efeitos de pares no desempenho escolar no âmbito das redes de amizade dos alunos. Os resultados apontam que um aumento do desvio-padrão na nota agregada de matemática de amigos, gera um aumento padronizado do rendimento individual em 6%, e quando os amigos indiretos são incluídos na análise, esse impacto aumenta para 45%. Os resultados apontam ainda que a influência do aproveitamento dos amigos é maior que o efeito da escolaridade dos pais.

Santos, Costa e Mariano (2019) buscam avaliar os efeitos da educação dos pais sobre o rendimento escolar de alunos do 5º ano do ensino fundamental da rede pública de ensino, em áreas urbanas do Brasil com base nos dados do SAEB, para o ano de 2015. Para tanto adotam um método de análise contrafactual de mediação causal que permite na perspectiva dos autores capturar a parcela do efeito da educação dos pais sobre o desempenho escolar dos filhos, associados ao nível socioeconômico da família, e o nível de escolaridade dos mesmos. Os resultados apontam que o efeito de mediação causal se potencializa a medida que o nível de educação dos pais aumenta, indicando que parcela do efeito da escolaridade dos pais é intensificado pelo nível socioeconômico familiar.

Ires, Mariano e Benevides (2020) criam um Índice de Motivação Multidimensional (IMM) considerando três dimensões: a motivação dos alunos, dos pais e dos professores, a partir

dos dados do PISA 2015 para verificar o efeito deste sobre o desempenho escolar e identificar qual das dimensões possui maior impacto na redução do diferencial de notas entre alunos com melhores e piores condições socioeconômicas. Para tanto, utilizam as metodologias de *Fuzzy Sets*, Balanceamento por Entropia e a decomposição de Oaxaca-Blinder. Os principais resultados observados indicam que o IMM foi positivo e significativo, apontando que um aumento médio de uma unidade deste indicador aumentaria a nota média em ciências, resultado semelhante foi identificado para as dimensões motivação dos alunos e dos professores.

Marioni, Freguglia e Menezes-Filho (2020) analisam se o capital humano do professor e as condições de trabalho impactam o desempenho do aluno, avaliando características do professor como educação, treinamento, renda familiar, experiência, estabilidade, outros empregos e horas trabalhadas. Os resultados demonstram que a jornada de trabalho do professor influencia positivamente no desempenho dos alunos nas provas de português e matemática. Por outro lado, os resultados indicam que professores que ensinam em mais de uma escola têm um impacto negativo nas pontuações dos testes de matemática. Indicando ainda que professores com mais de 15 anos de experiência na profissão têm um efeito positivo nas notas dos testes de português.

2.3 Diferenças no desempenho educacional urbano-rural.

Lounkaew (2013) busca explicar as diferenças urbano-rurais no sucesso educacional para estudantes da Tailândia avaliando os resultados do PISA 2009, para tanto utilizam um método de decomposição para demonstrar quanto a diferença urbano-rural do sucesso educacional era explicada por fatores características escolares não-mensuráveis. Os resultados demonstraram que o impacto das variáveis características da família do aluno e da escola variam ao longo das distribuições de aproveitamento do teste.

Ramos, Duque e Nieto (2012) examinam as diferenças no desempenho educacional entre estudantes das escolas nas zonas urbanas-rurais na Colômbia. Para tanto, utilizam os microdados de 2006 e 2009 do PISA avaliando os resultados obtidos em matemática, ciências e leitura, estimando uma função de produção educacional por meio da ampliação da metodologia desenvolvida em Juhn-Murphy-Pierce (1993) para analisar os efeitos da variação temporal nestas diferenças, que incluía variáveis de controle para características individuais, familiares e da escola, bem como a localização. Os resultados mostram que maior parte do diferencial de desempenho escolar urbano-rural está associada às características da família e não muito com as características da escola.

Gonzalez, Gómez e Gómez (2021) avaliam os determinantes do desempenho dos estudantes colombianos nos testes do PISA 2018, para alunos de 15 anos, para medir o nível de alfabetização em matemática, leitura e ciência, observando as lacunas rural-urbanas. Para tanto, inicialmente estimam uma regressão quantílica e método de decomposição de Oaxaca-Blinder para estimar quanto a diferença no desempenho entre os estudantes rurais-urbanos pode ser explicada por fatores como as características individuais, familiares e da escola. Os resultados apontam para diferenças médias significativas no desempenho dos alunos pela questão da localização urbano-rural, destacando dentre os fatores observáveis que as características da escola estavam entre os principais motivadores do *gap* de desempenho.

Sanfo e Ogawa (2021) adotam a decomposição da função de influência centralizada com dados do *Young Life* para estimar a função de produção da educação rural-urbana e decompor a lacuna de desempenho de aprendizagem rural-urbana na Etiópia. Os resultados indicam que a função de produção da educação rural-urbana é diferente entre as distribuições de desempenho. Os resultados apontam ainda que esta lacuna de desempenho rural-urbano é explicada pelas características do histórico do aluno, e as características não mensuráveis explicam proporcionalmente grande parcela desta lacuna.

Amini e Nivorozhkin (2015) analisam os determinantes do desempenho educacional rural-urbano para estudantes da Rússia. Para tanto, uma função de produção de educação é estimada com uma técnica de variável instrumental controlando a endogeneidade dos recursos escolares disponíveis por meio de um modelo multinível de dados hierárquicos. Os resultados das regressões estimadas indicam que melhores performances dos estudantes estariam associadas a um melhor nível socioeconômico. Já para o método de decomposição os resultados apontam que as diferenças urbano-rural no desempenho estudantil mais uma vez estariam associadas aos aspectos socioeconômicos da família que afeta os incentivos educacionais.

Rodrigues *et al* (2020) analisam os fatores que colaboram para determinação do desempenho escolar entre alunos em escolas da zona rural e urbana no Brasil. Para tanto, avaliam os resultados dos estudantes do 5º ano do Ensino Fundamental das escolas públicas, nas provas de Português e Matemática do exame Prova Brasil do ano de 2015. Os autores estimam regressões quantílicas incondicionais, decompondo o diferencial em seus determinantes com base na proposta de Firpo *et al* (2009) que estima uma *Regressão de Função de Influência Recentrada (RIF)* e no método de generalização da decomposição de Oaxaca (1973) aplicada aos quantis. Os resultados de estimação mostram que o efeito característica das escolas explica mais o diferencial urbano-rural que o efeito estrutural, observando ainda que as características das escolas frequentadas pelos alunos são os principais impulsionadores da diferença de notas, principalmente nos quantis superiores, sobretudo nas notas de Matemática.

Alencar *et al* (2021) analisam se as diferenças nas características dos professores explicariam a lacuna no desempenho educacional dos alunos do 9º ano do ensino fundamental entre escolas públicas nas zonas urbanas e rurais para o ano de 2017. Para tanto, definem uma Função de Produção Educacional e estimam regressões quantílicas incondicionais, por meio de uma combinação entre o método de decomposição de Oaxaca-Blinder e da *Regressão de Função de Influência Recentrada (RIF)* a fim de decompor os diferenciais nos diversos níveis de desempenho dos estudantes nos testes de Língua Portuguesa e Matemática da Prova Brasil-SAEB. Os resultados indicam que professores com licenciatura na área influenciam nos resultados de língua portuguesa e matemática de forma positiva e significativa, com maior efeito nas escolas rurais.

3 METODOLOGIA E ESPECIFICAÇÃO EMPÍRICA DO MODELO

3.1 Regressões Quantílicas Incondicionais e Efeitos de Tratamento de Quantis

A fim de avaliar os fatores determinantes do desempenho educacional entre zonas rurais e urbanas nos municípios brasileiros e avaliar a hipótese proposta neste estudo a estratégia empírica adotada é a de Regressão Quantílica Incondicional (“*Unconditional Quantile Regression – UQR*”), método introduzido por Firpo, Fortin e Lemieux (2009), que desenvolvem estimador para medir o impacto das mudanças na distribuição das variáveis explicativas (X), nos quantis marginais da variável resposta (Y). Para tanto, explicam que o método consiste na estimação de uma regressão de transformação, denominada Função de Influência Recentrada (“*Recentered Influence Function - RIF*”).

Como explica Rios-Ávila (2019) as regressões quantílicas incondicionais possibilitam ao pesquisador captar os efeitos parciais das variáveis explicativas em qualquer quantil incondicional da variável dependente.

A escolha por estimar Regressões Quantílicas Incondicionais se dá porque ela pode refletir a possível heterogeneidade nas relações estruturais entre o desempenho escolar e as covariáveis, pelo fato de considerar um número amplo de escolas dos municípios brasileiros, com múltiplas características diferenciadas e próprias e especialmente em zonas distintas, faz-se necessário controlar as características não-observáveis invariantes no tempo. Para tanto, a

Regressão Quantílica Incondicional a ser estimada com *dummies* de tempo e efeitos fixos individuais e comparação dos resultados com uma especificação de *pooling*.

De acordo com Firpo *et al* (2009) e Rios-Ávila (2019) a Regressão Quantílica Incondicional tem a vantagem de separar os efeitos de composição e estrutura da variável de interesse para qualquer estatística. O efeito de estrutura indica como a distribuição $F(Y/X)$ varia ao longo do tempo. As probabilidades de cada grupo obtidas no escore de propensão são reponderadas e usadas para estimar a regressão Função de Influência Recentrada (RIF). Esta regressão substitui a variável dependente pelo valor RIF estimado.

Desta forma, para Firpo *et al* (2009) e Rios-Ávila (2019) o modelo assume a estrutura geral:

$$RIF(y_i; v(F_y)) = X_i' \beta + \varepsilon_i. \quad (2)$$

Firpo, Fortin, and Lemieux (2009) e Rios-Avila (2019) mostram como identificar os efeitos parciais incondicionais de mudanças marginais nas covariáveis sobre a distribuição de v (por exemplo os quantis da variável de resposta). A regressão linear RIF-OLS usa o valor estimado $RIF(y_i; v(F_y))$ para cada observação y_i como uma variável de resposta de segunda etapa e faz a regressão contra o vetor covariável. Os efeitos parciais incondicionais podem ser mensurados calculando primeiro a expectativa incondicional da equação (1):

$$E[RIF(y_i; v(F_y))] = v(F_y) = E(X_i' \beta) + E(\varepsilon_i) = \bar{X}' \beta, \quad (3)$$

Assumindo $E(\varepsilon_i) = 0$. O efeito parcial incondicional para a variável independente x_k é dada por:

$$\frac{\partial v(F_y)}{\partial \bar{x}_k} = \beta_k \quad (4)$$

Assim, o coeficiente β_k expressa a mudança esperada na distribuição dv (no caso, algum quantil especificado) dado por uma mudança unitária na média incondicional de x_k .

De modo a avaliar os efeitos distributivos do tratamento (rural/ urbano), em outras palavras, comparar as distribuições de resultados no desempenho dos estudantes com a mesma distribuição de características, porém pertencentes a grupos diferentes, no caso deste trabalho, estudantes de escolas na zona rural e estudantes de escolas na zona urbana, a metodologia mais adequada para esta estimação, segundo Rios-Ávila e Maroto (2022) é a *Quantile Treatment Effects (QTE)*.

Rios-Ávila e Maroto (2022) apontam que as aproximações lineares das variações dos quantis incondicionais da variável dependente dada uma pequena mudança na distribuição das características das variáveis independentes. No entanto, Rios-Ávila e Maroto (2022) afirmam que tais aproximações poderiam ser inapropriadas ao se analisar grandes mudanças na distribuição das características, em especial, casos com faixa limitada de valores, ou quando a variável é binária. Caso que se aplica neste trabalho dado a categoria de comparação diferencial rural e urbano.

Rios-Ávila e Maroto (2022) consideram que o Tratamento de Efeitos dos Quantis (QTE) é uma harmonização (ajustamento) entre as Regressões Quantílicas Condicionais (CQR) e as Regressões Quantílicas Incondicionais (UQR) concentrando-se na análise do impacto da distribuição de uma única variável binária sobre o resultado de interesse mantendo-se constante a distribuição das outras possíveis características explicativas do modelo, desenvolvida a partir da metodologia de Firpo (2007) e Firpo e Pinto(2016)

Para este estudo no qual a localização da escola é medida como uma variável binária (x) e representa duas categorias, rural ou urbano. E assumindo que o desempenho na prova de matemática, enfrentam dois resultados potenciais, y_0 ou y_1 , dependendo da localização da escola frequentada pelos alunos, observaremos como em Rios-Ávila e Maroto (2022) apenas y :

$$y = y_1x + y_0(1 - x) \quad (5)$$

Nesta perspectiva, se é possível observar o potencial de desempenho de ambos, os tratamentos de efeitos dos quantis podem ser estimados como em Rios-Ávila e Maroto (2022), calculando a diferença nos quantis de ordem τ entre a distribuição de desempenho observada do rural e urbano:

$$\Delta_\tau = Q_\tau(y|x = 1) - Q_\tau(y|x = 0) \quad (6)$$

Como Rios-Ávila e Maroto (2022) explicam isto se torna possível, sob a suposição de atribuição aleatória, as distribuições observadas do desempenho entre o grupo tratado (ou não tratado) $F_{y|x=1}(F_{y|x=0})$ é como a mesma distribuição dos potenciais desempenhos $F_{y1}(F_{y0})$. Neste quadro, o tratamento dos efeitos pode ser estimado adotando a Regressão Quantílica Condicional (CQR), e a distribuição observada do desempenho, estimando a equação (6c):

$$Q_\tau(y|x) = Q(y|x = 1) * x_i + Q(y|x = 0) * (1 - x_i) \quad (6a)$$

$$Q_\tau(y|x) = Q(y|x = 0) + Q(y|x = 1) - Q(y|x = 0)x_i \quad (6b)$$

$$Q_\tau(y|x) = b_0(\tau) + b_x(\tau) x_i \quad (6c)$$

Rios-Ávila e Maroto (2022) apresentam uma estratégia alternativa para estimar o tratamento de efeito dos quantis utilizando uma extensão das regressões de influência recentrada (RIF). Eles estimam a equação (6c) via RIF na tentativa de aproximar o quantil condicional de y , para tanto, partindo de (6b) eles reordenam os termos do lado direito da equação, obtendo:

$$Q(y|x = 1)x_i + Q(y|x = 0)(1 - x_i) \quad (7)$$

A partir da equação (7), obtém-se a RIF, utilizando-a como variável dependente, em uma regressão similar a (6c) porém estimada por meio da OLS:

$$RIF(y_i, Q_\tau(\cdot), F_{y|x=1})x_i + RIF(y_i, Q_\tau(\cdot), F_{y|x=0})(1 - x_i) = b_0(\tau) + b_x(\tau)x_i + e_i \quad (8)$$

Considerando que neste estudo a condição do estudante estar matriculado em uma escola na zona urbana ou na zona rural não é um fator aleatório, torna-se necessário a realização de um procedimento de remodelação do modelo via construção de pesos de probabilidade inversa (IPW).

De acordo com Rios-Ávila e Maroto (2022), Firpo (2007) e Firpo e Pinto (2016), se a atribuição de tratamento não for aleatória, ainda que dependa somente de características observadas z (condicionalmente exógena) as estratégias de estimação apresentadas não identificarão corretamente os efeitos do tratamento. Isso ocorre porque as diferenças nos quantis capturados por $b_x(\tau)$ também podem explicar diferenças na distribuição de características entre o grupo tratado e não tratado, e a distribuição do resultado potencial será diferente da distribuição do resultado observado entre os tratados.

Desta forma, Rios-Ávila e Maroto (2022), Firpo (2007) e Firpo e Pinto (2016) sugerem que os efeitos do tratamento distribucional podem ser encontrados por meio de um procedimento de duas etapas. Na etapa inicial constrói-se um conjunto de pesos para igualar a distribuição observada de características z entre o grupo tratado e os grupos não tratados. Para tanto, estima-se inicialmente um escore de propensão $\hat{p}(z)$ que representa a probabilidade de uma observação fazer parte do grupo tratado, condicionada a z . Em seguida constrói-se os pesos de probabilidade inversa (IPW), da seguinte forma:

$$\omega_0 = \frac{1-x}{1-\hat{p}(z)} \text{ e } \omega_1 = \frac{x}{\hat{p}(z)} \quad (9)$$

Na segunda etapa, os pesos construídos são utilizados para construir a apropriada função reponderada RIF e estimar a equação (8) usando regressões RIF por OLS (FIRPO et al, 2018). De modo geral, como afirma Rios-Ávila e Maroto (2022), Firpo (2007) e Firpo e Pinto (2016) os pesos de probabilidade inversa (IPW) são adotados para remodelar a distribuição observada das características z , de modo que os grupos tratados (ou não tratados) se aproximem a toda população. Ao fazê-lo, a distribuição ponderada do resultado observado entre os tratados (ou não tratados) será similar à distribuição do resultado potencial sob tratamento para a amostra completa. Esta estimativa das distribuições de resultados potenciais pode então ser usada para identificar os efeitos de tratamento dos quantis.

A interpretação das estimativas de QTE para o nosso caso, tendo como variável de tratamento a localização da escola, pode ser avaliada ao comparar o nível da diferença nas distribuições de desempenho na prova de matemática ao comparar os resultados dos estudantes da zona urbana, com estudantes matriculados em escolas da zona rural, mantido todo o resto constante (efeito médio do tratamento).

Assim, como o RIF é construído para uma estatística de distribuição condicional a x , β_z não captura mais o efeito incondicional de z no τ -ésimo quantil de desempenho entre todos os estudantes. Em vez disso, ele captura a média ponderada dos efeitos incondicionais sobre os quantis de desempenho nos grupos de condicionamento, neste caso rural e urbano. Na prática, entretanto, tanto o UQR quanto o QTE via RIF fornecem estimativas de coeficientes semelhantes para variáveis não condicionantes.

Neste sentido, a equação do modelo estimado para avaliar o diferencial do efeito rural-urbano no desempenho escolar é especificado pela regressão de influência recentrada (RIF) com tratamento dos efeitos dos quantis (QTE) seguinte:

$$RIF(y_i, Q_\tau(.)) = \beta_0(\tau) + \delta(\tau)local_i + \beta_1(\tau)I_i + \beta_2(\tau)F_i + \beta_3(\tau)P_i + \beta_4(\tau)E_i + \eta_m + \varepsilon_{\tau i} \quad (10)$$

em que $RIF(y_i, Q_{\tau i})$ é a RIF do quantil incondicional τ da distribuição da pontuação da proficiência do aluno no teste de matemática na Prova Saeb; $\beta_{i\tau}$ é o vetor de coeficientes no quantil τ associados a cada vetor de características incluídas na regressão; $local$ é a *dummy* que identifica a localização da escola na zona rural ou urbana; I é o vetor de características individuais do estudante; F é o vetor de características do domicílio e *background* familiar do estudante; P é o vetor de características dos professores; E é o vetor de características da escola; η_m é o termo dos efeitos fixos a nível municipal; e ε_τ é o termo de erro no quantil incondicional τ .

3.2 Base de dados e descrição das variáveis.

Para o desenvolvimento da análise proposta foram utilizados os Microdados do Sistema de Avaliação da Educação Básica - Prova Saeb, dos anos com realização da prova no período de 2013 a 2017, para tanto são considerados os resultados dos estudantes do 5º ano do ensino fundamental na prova de matemática como indicador do desempenho escolar, diferenciados por grupos de zoneamento urbano ou rural, para todos os municípios brasileiros com resultados disponíveis.

Os testes do Saeb são realizados com o objetivo de mensurar os níveis de aprendizagem dos estudantes avaliados, sendo um dos principais instrumentos indicativos da qualidade do ensino básico brasileiro, possibilitando o monitoramento e subsídios para elaboração e aperfeiçoamento de políticas educacionais com base em dados.

Os testes dos Saeb são elaborados a partir de matrizes de referência, que avaliam as competências e habilidades que os estudantes deveriam ter desenvolvido na etapa da educação básica avaliada.

No caso deste trabalho decidiu-se avaliar os diferenciais de desempenho na proficiência do teste de matemática, dos alunos da rede pública no 5º ano do ensino fundamental. As escalas de proficiência do Saeb detalham o que será avaliado e os níveis de habilidades apresentados pelos estudantes. Esta classificação é importante para entendermos os resultados do desempenho dos estudantes não apenas na classificação rural-urbano, visto que a variável dependente do modelo desenvolvido neste trabalho é a proficiência do aluno em matemática, mas também por permitir compreender a qualidade do ensino-aprendizagem de matemática na educação básica brasileira de maneira geral.

Do ponto de vista do modelo econométrico, as variáveis de controle selecionadas representam quatro grupos de categorias que a literatura aponta como fatores importantes para o desempenho escolar: i) características dos estudantes; ii) grau de conhecimento (*background*) dos familiares; iii) características dos professores; iv) características das escolas. Elas estão descritas no Quadro 3.1.

QUADRO 3.1 – Descrição das Variáveis de Controle

(continua)

Grupo de Variáveis	Variável
Variável Dependente	Proficiência do aluno em Matemática.
Variável Independente	Localização da escola (0 – Urbana; 1 – Rural)
Variáveis de Controle	
Características dos Estudantes	Idade
	Sexo (1 – Masculino; 0 – Feminino)
	Raça (1- Branco; 0 – Não-branco)
	Reprovação (1- Nunca reprovou; 0 – Caso contrário)
	Abandono escolar (1- Nunca abandonou; 0 – Caso contrário)
	Distorção idade/série (Idade correta para a série (1 – Sim; 0 – Não))
	Hábitos de leitura (1 – ler frequentemente; 0 – Não)
	Hábitos culturais (1 – Tem algum hábito cultural; 0 – Caso contrário)
	Sempre faz os exercícios de casa de Matemática? (1 – Sim; 0 – Não)
	Tem um trabalho? (1 – Sim; 0 – Não)
Características do Domicílio e Background Familiar	Tarefas domésticas (1 – Sim; 0 – Não)
	Televisão em casa (1 – Sim; 0 – Não)
	Geladeira em casa (1 – Sim; 0 – Não)
	Máquina de lavar em casa (1 – Sim; 0 – Não)
	Família tem carro? (1 – Sim; 0 – Não)
	Computador em casa (1 – Sim; 0 – Não)
	Banheiro em casa (1 – Sim; 0 – Não)
	Quartos em casa (1 – Sim; 0 – Não)
	Quantas pessoas moram na casa? (Escala de 1 a 6 pessoas ou mais).
	Empregado(a) doméstico(a) na casa (1- Sim; 0 – não)
	Pai analfabeto ou ensino fundamental incompleto (1- Sim; 0 – não)
	Pai com ensino fundamental (1- Sim; 0 – não)
	Pai com ensino médio (1- Sim; 0 – não)
	Pai com ensino superior completo (1- Sim; 0 – não)
	Mãe analfabeta ou ensino fundamental incompleto (1- Sim; 0 – não)
	Mãe com ensino fundamental (1- Sim; 0 – não)
	Mãe com ensino médio (1- Sim; 0 – não)
	Mãe com ensino superior (1- Sim; 0 – não)
	Os pais (ambos) moram em casa (1-Sim; 0- Não)
	Os pais incentivam ir para escola e a realizar as atividades escolares (1-Sim; 0- Não)
Características dos Professores	Sexo
	Faixa etária
	Maior nível de escolaridade
	Área temática do curso de pós-graduação de mais alta titulação
	Salário bruto como professor
	Exercício de atividade externa que contribui para renda (1- Sim; 0 – não)
	Anos de trabalho como professor
	Anos de trabalho como professor na escola avaliada
	Anos de atuação na série/ turma atual
	Participação de atividade de aperfeiçoamento profissional nos últimos dois anos (1- Sim; 0 – não)

	Hábitos de leitura (1 – ler frequentemente; 0 – Não)
	Hábitos culturais (1 – Tem algum hábito cultural; 0 – Caso contrário)
	Proporção de professores com diploma superior
	Proporção de professores com mais de 6 anos de experiência
	Proporção de professores que ganham mais que 4 salários-mínimos correntes.
Características da Escola	Responsabilidade Administrativa (1 – Município; 0 – Estado)
	Número de alunos matriculados
	Relação aluno/ professor
	Tempo médio das aulas
	Falta de material escolar (1 – Sim; 0 – Não)
	Como o diretor é selecionado? (1 – Nomeado/ Eleito; 0 – Caso contrário.
	Número de professores é insuficiente (1 – Sim; 0 – Não)
	Taxa de Regularidade dos Professores
	Índice de Infraestrutura da Escola
	Indicador de Má Manutenção da Escola
	Nível socioeconômico da escola

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

4.1 Estatísticas descritivas do desempenho das escolas urbanas e rurais

A partir dos dados do Saeb apresenta-se um breve retrato do contexto educacional dos alunos do 5º ano do ensino fundamental brasileiro. Observa-se que em 2017, 13,5% dos alunos que frequentavam uma escola na zona rural, representando 354.588 alunos. Já, as escolas da zona urbana concentravam um quantitativo de 2.269.431 dos estudantes, uma proporção de 86,5%.

Uma característica importante observada refere-se a faixa etária do aluno, esta variável é importante por retratar os efeitos de potencial distorção idade-série, ou seja, estudantes que estejam fora da série adequada para a idade. O conjunto de dados revela que no Brasil, apenas 38,2% dos alunos da zona urbana e 33,4% na zona rural estavam na faixa de idade adequada para 5º ano.

Em 2017, grande parcela dos estudantes da zona rural, 72%, não tinha disponível em suas casas um computador, o proporcional de estudantes da zona urbana apesar de menor, também mostrava que o bem ainda não era acessado por 44% dos domicílios.

Os dados indicam que no grupo das mães ou mulheres responsáveis temos uma menor proporção de pessoas que não sabem ler ou escrever, sendo que na zona urbana apenas 3,5% eram analfabetas, já na zona rural este percentual era de 8,9%. Por outro lado, no grupo dos pais ou homens responsáveis esta proporção era maior, na zona urbana 7,4% deste grupo eram analfabetos, já na zona rural este proporcional representava aproximadamente 16%, um valor relativamente elevado.

Em relação as características dos professores, dados importantes a respeito do grau de formação docente revelam que nas escolas da zona rural 13,8% não possuíam diploma de nível superior em 2017, mostrando uma evolução em comparação a 2013, quando tínhamos aproximadamente 21% dos professores sem este grau de instrução. Outro ponto relevante está relacionado a remuneração do corpo docente, tanto na zona urbana quanto na rural uma elevada proporção recebe até 4 salários-mínimos na função, em 2017, este percentual era 88% dos professores da zona rural e 77% nas escolas urbanas.

Por fim, dentre algumas características das escolas, chama atenção a proporção de escolas rurais que não tem computadores para uso dos alunos, em 2017 este valor representava 49,3% das escolas. Complementar a este fato, 63% das não tinham acesso à internet.

Ao avaliar o diferencial de desempenho educacional urbano-rural no Brasil observa-se que as notas médias apresentaram uma evolução positiva, tanto na zona urbana, quanto na zona rural. Comparando, os resultados de 2013 e 2017, observa-se na Tabela 4.1 um aumento em

torno de 5% na nota média da zona urbana, já na zona rural este aumento de desempenho foi de aproximadamente 9,8%.

No entanto, os resultados mostram-se ainda bastante desiguais entre as duas localizações, se em 2013, a nota média para o Brasil na zona urbana foi de 210,06 pontos, o desempenho médio na zona rural foi quase 30 pontos a menos, com um resultado médio de 181,71 pontos. Ainda que tenha apresentado melhoras, em 2017 a nota média da zona rural (199,59) foi consideravelmente inferior a zona urbana (222,84), com uma diferença de 23 pontos.

As figuras 4.1, 4.2 e 4.3 mostram a distribuição das notas Escala de Proficiência de Matemática, respectivamente para os anos de 2013, 2015 e 2017. Os resultados mostram-se preocupantes acerca da lacuna de desempenho escolar dada a localização da escola. Em 2013, como pode ser observado na figura 4.1, aproximadamente 25,8% dos estudantes de escolas urbanas apresentaram um desempenho abaixo ou no nível 03 de proficiência, por outro lado, na zona rural os resultados revelam o contraste educacional na zona rural, dado que 48,45 dos estudantes desta localização apresentaram um desempenho inferior ou igual ao nível 03 de proficiência.

Tabela 4.1 – Proficiência média e desvio padrão na nota de Matemática por localização e Regiões – 2013 - 2017

Região	Ano e Localização					
	2013		2015		2017	
	Urbano	Rural	Urbano	Rural	Urbano	Rural
Brasil	210,96 (0,0378)	181,71 (0,1098)	218,66 (0,0334)	196,15 (0,0940)	222,84 (0,0339)	199,59 (0,0854)
Norte	193,32 (0,1043)	166,76 (0,2171)	203,75 (0,0912)	184,29 (0,1739)	208,92 (0,0981)	184,25 (0,1673)
Nordeste	188,33 (0,0697)	174,89 (0,1382)	200,98 (0,0624)	192,03 (0,1207)	205,39 (0,0662)	195,20 (0,1123)
Sudeste	223,14 (0,0598)	213,22 (0,3065)	229,32 (0,0525)	219,07 (0,2883)	232,94 (0,0517)	223,38 (0,2347)
Sul	225,68 (0,0905)	213,36 (0,4162)	229,11 (0,0821)	220,72 (0,3614)	233,53 (0,0833)	224,77 (0,2886)
Centro-Oeste	215,95 (0,1144)	197,83 (0,5044)	217,42 (0,1043)	205,69 (0,4703)	223,13 (0,1090)	207,76 (0,3915)

Fonte: Elaboração própria, com base nos dados do Saeb 2013, 2015 e 2017.

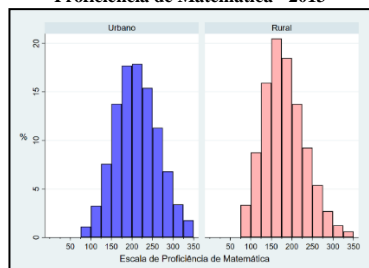
Outra diferença preocupante é o percentual de alunos que tiveram um resultado abaixo dos 125 pontos, classificados no nível 0 de proficiência, na zona urbana este percentual foi de aproximadamente 4,3%, contra aproximadamente 12% na zona rural.

As figuras 4.2 e 4.3 revelam um comportamento similar de distribuição dos resultados pela escala de proficiência matemática, nos anos de 2015 e 2017, ao observar o histograma os resultados para os alunos de escolas urbanas tendem a apresentar uma distribuição mais normalizada em torno da nota média com um considerável percentual de estudantes próximos dos valores do nível 04 e 05.

Já para a zona rural claramente percebe-se uma distorção para cauda direita da distribuição, com maior predominância dos valores mais baixos de notas. Em 2015, o percentual de alunos da zona rural que tiraram notas abaixo ou igual ao nível 03 de proficiência representava aproximadamente 60% dos alunos, no sentido oposto 62% dos alunos da zona urbana apresentaram um resultado maior ou igual ao nível 04 de proficiência.

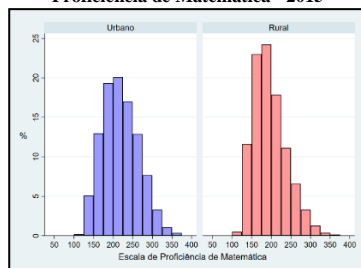
Em 2017, observa-se uma melhora muito sutil no percentual de alunos da zona rural no nível inferior ou igual a 03, representando aproximadamente 54,8% dos alunos. No entanto, estes resultados retratam as disparidades entre o processo de ensino-aprendizagem e o desempenho dos alunos em resolver problemas básicos da matemática.

Figura 4.1 – Distribuição das notas - Escala em Proficiência de Matemática - 2013



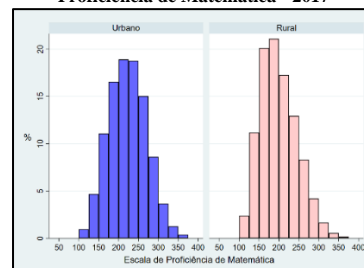
Fonte: Elaboração Própria. Microdados do Saeb 2013.

Figura 4.2 – Distribuição das notas - Escala em Proficiência de Matemática - 2015



Fonte: Elaboração Própria. Microdados do Saeb 2015.

Figura 4.3 – Distribuição das notas - Escala em Proficiência de Matemática - 2017



Fonte: Elaboração Própria. Microdados do Saeb 2017.

Ao avaliar os resultados nas esferas regionais, pode-se observar que estas diferenças de desempenho rural-urbano são potencializadas nas regiões tradicionalmente marcadas por maiores níveis de atraso socioeconômico.

Na região Norte, podemos observar que a evolução da nota média entre 2013-2017 na zona urbana foi considerável, se em 2013 a nota média foi de 193,32 pontos, no ano de 2017 a pontuação média foi 208,92. Por outro lado, na zona rural, apesar de uma relativa evolução nas notas médias quando comparados ao ano de 2013 (166,76 pontos), os resultados dos anos 2015 (184,29 pontos) e 2017 (184,25) ficaram estagnados, sendo a única região que não apresentou uma melhora na nota média para a zona rural no período.

Ao avaliar o desempenho médio em termos dos níveis da escala de proficiência de matemática os resultados revelam-se de maneira geral insatisfatórios em todas as regiões. Nos três períodos avaliados o desempenho médio das escolas rurais no Brasil estava na faixa de valores classificados no nível 03 de proficiência em matemática. No caso do urbano apesar de apresentar um desempenho levemente superior, o resultado médio encontrava-se situado no nível 04 de proficiência, numa escala que varia de 0 a 9. Na perspectiva regional, o Norte, o Nordeste e o Centro-Oeste continuam apresentando maiores defasagens em relação ao Sul e o Sudeste.

Em 2017, observa-se que no Norte e no Nordeste o desempenho médio das escolas rurais foi de 184,25 e 195,20 pontos respectivamente, classificadas no nível 03 de proficiência, enquanto no Sul e Sudeste o resultado médio para a zona rural já estavam em um patamar na fronteira entre o nível 04 e 05 de desempenho.

Para o Brasil, o diferencial percentual da média de notas da localização rural e urbana foi de -14% em 2013, e apesar da redução da diferença para os anos 2015 e 2017, este percentual era de -10% para a zona rural. Os resultados em termos percentuais para alguns estados no ano de 2017, superam o valor de 10% em termos de defasagem negativa da nota em alguns estados, a saber Mato Grosso do Sul (-10,7%), Tocantins (-12,6%), Roraima (-15%), Amazonas (-16,3%).

Diante deste cenário, a continuidade deste capítulo avalia a partir do tratamento empírico, controlando as características que a literatura aponta como fatores que implicam no sucesso acadêmico combinado à localização da escola, se há ou não uma persistência do diferencial negativo dos resultados acadêmicos dos estudantes da zona rural em relação ao estudante da zona urbana.

Ao aplicar o tratamento empírico via Efeito de Tratamento de Quantis e controle das características preconizadas na literatura espera-se que esta diferença percentual entre a desempenho dos estudantes dada a localização da escola deixe de existir.

4.2 Resultados das estimativas dos Efeitos de Tratamento dos Quantis

Tendo como principal objetivo isolar e avaliar o impacto distributivo da variável localização da escola no desempenho acadêmico mantida constante a distribuição das outras características, a estratégia empírica adotada neste trabalho tem como base a estimação de Regressões Quantílicas Incondicionais (Firpo *et al* 2009; Rios-Ávila 2019) com o auxílio do Efeitos de Tratamento de Quantis (Rios-Ávila e Maroto (2022)) considerando efeitos fixos para os municípios brasileiros a fim de observar a possibilidade de uma penalidade no desempenho acadêmico associado ao fato de estar na zona rural.

Os primeiros resultados estimados mostram a distribuição do desempenho dos estudantes baseados na sua condição de aluno da zona rural ou urbana, observa-se de maneira considerável um efeito negativo de estar numa escola rural, em todos os níveis da distribuição dos quantis. Os coeficientes do tratamento de efeito de quantil para o diferencial rural-urbano foram estimados por meio de modelos que adotam as seguintes estratégias: i) modelo com todos os controles e sem interação entre a variável localização da escola; ii) modelo com todos os controles e com interação entre a variável localização da escola; iii) modelos com tratamento de IPW; iv) modelos em *pooled* com *dummy* para os anos e interação com o tempo.

O primeiro conjunto de resultados refere-se às estimativas do QTE sem considerar a interação da *dummy* localização da escola com as variáveis de controle. Na Tabela 4.2, observa-se de modo geral a distribuição dos efeitos em cada período, a presença de uma penalidade aos estudantes da zona rural. No ano de 2013, evidenciou-se com alto nível de significância uma penalidade nas notas dos alunos da zona rural, associada ao fato de estar nesta localização, nos quantis 50, 75 e 90 há uma penalidade em torno de 2,1-2,3% na pontuação média do aluno.

Resultados similares são acompanhados nos anos 2015 e 2017, no mesmo intervalo quantílico, no entanto, observa-se que em 2015 há uma redução na penalidade para o aluno da zona rural no q50, que passa a ser próxima de 1%. Por outro lado, no ano de 2017 há um aumento considerável nesta penalidade no q50, onde a diferença negativa da pontuação do aluno da zona rural é de 2,3%, com uma redução no impacto no q90 que foi em torno de 1,3% a menos para o estudante da zona rural. Ainda na tabela 4.2, pode-se avaliar os resultados da estimação do *pooled* para os períodos 2013-2017, ao aplicar as *dummies* de tempo, e a interação com esta variável, os resultados mostram-se similares, o simples fato de estar localizado em uma escola de zona rural gera uma penalidade ao longo de praticamente toda distribuição quantílica.

Tabela 4.2 – QTE com todos os controles e sem interação

Quantil	QTE sem interação			
	2013	2015	2017	2013-2017
	<i>n. obs:</i> 751,557	<i>n. obs:</i> 1,003,512	<i>n. obs:</i> 898,526	<i>n. obs:</i> 2,653,619
	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$
5	-.01431*** (.00382)	.00558*** (.00202)	.00217 (.0023)	-.00146 (.00144)
10	-.01532*** (.00281)	.001625 (.00163)	-.00357** (.00177)	-.00388*** (.00177)
25	-.01686*** (.00189)	-.00559*** (.00128)	-.01286*** (.00132)	-.01147*** (.00080)
50	-.02152*** (.00159)	-.01079*** (.00115)	-.02301*** (.00112)	-.01868*** (.00070)
75	-.02306*** (.00165)	-.02301*** (.00123)	-.01915*** (.00108)	-.02181*** (.00071)
90	-.02105*** (.00188)	-.02018*** (.00142)	-.01381*** (.00124)	-.01704*** (.00080)
95	-.01639*** (.00213)	-.01580*** (.00163)	-.01044*** (.00142)	-.01470*** (.00091)

Notas: ***significativo a 1%; **significativo a 5%; *significativo a 10%.

Fonte: Elaboração própria a partir do *software* Stata 15.

Nestas primeiras estimativas sem interação, percebe-se que nos quantis inferiores a diferença da penalidade por estar numa escola da zona rural, é menor, que nos quantis superiores, e no caso do ano 2017 e no *pooled* o quantil mais baixo, o q5, foram não significativos. Neste sentido, quando estimados os tratamentos de efeitos de quantis com a interação da variável *dummy* localização da escola com as demais variáveis de controle, os resultados apontam um novo direcionamento da penalidade no desempenho do aluno que estuda em uma escola da zona rural, e pode ser visto na tabela 4.3. Diferente do primeiro conjunto de resultados iniciais, ao interagir a localização com outras características, a penalidade de estar na zona rural é suavizada e deixam de ser significativas em grande parte das estimativas para os quantis mais elevados (90 e 95).

Tabela 4.3 – QTE com todos os controles, com interação

Quantil	QTE com interação			
	2013	2015	2017	2013-2017
	<i>n. obs:</i> 751,557	<i>n. obs:</i> 1,003,512	<i>n. obs:</i> 898,526	<i>n. obs:</i> 2,653,619
	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$
5	-.06009*** (.00438)	-.02862*** (.00229)	-.03105*** (.00258)	-.03028*** (.00167)
10	-.04471*** (.00352)	-.02941*** (.00191)	-.03044*** (.00209)	-.03064*** (.00129)
25	-.03524*** (.00256)	-.02464*** (.00164)	-.02440*** (.00173)	-.02514*** (.00106)
50	-.02075*** (.00244)	-.01373*** (.00168)	-.01457*** (.00168)	-.01520*** (.00104)
75	-.00813*** (.00282)	-.00389*** (.00204)	-.00123 (.00176)	-.00358*** (.00119)
90	.00050 (.00348)	.00110 (.00255)	.00204 (.00216)	-.000619 (.00143)
95	.00879** (.00411)	.00016 (.00296)	.00161 (.00253)	.00077 (.00169)

Notas: ***significativo a 1%; **significativo a 5%; *significativo a 10%. %.

Fonte: Elaboração própria a partir do *software* Stata 15.

No entanto, os resultados dos quantis inferiores passam a ser significativos e há um claro aumento do percentual do impacto de penalidade por ser da zona rural associada as outras características destes alunos. Para o ano de 2013, o impacto negativo no aluno da zona rural no quantil mais baixo chegava a ser de 6%. Nos anos de 2015 e 2017 e no *pooled* os resultados mostravam-se também altamente significativos e com uma penalidade percentual em torno de 3% na nota do aluno da zona rural.

Este aumento de impacto negativo dos quantis inferiores também foram observados nos quantis 10 e 25, os resultados do *pooled* com interação revelam a magnitude desta multa aos estudantes da zona rural, para os alunos com baixo desempenho do q10 o impacto negativo é de 3% a menos na nota, e no q25 o impacto é 2,5% a menos para o aluno da zona rural.

Por fim, para obter estimadores mais eficientes combinou-se o modelo em regressões de influência recentrada (RIF) com estimadores ponderados por probabilidade inversa (IPW) para estimar os efeitos médios de tratamento do quantil (QTE), ao adotar pesos estimados, o IPW estima coeficientes de regressão que são ajustados para dados ausentes, além disso promove a estimação de *scores* de propensão eliminando os valores que estejam muito próximos de 0 ou 1. Os resultados desta proposição podem ser observados na tabela 4.4, o modelo sem interação da *dummy* e na tabela 4.5, para modelos que interagem a *dummy* de localização.

Pode-se observar uma melhora no coeficiente de avaliação de tratamento de efeito do quantil em ambas as estimativas com o IPW, estes coeficientes estão mais estáveis e com

resultados mais similares em termos de medidas de impactos percentuais na nota média do aluno dada a localização da escola em que estuda.

Concentrando a avaliação no *pooled* dos anos 2013-2017, podemos observar que os efeitos de penalidade sobre o desempenho dos estudantes da zona rural são evidenciados em praticamente toda a distribuição de quantis das notas, com elevado nível de significância. Uma pequena diferença acontece no quantil 5, no qual no modelo sem interação o efeito neste quantil é não significativo, já quando avaliado a interação da *dummy* este coeficiente passa a ser altamente significativo, com magnitude do percentual de impacto semelhante ao do modelo sem interação, no qual o fato de estar na zona rural penaliza o estudante em média 2,3% no quantil mais baixo.

Tabela 4.4 – QTE com todos os controles, sem interação e com IPW via logit

Quantil	Tratamento de Efeito dos Quantis (QTE)			
	2013	2015	2017	2013-2017
	<i>n. obs:</i> 750,898	<i>n. obs:</i> 1,002,887	<i>n. obs:</i> 898,029	<i>n. obs:</i> 2,651,838
	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$
5	-.03286*** (.00639)	-.01944*** (.00296)	-.01346*** (.00293)	-.02324 (.00267)
10	-.03450 (.00536)	-.01762 (.00235)	-.01677*** (.00251)	-.02560*** (.00232)
25	-.03078*** (.00383)	-.01630*** (.00204)	-.01861*** (.00211)	-.02419*** (.00183)
50	-.02827*** (.00320)	-.01482*** (.00191)	-.01706*** (.00187)	-.02111*** (.00153)
75	-.02657*** (.00321)	-.01391*** (.00194)	-.00893*** (.00197)	-.01830*** (.00151)
90	-.01717*** (.00342)	-.01600*** (.00224)	-.00022 (.00234)	-.01415*** (.00166)
95	-.01264*** (.00325)	-.01149*** (.00282)	-.00183 (.00262)	-.01199*** (.00184)

Notas: ***significativo a 1%; **significativo a 5%; *significativo a 10%.

Fonte: Elaboração própria a partir do *software* Stata 15.

Neste sentido, os demais resultados dos *pooled* chamam atenção por sua similaridade no percentual de diferencial do aluno na zona rural. Para os quantis superiores q90 e q95 o percentual médio de diferença foi entre 1,1-1,4%. Nos quantis 25 e 50, a diferença de pontos médios percentuais para o estudante da zona rural foi entre 2,1-2,4% negativos. Similarmente, estes valores são próximos aos evidenciados aos dos quantis inferiores.

De modo geral, os resultados aqui evidenciados revelam em grande parcela que há um efeito de diferencial de desempenho dos alunos da escola rural em relação ao estudante da escola urbana, isolando a variável relativa à localização da escola, pode-se captar que o simples fato de estudar na localização rural, impacta em termos médios em torno 2% a menos na pontuação da prova de matemática dos quantis de piores desempenho (q5 e q10).

Nos quantis mais centrais (q25, q50 e q75) este diferencial médio percentual é de uma penalidade em torno de 2,5% para alunos da zona rural, em alguns anos o resultado diferencial se aproxima de 3%. Revelando de maneira importante que em toda a distribuição quantílica os coeficientes apresentam-se altamente significativos em afirmar a hipótese que estudar em uma escola rural gera uma penalidade no desempenho do estudante.

Tabela 4.5 – QTE com todos os controles, com interação e com IPW via logit

Quantil	Tratamento de Efeito dos Quantis (QTE)			
	2013	2015	2017	2013-2017
	<i>n. obs:</i> 750,898	<i>n. obs:</i> 1,002,887	<i>n. obs:</i> 898,029	<i>n. obs:</i> 2,653,619
	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$	$\delta(\tau)_{local}$
5	-.02949*** (.00677)	-.02239*** (.00344)	-.01383*** (.00335)	-.02376*** (.00286)
10	-.03285*** (.00582)	-.01775*** (.00272)	-.01985*** (.00312)	-.02632*** (.00259)
25	-.03043*** (.00409)	-.01597*** (.00240)	-.01931*** (.00272)	-.02484*** (.00203)
50	-.02957*** (.00354)	-.01479*** (.00231)	-.01623*** (.00247)	-.02123*** (.00173)
75	-.02737*** (.00366)	-.01382*** (.002424)	-.00770*** (.00258)	-.01714*** (.00175)
90	-.01697 (.00397)	-.01656*** (.00280)	.00016 (.00302)	-.01359*** (.00192)
95	-.01175*** (.00389)	-.01238*** (.003512)	-.00134 (.00334)	-.01140*** (.00212)

Notas: ***significativo a 1%; **significativo a 5%; *significativo a 10%.

Fonte: Elaboração própria a partir do *software* Stata 15.

Quando fazemos o comparativo do diferencial da média, os resultados recentes para o Brasil revelam uma diferença em torno de 10% negativos para os alunos da zona rural, porém feitas as regressões estes resultados apresentam-se ainda algum resíduo de diferença percentual negativa em torno de 3%, na nossa hipótese este percentual deveria desaparecer a medida que todas as características são controladas, mesmo que ele reduza, ele não desaparece.

5 CONSIDERAÇÕES FINAIS

O presente estudo examinou se a localização da escola na zona rural implica em uma penalidade ao desempenho educacional de seus estudantes. Conhecida as discussões da literatura especializada que indicam que não é a localização em si que afetaria as diferenças de resultados, mas o contexto socioeconômico destas áreas associado a fatores como as características dos alunos, das famílias, das escolas, dentre outras, que causariam esta diferença. Desta forma, foi estabelecida uma estratégia empírica via Regressões Quantílicas Incondicionais e Tratamento de Efeito dos Quantis para avaliar a hipótese que dado um controle das múltiplas dimensões associadas ao processo do desempenho educacional admitidas na literatura, o fator localização seria significativo na determinação da lacuna de desempenho educacional entre as zonas urbanas e rurais.

Ao examinar o diferencial de desempenho dos resultados dos alunos na prova de matemática, é possível observar que houve uma evolução em torno de 5% na nota média dos estudantes da zona urbana, e uma melhora de desempenho de 9,8% na zona rural.

No entanto, apesar dos avanços recentes, os resultados mostram-se ainda desiguais entre as duas localizações, visto que em 2017 a nota média da zona rural foi muito inferior à da zona urbana, com uma diferença de 23 pontos.

Quando avaliada a distribuição de notas pela escala de proficiência, os resultados mostraram-se preocupantes. Em 2015, o percentual de alunos da zona rural que tiraram notas abaixo ou igual ao nível 03 de proficiência representava aproximadamente 60% dos alunos, no sentido oposto 62% dos alunos da zona urbana apresentaram um resultado maior ou igual ao nível 04 de proficiência. Em 2017, observa-se uma melhora muito sutil no percentual de alunos da zona rural no nível inferior ou igual a 03, representando aproximadamente 54,8% dos alunos.

Diante do contexto evidenciado, o tratamento empírico proposto neste trabalho teve como objetivo isolar e avaliar o impacto distributivo da variável localização da escola no desempenho acadêmico. Ao realizar o procedimento empírico por QTE, espera-se que os efeitos da localização sejam zerados. Ou seja, que não haja uma penalidade no desempenho acadêmico por estar na zona rural.

Os primeiros resultados referem-se às estimativas do QTE sem considerar a interação da *dummy* localização da escola com as variáveis de controle, indicando a presença de uma penalidade aos estudantes da zona rural em todos os níveis da distribuição dos quantis. Chama atenção que os resultados dos quantis superiores 75, 90 e 95 em todos os períodos foram altamente significativos e em parcela considerável dos resultados o percentual de penalidade da nota eram maiores. Ao avaliar os resultados da estimação do *pooled* para os períodos 2013-2017, ao aplicar as *dummies* de tempo e a interação com esta variável, o coeficiente do menor quantil não foi significativo, já para os quantis acima do q50, a diferença percentual dos pontos médios variou em torno de -1,4 a -2,1% na pontuação média do aluno.

A medida que avançamos na estratégia empírica proposta, quando estimados os tratamentos de efeitos de quantis com a interação da variável *dummy* localização da escola com as demais variáveis de controle, os resultados mais robustos apontam um novo direcionamento da penalidade no desempenho do aluno que estuda em uma escola da zona rural. Ao interagir a localização com outras características, a penalidade de estar na zona rural é suavizada e deixa de ser significativa em grande parte das estimativas para os quantis mais elevados (90 e 95).

No entanto, os resultados dos quantis inferiores passam a ser significativos e há um claro aumento do percentual do impacto de penalidade por ser da zona rural associada às outras características destes alunos. Para os alunos com baixo desempenho do q10, o impacto negativo é de 3% a menos na nota e, no q25, o impacto é 2,5% a menos para o aluno da zona rural.

Neste sentido, esta mudança de resultado pode ser potencialmente explicada pelo seguinte aspecto, os alunos que já apresentam melhor nível de desempenho, alunos com melhores habilidades formativas não sofreriam os efeitos da penalidade caso estivessem em uma escola rural ou urbana. Diferente dos alunos que já apresentam pior desempenho educacional-formativo, caso estes alunos com maior déficit matemático estivessem em uma escola da zona rural o desempenho na nota da prova Saeb sofreria uma penalidade em torno de -3% da nota.

Por fim, observou-se uma melhora no coeficiente de avaliação de tratamento de efeito do quantil em ambas as estimativas com o IPW, estes coeficientes estão mais estáveis e com resultados mais similares em termos de medidas de impactos percentuais na nota média do aluno dada a localização da escola em que estuda.

A avaliação do *pooled* dos anos 2013-2017 mostra que os efeitos de penalidade sobre o desempenho dos estudantes da zona rural são evidenciados em praticamente toda a distribuição de quantis das notas, com elevado nível de significância. Para os quantis superiores, q90 e q95, o percentual médio de diferença foi entre -1,4 e -1,1%. Nos quantis 25 e 50, a diferença de pontos médios percentuais para o estudante da zona rural foi entre 2,1-2,4% negativos. Similarmente, estes valores são próximos aos evidenciados nos quantis inferiores.

De maneira geral, as estimativas desenvolvidas neste trabalho indicam que existe uma multa educacional ao estudante da zona rural, se compararmos os resultados das estatísticas descritivas em especial a diferença percentual, pode-se observar que, mesmo que as regressões controlem os potenciais efeitos das variáveis preconizadas na literatura e suavizem o diferencial percentual negativo nas notas, esta penalidade não é mitigada, ou seja, esperava-se que o efeito do rural fosse zerado, o que não aconteceu em nenhuma das estimativas realizadas.

Demonstrado que a localização da escola penaliza a nota do aluno, é preciso em termos propositivos direcionar não só investigações futuras sobre o tema, mas também as possíveis ações de políticas públicas a respeito do que ainda poderia estar afetando o desempenho destes

estudantes. Para além das variáveis aqui já avaliadas, dado que nosso modelo controla um conjunto amplo de características do estudante, do *background* familiar e dos atributos relacionados com o perfil dos professores e das escolas.

Uma hipótese razoável de ser avaliada para efeitos de políticas públicas, seria uma avaliação longitudinal da evolução do processo educacional, ou seja, do *background* do aluno e seus impactos na sequência formativa e no processo de aprendizagem.

Desta forma, deve-se estabelecer parâmetros que relacionem a carga educacional no desenvolvimento da formação básica do estudante, seja pelas etapas iniciais anteriores ao 5º ano do ensino fundamental, seja pela compreensão do processo de alfabetização deste aluno e seus efeitos sobre a aquisição de competências fundamentais para o desempenho deles ao longo do percurso escolar. Neste caso, o processo formativo inicial não é controlado no nosso modelo e pode ser um gerador desta permanência da penalidade.

Este aspecto apresenta-se também em certa medida como um fator limitador a este trabalho, dada a dificuldade de implementação de um painel para os indivíduos, já que para a cada ano avaliado, estes alunos não estariam na mesma série do ensino fundamental. O conjunto de microdados públicos não permitem compreender o progresso acadêmico deste estudante, seja na perspectiva do *background* formativo, seja na perspectiva do avanço do aluno na sua trajetória educacional.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALENCAR, N. S.; ARAUJO, J. A.; JUSTO, W. R. e SOBREIRA, D. B. Diferentes características dos professores explicam o gap educacional entre escolas urbanas e rurais no Brasil? **Estudios Económicos**, v. 38, nº 76, p. 45-68, jan-jun 2021.

AMINI, C., & NIVOROZHKIN, E.. The urban–rural divide in educational outcomes: Evidence from Russia. **International Journal of Educational Development**, 44, 118–133, 2015.

BARROS, R. P.; MENDONCA, R. S.; SANTOS, D.; QUINTAES, G.. Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**. Rio de Janeiro, v. 31, n.1, p. 1-42, 2001.

BLINDER, A. S. “Wage Discrimination: Reduced Form and Structural Estimates.” **The Journal of Human Resources**, 8 (4): 436–55, 1973.

BOARDMAN, A., & MURNANE, R. (1979). Using Panel Data to Improve Estimates of the Determinants of Educational Achievement. **Sociology of Education**, 52(2), 113-121. <https://doi.org/10.2307/2112449>

CARD, D. E KRUEGER, A. B. “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.” **Journal of Political Economy** 100 (1): 1–40.1992.

CARD, D. “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems.” **Econometrica** 69 (5): 1127–1160. 2001

CHEVALIER, A. e LANOT, G. The relative effect of Family characteristics and financial situation on educational achievement. **Education Economics**, vol. 10, n. 2, pp. 165-181, 2002.

COLEMAN, J. S *et al.* **Equality of educational opportunity**. National Center for Educational Statistics. U.S. Government Printing Office: Washington: 1966.

FRANCO, A. M. P. e MENEZES-FILHO, N. A. Os determinantes do aprendizado com Dados de um painel de escolas do SAEB. **Economia Aplicada**, v. 21, n. 3, pp.252-548, 2017.

- FIRPO, S. “Efficient Semiparametric Estimation of Quantile Treatment Effects.” **Econometrica** 75(1):259-76, 2007. Disponível em: <http://hdl.handle.net/10438/12995>
- FIRPO, S. P., FORTIN, N. M.; LEMIEUX, T. Unconditional Quantile Regressions. **Econometrica**, v. 77, n. 3, pp. 953-973, 2009.
- FIRPO, S. e PINTO, C. 2016. “Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures.” **Journal of Applied Econometrics** 31(3):457-86, 2016.
- GOMEZ-GONZALEZ, Jose Eduardo; RODRÍGUEZ GÓMEZ, Wilson; RODRÍGUEZ GÓMEZ, Efrén. **Explaining the Rural-Urban Student Performance Gap for Different Distribution Quantiles in Colombia**. Documentos de Trabajo, 2021.
- HANUSHEK, Eric A. Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data. **American Economic Review**, 61(2), pp. 280-288, May 1971.
- HANUSHEK, Eric A. Conceptual and empirical issues in the estimation of Educational Production Function. **The Journal of Human Resources**, vol. 14, n. 3, p. 351-388, 1979.
- HANUSHEK, Eric A. Educational Production Functions. *In*: PSACHAROPOULOS, George **Economics of Education**, Pergamon, 1987, p. 33-42.
- HANUSHEK, Eric A. e WOESSMANN, Ludger. Chapter 14 - Education, knowledge capital, and economic growth. *In*: BRADLEY, Steve e GREEN, Colin (editores). **The Economics of Education**. 2.ed. Academic Press, p. 171-182. 2020a.
- HANUSHEK, Eric A. Chapter 13 – Education production functions. *In*: BRADLEY, Steve e GREEN, Colin (editores). **The Economics of Education**. 2.ed. Academic Press, p. 161-170, 2020b.
- HANUSHEK, Eric A. e WOESSMANN, Ludger. **Education Quality and Economic Growth**. Washington, DC: World Bank, Julho, 27 pg. 2007.
- HÄKKINEN, I., KIRJAVAINEN, T. e UUSITALO, R., School resources and student achievement revisited: new evidence from panel data. **Economics of Education Review**, 22, pp. 329-335, 2003
- IRES, F. N., MARIANO, F. Z. e BENEVIDES, A. A. Índice de motivação multidimensional e desempenho escolar no Brasil. *Revista Brasileira de Economia*. Vol. 74, n. 3, Jul–Set, 2020.
- JUHN, C., MURPHY, K. M. e PIERCE, B., Wage Inequality and the Rise in Returns to Skill. **Journal of Political Economy**, Jun. vol. 101, nº 3, pp. 410-442, 1993.
- KRUEGER, A. B; Experimental estimates of education production functions. **The Quarterly Journal of Economics**, v. 114, n.2, p. 497-532, May. 1999.
- LAVY, V. Expanding School Resources and Increasing Time on Task: Effects on Students’ Academic and Noncognitive Outcomes. **Journal of the European Economic Association**, Volume 18, Issue 1, Pages 232–265, February 2020.
- LEE, J.W., e BARRO, R. J. Schooling quality in a cross-section of countries. **Economica**, v. 68, n. 272, p. 465-488, 2001.
- LOUNKAEW, Kiatanantha. Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. **Economics of Education Review**, 37, 213–225, 2013.

- MARIONI, L. S., FREGUGLIA, R. S., e MENEZES-FILHO, N. A. The impacts of teacher working conditions and human capital on student achievement: evidence from Brazilian longitudinal data. **Applied Economics**, 52: 6, 568-582, 2020.
- MARTINS, L. e VEIGA, P. Do inequalities in parents' education play an important role in PISA students' mathematics achievement tests core disparities? **Economics of Education Review**, vol.29, n.6, pp.1016-1033, 2010.
- MENEZES-FILHO, Naécio. **Os determinantes do desempenho escolar no Brasil** [Sumário Executivo]. Instituto Futuro Brasil, IBMEC São Paulo e Faculdade de Economia e Administração da Universidade de São Paulo, 2007
- OAXACA, R. "Male-Female Wage Differentials in Urban Labor Markets." **International Economic Review**, 14 (3): 693–709, 1973.
- RAMOS, R., DUQUE, J.C. e NIETO, S. Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata. IZA Discussion Paper, nº 6515, apr. 2012.
- RAPOSO, I. P. de A.; ALMEIDA DE MENEZES, T.; CARVALHO DE ANDRADE LIMA, R.; ZIMMERLE DA NÓBREGA, R. Peer effects and scholastic achievement: spatial models estimates using the student friendship network at the classroom level. **Economia Aplicada**, [S. l.], v. 23, n. 1, p. 5-24, 2019. DOI: 10.11606/1980-5330/ea134204.
- RIOS-AVILA, F. **Recentered Influence Functions in Stata: Methods for Analyzing the Determinants of Poverty and Inequality**. Working paper 927, Levy Economics Institute. 2019. Disponível em: http://www.levyinstitute.org/pubs/wp_927.pdf
- RIOS-AVILA, F e MAROTO, M. L. **Moving Beyond Linear Regression: Implementing and Interpreting Quantile Regression Models with Fixed Effects**. Sociological Methods & Research. p. 1-44, feb. 2022.
- RIVKIN, S. G., HANUSHEK, E. A. e KAIN, J. F. Teachers, Schools, and Academic Achievement. **Econometrica**, 73(2), pp. 417-458, Mar. 2005.
- RODRIGUES, C. R.; RIOS-NETO, E. L. G.; PINTO, C. C. X. . Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico para o período de 1997 a 2005. **Revista Brasileira de Estudos de População** (Impresso), v. 28, p. 5-36, 2011.
- RODRIGUES, L. O; COSTA, E. M.; SILVA; V. H. M. C.; MARIANO, F. Z; FILHO, J. J. A note on performance differences between urban and rural schools in Brazil. **Revista Brasileira de Economia**, v. 74, nº. 4, p.494–507, Out–Dez 2020.
- SANFO, J. B. M.B. e OGAWA, K. Explaining the rural-urban learning achievements gap in Ethiopian primary education: a re-centered influence function decomposition using Young Lives data. **Education Economics**, v.29, nº 3, p.269-297, 2021.
- SANTOS, M. M. dos; MARIANO, F. Z.; COSTA, E. M. Efeitos da educação dos pais sobre o rendimento escolar dos filhos via mediação das condições socioeconômicas. **Economia Aplicada**, [S. l.], v. 23, n. 2, p. 145-182, 2019. DOI: 10.11606/1980-5330/ea144751.
- TODD, P. E., e K. I. WOLPIN. "On the Specification and Estimation of the Production Function for Cognitive Achievement." **The Economic Journal**, 113 (485): F3–F33. 2003.
- WOESSMANN, L. "The Importance of School Systems: Evidence from International Differences in Student Achievement." **Journal of Economic Perspectives**, 30 (3): 3-32, 2016.