

O impacto da escola pública brasileira sobre o desempenho educacional: uma análise a partir de um modelo de Regressão de Quantis Extremos^a

The Impact of Brazilian Public Schools on Educational Performance: an Analysis Based on an Extreme Quantile Regression

Thiago Costa Carvalho^b 

Universidade Federal Rural do Semi-Árido, Laboratório de Engenharia Econômica e Finanças, Mossoró (RN), Brasil

Wallace Patrick Santos de Farias Souza^c 

Universidade Federal da Paraíba, Departamento de Economia, João Pessoa (PB), Brasil

Erik Figueiredo^d 

Instituto Mauro Borges (IMB), Goiânia (GO), Brasil

Resumo: Debater desempenho educacional e mensurar os condicionantes que influenciam o resultado dos alunos é condição primordial para políticas educacionais que pretendem reduzir o quadro de desigualdades no Brasil. Este ensaio tem como objetivo avaliar a contribuição da escola pública nos diferentes intervalos de resultados dos alunos no Enem, buscando captar as diferenças de desempenho a partir da análise por quantil. Para isso, é utilizado um modelo baseado em regressão quantílica para captar a diferença dos condicionantes ao longo da distribuição dos resultados e aplicado o modelo semi-paramétrico de seleção de amostra baseado em regressão por quantis extremos para superar o problema da endogeneidade. Os resultados apontam uma significativa diferença de oportunidades educacionais em desfavor do aluno proveniente da escola pública.

Palavras-chave: Desempenho educacional. Escola pública. Regressão de Quantis Extremos.

Editor responsável: Diogo de Prince

^a Submissão: 19/07/2021 | Aprovação: 18/06/2022 | DOI: 10.22456/2176-5456.116732

^b thiago.costa@ufersa.edu.br

^c wpsfarias@gmail.com

^d eafigueiredo@gmail.com

Todos os autores foram responsáveis pela concepção, pesquisa de dados e/ou documentos, análise dos dados e/ou documentos, participação ativa na discussão dos resultados e revisão e aprovação da versão final.



Esta publicação está licenciada sob os termos de Creative Commons Atribuição 4.0 Internacional

Abstract: Discussing educational performance and measuring the conditions that influence the students' results is essential for educational policies that aim to reduce inequalities in Brazil. This paper aims to evaluate the contribution of public schools in the different ranges of student results in Enem, seeking to capture differences in performance based on quantile analysis. To this end, a model based on quantile regression is used to capture the difference in conditioning factors along the distribution of results and the semi-parametric model of sample selection based on extremal quantile regression is applied. The results point to a significant difference in educational opportunities to the disadvantage of the student coming from public school.

Keywords: Educational Performance. Public School. Extremal Quantile Regression.

JEL: I21. I28. P36.

1. Introdução

No debate sobre políticas e estratégias que impulsionam o desenvolvimento de um país há um consenso ligado aos efeitos positivos da educação, como é possível ver em Anderson, Leo e Muelhaupt (2009); Anderson e Leo (2017); Hanushek (1970) e Heckman e Masterov (2004). Problemáticas relevantes na educação brasileira como analfabetismo, evasão escolar, formação docente e aprendizagem persistem, apesar do crescente aporte orçamentário em todas as esferas de governo e avanços institucionais. Destes avanços, é possível destacar a Lei de Diretrizes e Bases da Educação (LDB/1996); Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (FUNDEF/1996); Fundo Nacional de Desenvolvimento da Educação Básica (FUNDEB/2007); Base Nacional Comum Curricular (BNCC/2017), além do Exame Nacional do Ensino Médio (Enem/1998).

O Brasil é um país reconhecido pela intrínseca desigualdade social, refletida também na educação. Essa desigualdade educacional se apresenta de diversas formas, tais como: o acesso ao sistema escolar, a diferença na formação dos docentes, a disparidade de infraestrutura, desigualdade de oportunidades e a disparidade nos processos de aprendizagem.

Mensurar o desempenho educacional e definir os fatores que influenciam esses resultados é ponto crucial para políticas educacionais que possibilitem a redução das desigualdades econômicas e sociais. Portanto, compreender o processo educacional permite que investimentos na área gerem efetivos retornos para a sociedade (Hanusker, 1979). O desempenho depende de uma combinação de variáveis sujeitas às escolhas dos indivíduos, como o nível de esforço e a decisão de migrar, e às condições fora desse controle decisório, como as características familiares, etnia, educação parental, tipo de escola, entre outras. Os resultados que diferem devido ao esforço individual podem representar uma desigualdade natural, porém, se a variação do resultado econômico deriva de circunstâncias impostas ao indivíduo, como raça ou gênero, pode ser considerada uma desigualdade socialmente inaceitável (Hild; Voorhoeve, 2001).

Na medida em que as circunstâncias de um indivíduo estão relacionadas com seu *background* familiar, as políticas sociais necessitam buscar uma condição onde os indivíduos tenham acesso a serviços públicos com qualidade similar. Ao mesmo tempo, devem ser tratados de forma análoga, sem limitações impostas por

obstáculos artificiais, preconceitos ou preferências, o que atende ao conceito de igualdade de oportunidades (Anderson; Leo; Muelhaupt, 2009). Definido por Roemer (1996), esse conceito pondera que resultados econômicos distintos entre dois indivíduos são decorrentes de fatores de responsabilidade e de não-responsabilidade.

Igualdade de oportunidades está intrinsecamente relacionada ao desempenho educacional quando se percebe que um indivíduo com baixo *background* (fatores de não responsabilidade) sofre grande desvantagem para alcançar estrato social superior (Figueiredo; Nogueira; Santana, 2014). Isso é identificado na realidade brasileira quando dada a existência de expressiva desigualdade de aprendizagem entre os alunos do sistema público e do privado (Delgado; Ribeiro; Soares, 2012; Mediavilla; Duque, 2012). Essa desigualdade também pode ser visualizada no ensino público, conforme estudos sobre escolas militares (Benevides e Soares, 2020) e institutos federais (Dutra, 2019).

Uma condição socioeconômica vantajosa que permite ao aluno frequentar uma escola privada gera a possibilidade de compensar efeitos negativos de outros fatores sobre o desempenho, expondo o abismo existente entre o rendimento dos alunos por tipo de escolas (De Sousa *et al.*, 2020) (Lobo; Cassuce; Cirino, 2017). Posto isso, uma possível forma de atender ao conceito de igualdade de oportunidades, segundo Grawe (2007), está na ampliação do acesso e melhoria da qualidade da escola pública, o que pode resultar em maior persistência de ganhos e em maior e melhor mobilidade social.

Esse ensaio pretende avaliar como a escola pública brasileira afeta os diferentes intervalos de desempenho dos alunos de ensino médio no Enem 2017. Tradicionalmente, estudos sobre performance educacional, como De Souza *et al.* (2020), Dutra *et al.* (2019), Carvalho (2017) e Travitzki, Calero e Boto (2014) não abordam os efeitos dos condicionantes do desempenho educacional nos intervalos da distribuição da variável de resposta o que se apresenta como lacuna para uma investigação empírica. Além disso, trabalhos como Arraes e Mariano (2019) e Feijó e França (2021) que também propõem uma análise por quantil não é abordada a questão da engoneidade. Esse problema incide em casos onde há correlação entre uma variável omitida e algum dos regressores o que leva basicamente a correlação entre a variável explicativa e erro, podendo ocorrer devido a fatores como a omissão de variáveis.

Ao mesmo tempo, foram incluídas na base dados informações que buscam captar informações sobre as oportunidades educacionais que os alunos tiveram acesso durante o ensino médio. Entre essas estão o acesso a escolas federais, bolsas de estudo, recursos tecnológicos em casa e na escola, entre outros. Por fim, é proposto um indicador de *background* familiar que considera a escolaridade e ocupação dos pais e uma análise de efeitos migratórios.

Utilizando uma modelagem baseada em regressões quantílicas, nesse trabalho busca-se traçar um comparativo do comportamento das variáveis dividindo os alunos por quantis de desempenho. Conforme Koenker e Bassett (1978), a regressão quantílica é robusta quanto à presença de *outliers* e de erros não distribuídos normalmente, dado que o método investiga componentes particulares da distribuição condicional da variável de resposta, é possível capturar diferenças nos efeitos conforme diversos quantis. De modo a resolver o problema de endogeneidade é utilizada a proposta de um modelo semi-paramétrico de seleção de amostra idealizada por D'Haultfoeuille, Maurel e Zhang (2018). Este procedimento é fundamentado em Regressão de Quantis Extremos (Chernozhukov *et al.*, 2005; Chernozhukov; Fernández-Val; Kaji, 2016) e pode ser usado na ausência de uma variável instrumental, que é uma variável correlacionada com o regressor, porém não correlacionada com os fatores não observáveis que determinam a variável dependente.

O trabalho está dividido em cinco seções, sendo a primeira esta introdução e seguida pela seção 2 que apresenta uma breve revisão de literatura sobre desempenho educacional. A seção 3 é destinada ao modelo teórico, métodos de estimação e descrição do banco de dados; a seção 4, que analisa os resultados; e a seção 5, com as considerações finais.

2. Revisão de literatura: Condicionantes do desempenho educacional

Os efeitos econômicos da educação estão presentes em grandes discussões, como em Romer (1994), Mankiw, Romer e Weil (1992) e Barro (1991), os quais explicam a diferença de renda entre países a partir da implementação de políticas educacionais. Mesmo considerando que o investimento educacional apresenta retorno positivo para a sociedade, Marconi (2016) alerta que se faz necessário adotar um ponto de vista de longo prazo ao decidir investir em educação, seja avaliando a perspectiva do indivíduo ou de uma nação.

Há evidências robustas na literatura, como em Barro (1991); Heckman e Masterov (2004) e Jiménez e Jiménez (2019) sobre o retorno do investimento em educação, que consideram especialmente regiões com baixos níveis de escolaridade. Esses retornos também podem ser vislumbrados em efeitos de produtividade social quando há investimentos direcionados a crianças em ambientes desfavorecidos (Heckman; Masterov, 2004). Além disso, tais investimentos devem ter como base a igualdade de oportunidades e possibilitar que indivíduos diferentes, independentemente das circunstâncias fora de seu controle, tenham as mesmas possibilidades para alcançar resultados através de suas escolhas (Roemer, 1996).

Uma evolução nesse debate está no conceito de igualdade de oportunidades qualificada destacado em Anderson, Leo e Muelhaupt (2009), Anderson e Leo (2017) e Jiménez e Jiménez (2019), que se concentra apenas em elevar as perspectivas dos indivíduos que apresentam características desfavoráveis sem diminuir os efeitos das características positivas. Nesse caso, as políticas devem possibilitar ganhos para indivíduos com a chamada “herança pobre”, tal qual a baixa escolaridade dos pais. Contudo, essas políticas devem apresentar mecanismos que não reduzam as oportunidades dos indivíduos com uma “herança rica”.

Ao abordar políticas que permitam a igualdade de oportunidades é inevitável abordar a desigualdade educacional. Travetzki *et al.* (2016) destacam que a performance dos alunos do ensino médio é sensível a condições sociais, espaciais e raciais, e que a redução da influência das características familiares pode estar relacionada ao impacto positivo da escola sobre o aluno. Isso foi perceptível a partir da decomposição da variância do desempenho educacional no nível de ensino pré-universitário para diferentes estados através de um modelo multinível.

O entendimento de que as circunstâncias (renda familiar, tipo de escola, escolaridade do pai) são significativas no desempenho escolar também foi resultado de Nogueira e Figueiredo (2019). Utilizando variável instrumental para corrigir o problema da omissão de variável de esforço, foi identificado que indivíduos de baixo *background* social precisam de quase o dobro de empenho para alcançarem as melhores notas.

Dito isso, seria possível intervir no círculo vicioso da desigualdade educacional? Para Vernier, Bagolin e Jacinto (2015) e Riane e Rios-Neto (2008), melhorar a rede escolar de ensino médio representa um fator que permite diminuir a estratificação educacional, o que possibilita um menor impacto de fatores

relacionados ao ambiente familiar do aluno. Escolas com desempenho alto em exames educacionais apresentam número menor de determinantes e peso maior para as características dos professores, dos diretores e da escola.

Parman (2011) destaca que o retorno marginal de ensino médio é mais robusto que nos demais níveis educacionais. Essas mudanças têm maiores efeitos sobre o bem-estar econômico do que modificações na escolaridade total, assim ampliar o acesso e melhorar a qualidade no ensino médio pode gerar respostas distintas em indivíduos de diferentes níveis de renda.

É importante do mesmo modo aprofundar a discussão do efeito do tipo de escola sobre o desempenho educacional, dado que essa variável está presente em diversos estudos da Economia da Educação. Lobo, Cassuce e Cirino (2017) destacam que parte da variância do desempenho está relacionada ao tipo de escola, tendo a escola pública impacto negativo. A partir de modelos hierárquicos com as características socioeconômicas do aluno e o tipo de escola, reforça-se o impacto negativo da pública em comparação à privada.

Comparando escolas públicas e privadas de nível médio, Diaz *et al.* (2012) apontam a estabilidade da desigualdade de oportunidades oriunda de melhorias dos indicadores de distorção idade-série e do aumento da participação das escolas públicas no número de concluintes do ensino médio no Brasil. A redução da idade dos alunos e da demanda por escolas privadas é acompanhada pela redução do nível de desigualdade de oportunidades.

Além do debate entre escolas públicas e privadas, merece destaque outros tipos de escola que apresentam características específicas podem impactar diretamente o desempenho do aluno. Benevides e Soares (2020) compararam os resultados das escolas públicas militares com as escolas públicas tradicionais utilizando modelos de *propensity score matching* (PSM) e regressões lineares multivariadas. Verificou-se que apesar da falta de controle de viés de seleção o efeito da escola sobre o desempenho é significativo nas unidades educacionais militares. De Sousa *et al.* (2020) também destacam que escolas que adotam um viés religioso também apresentam diferencial positivo no desempenho educacional.

Em Arraes e Mariano (2019) e Feijó e França (2021) discutiram os diferenciais entre escolas públicas e privadas adotando métodos de decomposição do resultado. Cabe destacar desses trabalhos que há evidências de desempenho superior por parte dos alunos de escolas privadas em comparação a escolas públicas,

em especial nos quantis inferiores de desempenho. O que pode justificar essa situação reside em questões como a escolaridade e renda dos pais, o background da turma e as características dos docentes e da escola.

Outro diferencial no processo educacional reside na possibilidade de o aluno dedicar o máximo de tempo a escola. Em virtude das circunstâncias familiares muitos alunos de escolas públicas são obrigados a dividir o tempo entre trabalho e escola. Em Harkreader (2008) e Figlio e Karbownik (2016) é demonstrado que quanto maior o tempo de estudo do aluno melhor o desempenho em exames educacional e que políticas de bolsas de estudo podem auxiliar na manutenção e dedicação a vida escolar.

Lobler *et al.* (2010) chamam atenção para o acesso a novas tecnologias no processo educacional e como isso impacta no desempenho dos alunos. Pelo exposto é possível exemplificar que incluir acesso a computadores, *internet* e outros instrumentos tecnológicos no processo educacional têm efeitos positivos sobre os alunos de melhor desempenho.

É possível presumir que melhorar a qualidade da escola pública pode resultar em maior persistência de ganhos, enquanto aumento na quantidade de escolas públicas leva a maior mobilidade social (Grawe, 2007). É preciso estar claro que a política educacional, quando bem elaborada e executada, tem a faculdade de aperfeiçoar a mobilidade social desde que reduza as lacunas educacionais para fortalecer a coesão social, tornando possível a redistribuição de renda na idade adulta (Kim, 2014). Em síntese, o debate sobre desempenho educacional continua pertinente e atual na literatura econômica, o que justifica avançar na pesquisa sobre os determinantes do desempenho educacional e qual a contribuição da escola pública.

3. Estratégia empírica

Este ensaio propõe uma análise com base no conceito de igualdade de oportunidades qualificada empregando uma abordagem que permite separar as melhores e piores performances no Enem. Isso tornará possível observar os efeitos dos condicionantes nos diferentes níveis de desempenho dos alunos, em especial o impacto da escola pública.

Essa seção apresenta uma primeira subseção com o modelo teórico seguida pela apresentação da ideia geral da regressão quantílica. Às três últimas subseções destacam conceito de regressões de quantis extremos, o modelo de seleção de amostra com base em regressões quantis extremos e por último a base de dados.

3.1 Modelo teórico

É recorrente na literatura o emprego da Função de Produção Educacional (FPE) em estudos sobre o desempenho educacional. Essa abordagem permite relacionar insumos educacionais a uma variável que mensure o resultado do processo educacional. A FPE estuda a relação de produtividade entre os insumos e produtos do processo educacional, objetivando entender a tecnologia de combinar os insumos escolares e familiares de forma que o resultado educacional seja maximizado. Dessa forma, a mensuração dos efeitos dos fatores sobre o desempenho educacional possibilita visualizar as causas da estratificação educacional (Hanushek, 1970; 1979; 1996; Hanushek; Wobmann, 2007). A FPE é dada por:

$$A_{it} = f(B_i^t, P_i^t, S_i^t, \varepsilon^t) \quad (1)$$

Em que A_{it} representa o rendimento escolar do aluno, que pode ser quantificado por exames de desempenho, média de anos de estudo, taxa de repetência ou evasão. Essa variável está em função de grupos distintos de fatores, sendo:

- ✓ Características familiares (B_i^t): *background* familiar;
- ✓ Influência dos pares (P_i^t): Características da turma (a literatura destaca a dificuldade de separar o efeito da turma do efeito individual);
- ✓ Características escolares (S_i^t): nível educacional, experiência do professor, taxa professor/aluno; gasto por aluno, salário dos professores; salário ou qualidade dos administradores e instalações;
- ✓ Erro aleatório (ε^t): reflete os fatores não mensurados.

Buscando tornar robusta a análise, a FPE pode ser estimada através de regressão quantílica, o que torna possível identificar variações inter e intraquartis, não comportadas pela estimação de OLS, uma vez que essa é baseada na média da

distribuição condicional. Além de caracterizar toda a distribuição condicional, esse método é robusto a *outliers* e permite estimar os intervalos de confiança dos parâmetros e do regressando diretamente dos quartis condicionais. O fato dos erros não possuírem distribuição normal permite que os estimadores sejam mais eficientes (Vernier; Bagolin; Jacinto, 2015).

A estimação via regressão quantílica permite debelar esses problemas tornando os resultados robustos, conforme Cheety *et al.* (2014). Assim é proposto para esse estudo um modelo baseado na equação a seguir:

$$Y_i = \alpha + \beta_1 X + \beta_2 F + \beta_3 E + \mu \quad (2)$$

Em que “Y” refere-se à medida de desempenho educacional (Resultado no Enem); “X” representa o vetor de atributos do indivíduo (idade, sexo, raça); “F” o vetor de *background* familiar (escolaridade dos pais, ocupação dos pais, status social, tamanho da família) e “E” características de oportunidade educacional (pública, privada, rural, urbana, acesso à internet). Os parâmetros são definidos por α como a constante do modelo, os β serão vetores de coeficientes a serem estimados e μ é o termo de erro.

O modelo empírico será estimado por OLS (média), quantil 10, quantil 50 (mediana) e quantil 90. Isso permitirá comparar como os coeficientes se comportam ao longo da distribuição da variável dependente. A principal preocupação é observar os extremos da distribuição. Neste caso, o quantil 10 representa os alunos de pior desempenho e o quantil 90 os alunos de melhor desempenho.

A forma funcional do modelo segue o padrão *log-lin*, na qual a variável resposta é apresentada na forma logarítmica e as variáveis independentes em seus valores lineares. Figueiredo, Lima e Schaur (2014) sugerem que o modelo log-linear pode ser estimado pelo método de regressões quantílicas, reforçando que a propriedade de equivariância dos quantis soluciona o problema de identificação dos parâmetros. Essa forma descreve padrões de associação entre variáveis categóricas dado que os coeficientes estimados medem a variação relativa constante na variável dependente y para uma variação absoluta do regressor.

Entre as peculiaridades que podem ser observadas no Enem está uma expressiva heterogeneidade entre os participantes. É possível em uma mesma sala

de aplicação do exame encontrar um adolescente e uma pessoa da terceira idade; um indivíduo que ainda está cursando o ensino médio e outro fora da escola há mais de uma década. Essa condição pode implicar em um problema de heterocedasticidade e de *outliers*. Para tornar a amostra mais homogênea, com característica semelhantes e captar melhor o efeito da escola sobre o desempenho foram selecionados apenas os alunos entre 15 e 19 anos⁵, faixa-etária que corresponde à idade-série ideal para o ensino médio definido pelo Ministério da Educação. Assim busca-se evitar condições que poderiam comprometer a eficiência e a consistência dos estimadores.

Koenker e Bassett (1978) e Koenker e Hallock (2001) apresentam os Modelos de Regressão Quantílicas que estimam a resposta da variável dependente em termos dos valores das covariáveis a partir de distribuição por quantil. Diferentemente de uma regressão linear que apresenta estimações a partir da média, esses modelos apresentam estimações pela mediana ou outros quantis definidos na pesquisa. Essa abordagem permite identificar diferenças nas associações entre os quantis das variáveis de resposta e as variáveis explicativas, o que torna o instrumento importante para comparações, não só pelos valores médios, pois é possível estimar as causas para cada quantil modelado a partir das estimativas dos coeficientes de regressão.

Ao contrário da regressão de Mínimos Quadrados Ordinários (OLS), a regressão quantílica do estimador mediano da regressão minimiza a soma dos erros absolutos em detrimento aos erros ao quadrado. Essa proposta é considerada uma técnica de estimativa robusta substancialmente menos sensível aos valores extremos que as técnicas de regressão usuais (Fertig, 2003).

3.2 Regressão de Quantis Extremos

A regressão quantílica extrema pode ser entendida como o emprego da regressão quantílica às caudas da distribuição condicional através de aproximações baseadas na Teoria de Valores Extremos. Essa estratégia empírica fornece aproximações mais precisas que as leis gaussianas nas caudas, além de desenvolver estimadores corrigidos de viés e métodos de inferência ao usar simulação e variações adequadas de *bootstrap* e sub-amostragem.

⁵ Ver <http://portal.mec.gov.br/brasil-alfabetizado/195-secretarias-112877938/seb-educacao-basica-2007048997/13559-ensino-medio-introducao>

A proposta apresentada em Chernozhukov (2005) e Daouia *et al.* (2013) combina restrições da teoria de valores extremos com especificações lineares de homocedásticas e heterocedásticas na análise de regressão. Dessa forma, assume-se que as caudas da distribuição dos retornos (Y) comportam-se aproximadamente como uma função de potência. Considerando grandes amostras, os quantis de regressão de ordem extrema convergem fracamente para funções integrais estocásticas de processos de Poisson, enquanto os quantis de regressão centrais convergem para vetores normais com matrizes de variância dependentes dos parâmetros da cauda. Dessa forma é possível que os regressores apresentem impactos nos quantis da cauda condicional de maneira diferente dos quantis centrais considerando a mesma variável de resposta.

Em Chernozhukov, Fernández-Val e Kaji (2016), é demonstrado que são incorporadas as caudas da distribuição da variável resposta às especificações lineares para estimar a função de regressão, com base nas condições de Pareto, dadas por: $x \mapsto Q_v(\tau/x)$. Definida uma variável U por $U := Y$, se o ponto final inferior do suporte de Y for $-\infty$ e por $U := Y - Q_v(0)$ se o ponto final inferior do suporte de Y for finito. Em linhas gerais U é uma duplicação deslocada de Y cujo limite termina em $-\infty$ ou 0. A suposição de que a variável aleatória U exibe uma cauda do tipo Pareto é declarada pelas duas seguintes condições equivalentes:

$$Q_u(\tau) \sim L(\tau) \cdot \tau^{-\xi} \quad \text{com } \tau \searrow 0 \quad (5)$$

$$F_u(u) \sim \bar{L}(u) \cdot u^{-\frac{1}{\xi}} \quad \text{com } u \searrow Q_u(0) \quad (6)$$

Para todo $\xi \neq 0$, em que $\tau \mapsto L(\tau)$ é uma função não-paramétrica de variação lenta em 0, e $u \mapsto \bar{L}(u)$ é uma função não-paramétrica de variação lenta em $Q_u(0)$, essas duas funções de variação lenta podem ser do tipo constante ou logarítmica.

É assumida a existência de um parâmetro de regressão auxiliar $\beta_e \in R^{k_x}$ tal que o distúrbio de erro é dado por $V := Y - X'\beta_e$ tem ponto final inferior $s = 0$ ou $s = -\infty$ a.s., e sua função quantílica condicional $Q_v(\tau/x)$ satisfaz a relação' de equivalência de cauda $Q_v(\tau/x) \sim_x \gamma Q_v(\tau)$ dado que $\tau \searrow 0$ uniformemente em $x \in \subseteq R^{k_x}$.

Uma função quantílica $Q_v(\tau)$ que exibe uma cauda do tipo Pareto apresenta $Q_v(\tau) \sim L(\tau) \cdot \tau^{-\xi}$ com $\tau \searrow 0$ com EV índice ξ , e um parâmetro vetorial γ tal que

$E[X']\gamma = 1$ e $X'\gamma > 0$ a.s. Essa condição afeta apenas as caudas distantes e, portanto, permite que as covariáveis afetem os quantis extremos e centrais distintamente. EV corresponde ao índice de valor extremo ou índice de cauda e o valor absoluto de ξ corresponde ao peso da cauda na distribuição e tem necessariamente um limite inferior finito se $\xi < 0$ e um limite inferior infinito se $\xi > 0$. O efeito local das covariáveis é aproximado pela condição onde $\beta(\tau) \approx \beta_c + \gamma Q_v(\tau)$, que pode ser heterogêneo entre os quantis extremos.

A importância de assumir que as caudas da distribuição são do tipo Pareto significa que essas decaem aproximadamente como uma função potência, ou mais formalmente, uma função que varia regularmente. Essas restrições são impostas de uma maneira que permitem que os regressores impactem os quantis da cauda condicional da variável de resposta de maneira diferente dos quantis centrais. A modelagem cobre modelos de regressão de deslocamento de localização convencionais, bem como modelos de regressão de quantis mais gerais (Chernozhukov, 2005).

3.3 Modelo de seleção de amostra com base em Regressões de Quantis Extremos

A proposta de D'Haultoeuille, Maurel e Zhang (2014) estima e fornece inferência para um modelo semi-paramétrico de seleção de amostra baseado em regressão de quantis extremos e pode ser usado na ausência de um instrumento ou de um grande regressor de suporte. Presume-se que as variáveis de interesse tenham efeitos homogêneos sobre o resultado em sua distribuição. Métodos anteriores que propõem estimar modelos de seleção de amostras endógenas recomendam a utilização variáveis instrumentais ou grandes regressores de suporte.

Esse método se apresenta como uma opção para situações onde há dificuldade na definição de instrumentos válidos. A estratégia proposta se baseia na ideia de que, desde que a seleção seja endógena, pode-se esperar que o efeito do resultado na seleção domine as das covariáveis (D'haultoeuille; Maurel; Zhang, 2018).

Dentre as características do modelo destaca-se a não obrigatoriedade da normalidade do termo de erro na equação de seleção, nem linearidade da expectativa condicional do termo de erro nas equações de resultado. Além disso, o método não restringe o processo de seleção, além de uma independência na

condição de infinito e permite efeitos distributivos heterogêneos de outras variáveis de controle.

A intuição inicial é que, se a seleção for verdadeiramente endógena, se pode esperar que o efeito do resultado sobre a seleção domine o das covariáveis para valores suficientemente grandes do resultado. Seguindo essa ideia, a identificação ocorre sob a condição de que a seleção se torna independente das covariáveis no limite, ou seja, quando o resultado tende para o limite superior de seu suporte (D'haultoeuille; Maurel; Zhang, 2014). Considerando a seguinte função:

$$Y^* = X_1' \beta_1 + \varepsilon \quad (10)$$

onde $Y^* \in R$ e $X_l \in R^{k_l}$ são o resultado e as covariáveis de interesse, respectivamente. Para estimar β_1 se faz necessário que $\tau \in (0, 1)$ onde o τ -ésimo quantil condicional de ε satisfaz:

$$Q_{\varepsilon|X}(\tau|X) = \beta_0(\tau) + X_2' \beta_2(\tau) \quad (11)$$

onde $X = (X_1', X_2')$ e X_2 denota outras covariáveis, então:

$$Q_{Y^*|X}(\tau) = X_1' \beta_1 + \beta_0 + X_2' \beta_2(\tau) \quad (12)$$

O efeito de X_l é assumido como homogêneo em diferentes índices de quantis, enquanto o efeito das outras covariáveis X_2 pode ser heterogêneo ao longo da distribuição de Y^* . Essa variável de resposta Y^* não é observada diretamente e sim denotada uma *dummy* de seleção onde $Y = DY^*$. Uma segunda condição de estimação considera que a seleção é independente das covariáveis, assumindo que existe uma constante $h \in (0,1]$ para todo $X \in \text{Supp}(X)$,

$$\lim_{y \rightarrow \infty} P(D = 1 | X = x, Y^* = y) = h \quad (13)$$

Cabe destacar que o desempenho dos estimadores de quantis extremos depende de um *trade-off* entre viés e variância, que é determinado pelo índice de quantis τ_n usado na regressão. Considerando que o banco de dados disponibilizado pelo Enem não apresenta indicadores para variáveis de esforço e opções para variáveis instrumentais, essa solução aqui apresentada se qualifica a atender os objetivos da pesquisa.

3.4 Base de dados

Segundo a LDB⁶ a educação básica é formada pela educação infantil, ensino fundamental e ensino médio, e sua avaliação é de responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Como já citado anteriormente a partir de meados dos anos 1990, uma série de mudanças institucionais ocorreram na educação brasileira, entre estas mudanças está a criação do Enem.

O Enem é o maior exame educacional brasileiro e teve seu início, no ano de 1998, visando avaliar se os alunos que concluíram o ensino médio apresentam domínio sobre conteúdo das áreas de linguagem, matemática, ciências humanas e naturais. É aplicado aos alunos das redes públicas e privadas, em áreas urbanas e rurais, que concluíram a 3ª série do ensino médio. Além disso, possibilita analisar o rendimento dos alunos e é a principal porta de acesso ao sistema universitário brasileiro.

Este trabalho utiliza os microdados do Enem e do Censo Escolar para o ano de 2017 disponibilizados pelo INEP⁷. Esses dados agrupam informações sobre as características dos participantes, suas famílias, escolas, perfil socioeconômico e o resultado da avaliação, conforme o Quadro 1.1. A amostra selecionada apresenta o total de 912.866 participantes do exame, desses 75,9% são oriundos de escola pública e 24,1% são alunos de escolas privadas.

⁶ LDB - Leis de Diretrizes e Bases. Lei nº 9.394. 1996.

⁷ <http://inep.gov.br/microdados>

Quadro 1.1 - Descrição das variáveis adotadas

Grupo	Variável	Descrição
Desempenho Educacional	Prova de redação	Logaritmo natural da pontuação que está entre zero e 1.000
Características Individuais	Idade	Anos de idade
	Sexo	1 para mulher e 0 para homem
	Raça	1 para branco ⁸ e 0 para não-branco.
	Migração	1 para aluno que mora em cidade diferente do seu nascimento; 0 para aluno que mora na mesma cidade do seu nascimento.
	Migração escolar	1 para aluno que estuda em cidade diferente da sua residência; 0 para aluno que estuda na mesma cidade em que reside.
Características Familiares	<i>Background</i> familiar	Variável composta pela soma dos valores das variáveis escolaridade entre o pai, escolaridade da mãe, grupo ocupacional entre o pai e grupo ocupacional da mãe (Escala de 1 a 7) ⁹
	Tamanho da família	Número de pessoas que moram na mesma residência do aluno
	Renda familiar	Renda familiar em salários mínimos (Escala de 1 a 20)
	Computador residencial	1 para aluno que teve acesso à computador em casa e 0 caso contrário
	Internet residencial	1 para aluno que teve acesso à internet em casa e 0 caso contrário.
Oportunidades Educacionais	Escola pública	1 para escola pública e 0 para escola privada
	Escola rural	1 para escola rural e 0 para escola urbana
	Escola federal	1 para escola federal e 0 caso contrário
	Internet escolar	1 para acesso à internet na escola e 0 sem acesso
	Bolsista	1 para aluno que recebeu bolsa de estudos no ensino médio em escola pública ou privada e 0 para não-bolsista
	Ensino médio híbrido	1 para aluno que frequentou ensino médio parte em escola pública e parte em escola privada; 0 para aluno que frequentou ensino médio totalmente em escola pública ou escola privada.
	Atividades complementares ¹⁰	1 para aluno frequentou atividades complementares no ensino médio, 0 caso contrário.
	Ensino técnico integrado	1 para aluno que cursou ensino técnico integrado ao ensino médio, 0 caso contrário.
Região	Norte	1 para estudante da região norte e 0 do contrário.

⁸ Autodenominação dos indivíduos quanto a sua raça/cor.⁹ A variável “escolaridade” segue uma escala crescente de 1 a 7 e a variável “grupo ocupacional” segue uma escala crescente de 1 a 5 conforme definição do INEP.¹⁰ Atividades de livre escolha que complementam a escolarização e o currículo obrigatório (BRASIL, 2017).

	Nordeste	1 para estudante da região nordeste e 0 do contrário.
	Centro-oeste	1 para estudante da região centro-oeste e 0 do contrário.
	Sudeste	1 para estudante da região sudeste e 0 do contrário.
	Sul	1 para estudante da região sul e 0 do contrário.

Fonte: INEP (2017) e Brasil (2017).

Devido às inconsistências detectadas nas informações de alguns participantes do Enem e de algumas escolas foram necessários ajustes a base de dados. O primeiro deles foi a exclusão dos indivíduos e das escolas que não apresentaram o preenchimento correto das variáveis selecionadas. Também ocorreu a exclusão de alunos que apresentaram nota na redação igual a zero e dos que participaram do exame sem estarem no ano de conclusão do ensino médio.

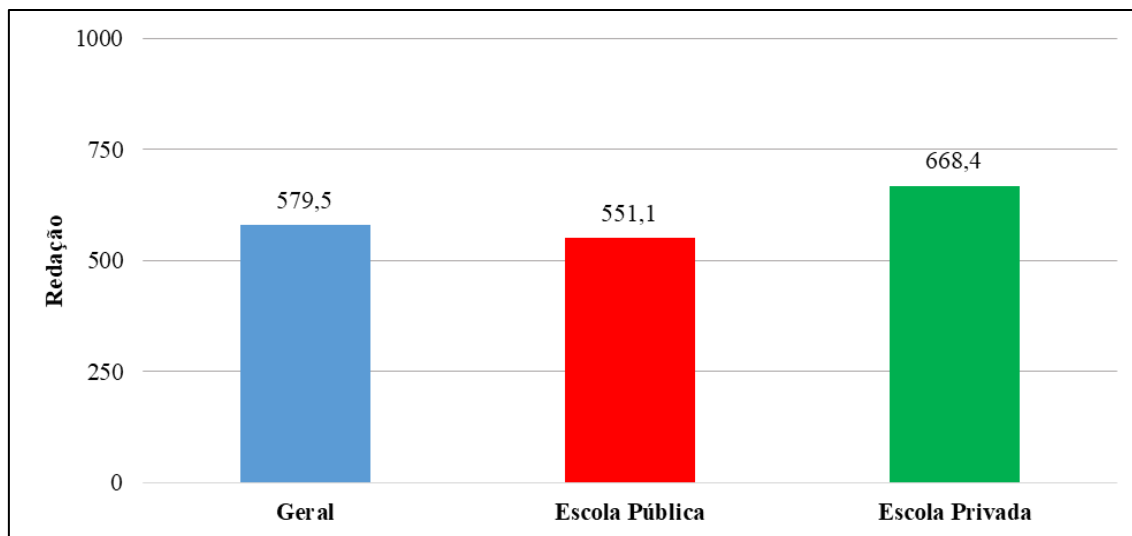
4. Resultados

4.1 Análise descritiva

As estimações tomam como medida para o desempenho educacional o logaritmo da nota na avaliação de redação que possibilita medir capacidades não captadas pelas avaliações que envolvem questões de múltipla escolha. O uso dessa variável apresenta um desempenho particular do aluno, diferentemente das demais avaliações, nas quais a performance do aluno depende do resultado obtido pelos demais participantes. Em Carvalho (2017), destaca-se que essa avaliação permite aferir o domínio da língua portuguesa, compreensão da proposta, seleção das informações, construção da argumentação e proposta de intervenção, o que denota maturidade para a vida acadêmica e profissional.

A Figura 1.1 apresenta a nota média para avaliação de redação no Enem, considerando toda a amostra o resultado foi de 579,5 pontos por aluno. Cabe destacar a diferença de notas por tipo de escola: enquanto o aluno de escola pública obtém nota média igual a 551,1, o aluno de escola privada apresenta 668,4 em média, o que equivale à diferença de 17,54%.

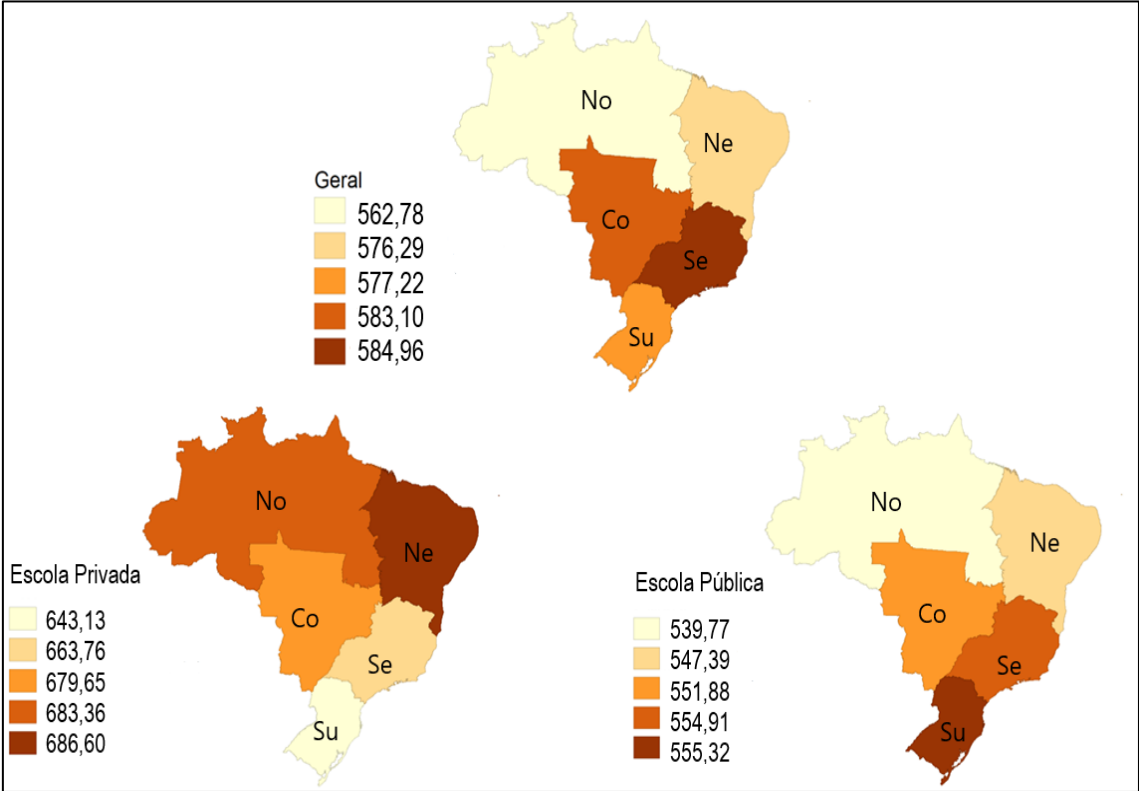
Figura 1.1 – Nota média na prova de redação do Enem, geral e por tipo de escola (Brasil, 2017)



Fonte: elaboração própria com base nos microdados do Enem 2017.

Outra observação importante que reforça o desequilíbrio educacional no Brasil está presente na figura 1.2, que apresenta o resultado médio na prova de redação por região e por tipo de escola. Nos mapas é possível observar que a região sudeste é a que apresenta o maior resultado médio na avaliação de redação do Enem. Esse resultado é 3,9% maior que a região norte (região com a menor média geral). Com relação às escolas públicas, a região sul apresenta desempenho mais expressivo, seguida por sudeste e centro-oeste. Por último, ao observar o desempenho regional das escolas privadas, a região nordeste apresenta-se em primeiro lugar, seguida por norte e centro-oeste.

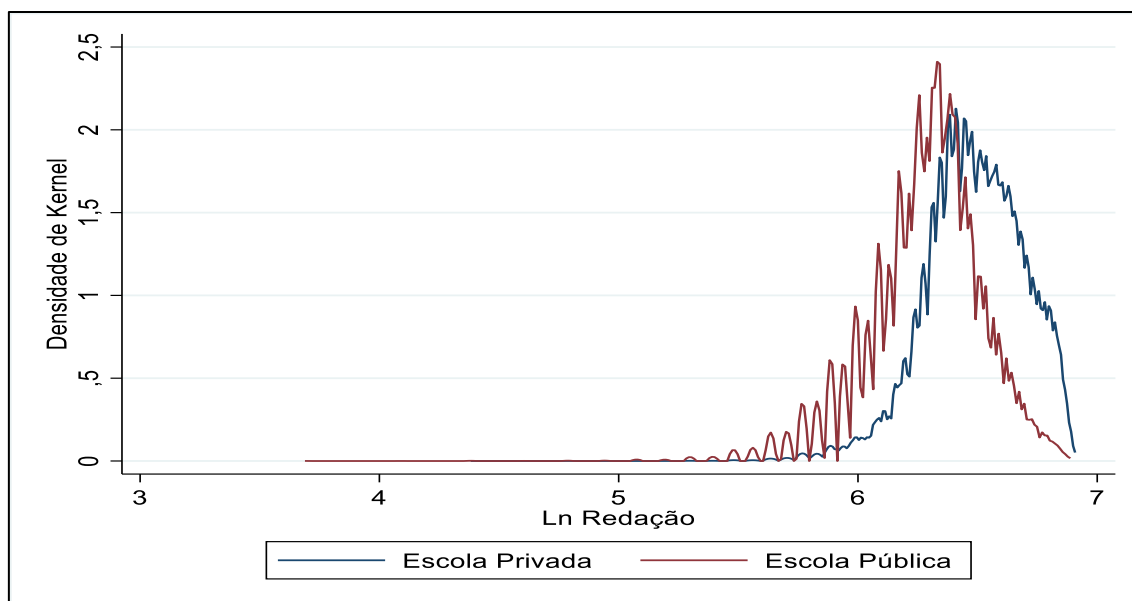
Figura 1.2 – Desempenho médio por região e por escola (Brasil, 2017)



Fonte: elaboração própria com base nos microdados do Enem 2017.

Reforçando as evidências quanto à existência de diferencial de desempenho entre escolas públicas e privadas, a Figura 1.3 apresenta a Densidade de Kernel para a distribuição logaritmo natural das notas em redação. O desempenho da escola privada (curva azul) está deslocado mais à direita, e mais próximo ao valor máximo (7), enquanto o desempenho da escola pública (curva vermelha), mais à esquerda, apresenta o ponto máximo próximo ao valor (6). Considera-se então que alunos da escola privada, em média, conseguem melhores resultados, ressaltando que não há controle nenhum sobre a variável resposta.

Figura 1.3 - Densidades Estimadas de Kernel do desempenho por tipo de escola (Brasil, 2017)



Fonte: elaboração própria com base nos microdados do Enem 2017.

Os valores médios das demais variáveis do banco de dados estão expostos na Tabela 1.1. Comparando-se as escolas pública e privada é perceptível que o aluno da primeira rede de ensino apresenta maior média de idade, maior participação do gênero feminino e são maioria autodeclarados “não-brancos”. Tanto o pai, quanto a mãe desse aluno apresentam menor escolaridade e sua família um maior número de membros. Outro indicativo de desigualdade está no fato da renda familiar do aluno da escola pública representar em média apenas 45,48% da renda familiar do aluno da escola privada.

A disponibilidade de computadores e internet na residência dos estudantes também reforçam a disparidade entre os estudantes. Dos alunos de escolas privadas, cerca de 90% dispõem de computadores em suas residências e 95% dispõem de internet. Esses percentuais, quando observados os alunos de escolas públicas, são reduzidos aproximadamente para 60% e 70% respectivamente.

Compete destacar outras informações sobre as oportunidades educacionais. Mais de três quartos dos alunos são oriundos de escolas públicas, desses menos de 3% estão em unidades federais. Das escolas localizadas em zona rurais a quase totalidade das são públicas. Dos alunos oriundos das escolas privadas, 18% receberam bolsa em algum momento do ensino médio em quanto na escola pública

esse percentual é reduzido para pouco mais de 9,5%. Ao mesmo tempo o percentual dos que cursaram o ensino médio em estabelecimentos privados com acesso à *internet* é o dobro dos matriculados em escolas públicas. Com relação às atividades complementares o percentual de alunos de escolas privadas que tiveram acesso à atividade complementares no ensino médio é 7,5 vezes maior do que o número de alunos de escolas públicas. Por último considerando de ensino técnico integrado ao ensino médio o percentual de alunos procedentes de escolas públicas é 25 vezes maior do que o percentual de alunos de escolas privadas.

Tabela 1.1 – Estatística descritiva das variáveis

Variável		Geral	Escola Pública	Escola Privada
Características Individuais	Idade (Anos)	17,45	17,51	17,29
	Mulher	58,55%	59,56%	55,36%
	Branco	43,05%	36,72%	62,96%
	Migração	37,21%	38,82%	32,15%
	Migração escolar	8,05%	7,14%	10,92%
Características Familiares	<i>Background</i> familiar	14,17	12,87	18,26
	Tamanho da família (Membros)	4,12	4,20	3,85
	Renda familiar (S/M)	5,22	4,05	8,92
	Computador residencial	66,83%	59,58%	89,66%
	Internet residencial	76,16%	70,23%	94,84%
Oportunidades Educacionais	Pública	75,89%	-	-
	Escola rural	2,97%	3,69%	0,72%
	Escola federal	3,33%	100%	0,0%
	Internet escolar	1,5%	1,15%	2,59%
	Bolsista	5,07%	9,55%	18,03%
	Ensino médio híbrido	5,13%	3,88%	9,08%
	Atividades complementares	2,88%	1,15%	8,34%
	Ensino técnico integrado	1,35%	1,76%	0,07%
Região	Norte	8,57%	9,48%	5,69%
	Nordeste	27,76%	28,97%	23,89%
	Centro-oeste	8,27%	8,24%	8,38%
	Sudeste	42,65%	40,68%	48,84%
	Sul	12,74%	12,60%	13,17%

Fonte: elaboração própria com base nos microdados do Enem 2017.

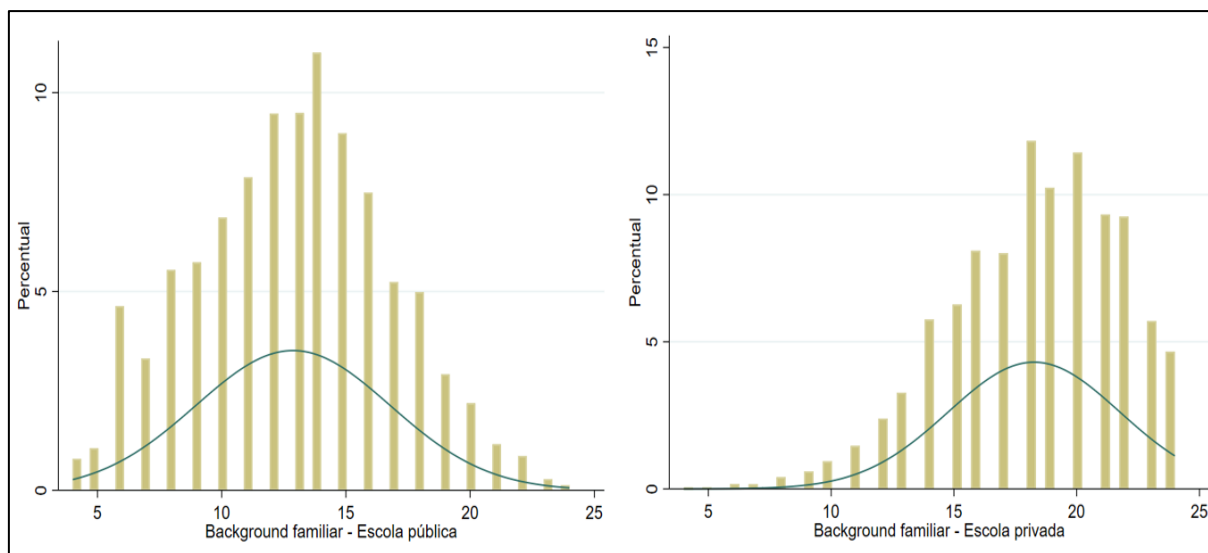
Pelos dados da Tabela 1.1, existe uma forte disparidade do *background* familiar que mensura as características educacionais e profissionais dos familiares dos estudantes. É perceptível que alunos de escolas privadas apresentam um valor

nesse indicador 46,4% maior que os alunos das escolas públicas. Essa discrepância é melhor entendida quando detalhadas as informações na figura 1.4.

Observando a distribuição do indicador de background familiar proposto é visível que as famílias de alunos de escola pública têm uma distribuição levemente assimétrica positiva onde a maioria dos dados está localizada abaixo da média da distribuição total. No caso dos alunos provenientes das escolas privadas a distribuição se caracteriza como assimétrica negativa com uma concentração maior na extremidade direita dos dados.

Essa desigualdade do indicador em desfavor dos alunos da escola pública ocorre devido a alguns fatos que podem ser destacados. Menos de 20% dos alunos da escola pública apresentam pelo menos um dos pais com ensino superior, enquanto na escola privada esse valor está acima de 60%. Além disso, quando observada a distribuição dos alunos por tipo de ocupação dos pais no mercado de trabalho fica reforçada a desigualdade de condições entre os alunos quando 67,25% dos pais dos oriundos da escola pública estão concentrados nos três grupos de menor status social e 82,71% dos genitores de alunos da escola privada estão reunidos nos dois grupos de maior estrato social.

Figura 1.4 - Distribuição do *background* familiar por tipo de escola (Brasil, 2017)



Fonte: elaboração própria com base nos microdados do Enem 2017.

4.2 Resultados das estimações

A partir dos resultados das estimações foi elaborada a Tabela 1.2 que apresenta os resultados pelos métodos de OLS e Regressão Quantílica. A Figura 1.5 expõe o comportamento dos coeficientes das variáveis utilizadas no modelo a partir da variação nos quantis, incluído o intercepto. Em cada gráfico está presente uma linha tracejada no centro que representa a regressão por OLS acompanhada dos respectivos intervalos de confiança ilustradas por linhas pontilhadas paralelas. A linha contínua representa a estimacão por quantil sombreada pelo intervalo de confiança.

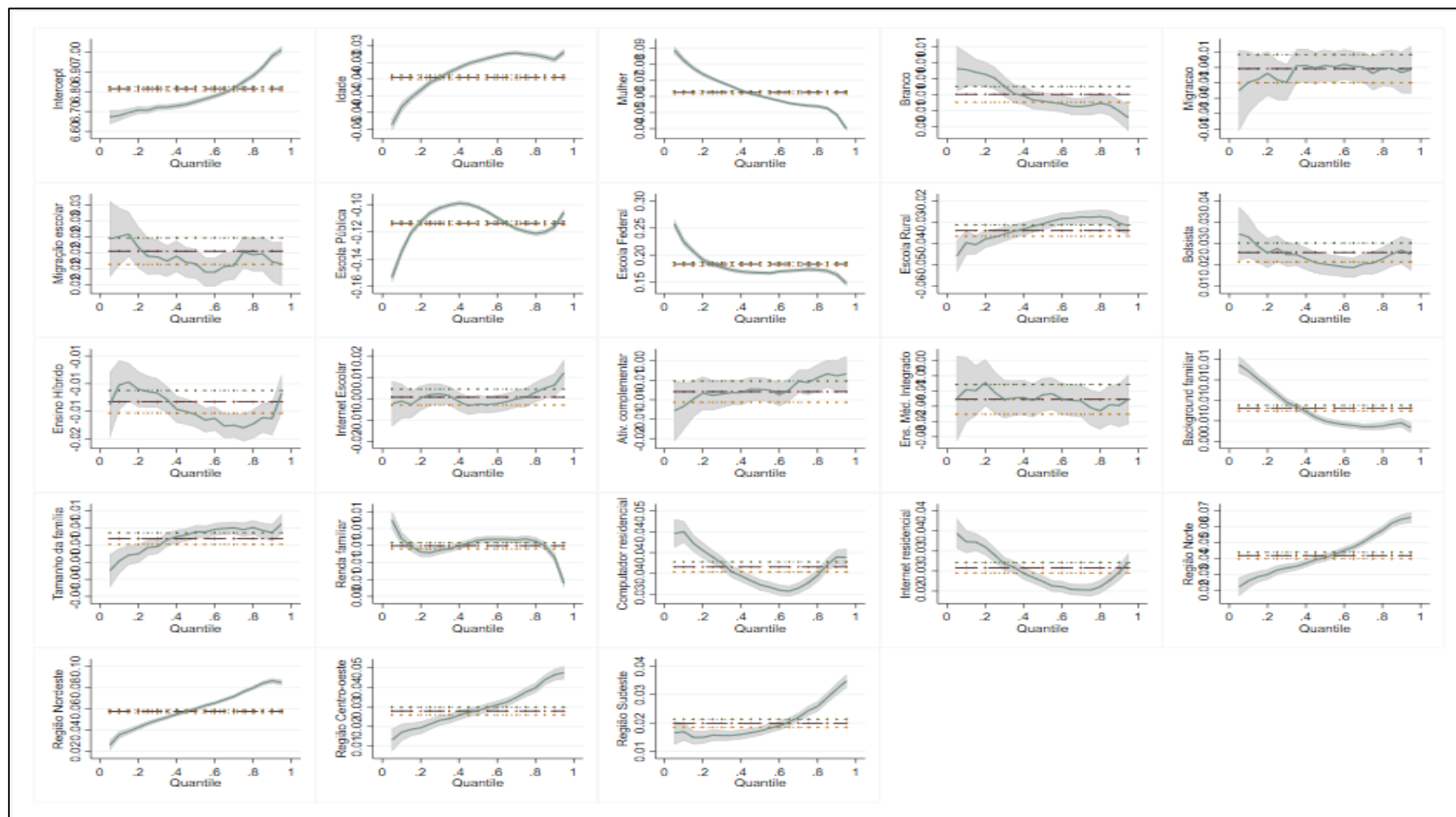
Tabela 1.2 – Estimacão do desempenho educacional (Brasil, 2017)

Variável		OLS	Regressão Quantílica		
			q10	q50	q90
Características Individuais	Idade	-0,0344*** (0,0003)	-0,0434*** (0,0007)	-0,0295*** (0,0004)	-0,0291*** (0,0005)
	Mulher	0,0625*** (0,0005)	0,0826*** (0,0010)	0,0603*** (0,0005)	0,0486*** (0,0007)
	Branco	0,0080*** (0,0005)	0,0111*** (0,0011)	0,0072*** (0,0006)	0,0059*** (0,0008)
	Migração	-0,0071*** (0,0005)	-0,0080*** (0,0010)	-0,0069*** (0,0005)	-0,0073*** (0,0007)
	Migração escolar	0,0222*** (0,0008)	0,0240*** (0,0019)	0,0206*** (0,0009)	0,0209*** (0,0013)
	Background Familiar	0,0056*** (0,0001)	0,0074*** (0,0002)	0,0050*** (0,0001)	0,0049*** (0,0001)
Características Familiar	Tamanho da Família	-0,0066*** (0,0002)	-0,0079*** (0,0004)	-0,0062*** (0,0002)	-0,0063*** (0,0003)
	Renda Familiar	0,0070*** (0,0001)	0,0074*** (0,0002)	0,0073*** (0,0001)	0,0063*** (0,0001)
	Computador residencial	0,0366*** (0,0006)	0,0450*** (0,0013)	0,0322*** (0,0007)	0,0389*** (0,0009)
	Internet residencial	0,0258*** (0,0007)	0,0323*** (0,0015)	0,0224*** (0,0007)	0,0247*** (0,0010)
Oportunidades Educacional	Pública	-0,1137*** (0,0007)	-0,1346*** (0,0016)	-0,1019*** (0,0008)	-0,1164*** (0,0011)
	Federal	0,1832*** (0,0017)	0,2241*** (0,0036)	0,1672*** (0,0018)	0,1639*** (0,0025)
	Rural	-0,0339*** (0,0013)	-0,0396*** (0,0029)	-0,0307*** (0,0015)	-0,0303*** (0,0020)
	Bolsista	0,0229*** (0,0011)	0,0266*** (0,0025)	0,0202*** (0,0012)	0,0235*** (0,0017)
	Ens. Méd. Híbrido	-0,0133*** (0,0010)	-0,0102*** (0,0023)	-0,0156*** (0,0012)	-0,0163*** (0,0016)

Ativ. Complementares		-0,0079*** (0,0014)	-0,0117*** (0,0031)	-0,0073*** (0,0015)	-0,0038* (0,0021)
Ens. Méd. Integrado		-0,0128*** (0,0025)	-0,0095* (0,0054)	-0,0113*** (0,0027)	-0,0147*** (0,0038)
Internet Escolar		0,0007 (0,0019)	-0,0011 (0,0042)	-0,0025 (0,0021)	0,0064*** (0,0029)
Região	Norte	0,0139*** (0,0011)	0,0259*** (0,0023)	0,0405*** (0,0012)	0,0648*** (0,0016)
	Nordeste	0,0295*** (0,0009)	0,0356*** (0,0018)	0,0599*** (0,0009)	0,0862*** (0,0013)
	Centro-oeste	(omitted) -	0,0170*** (0,0022)	0,0281*** (0,0011)	0,0464*** (0,0015)
	Sudeste	-0,0080*** (0,0009)	0,0169*** (0,0016)	0,0172*** (0,0008)	0,0319*** (0,0011)
	Sul	-0,0279*** (0,0010)	(omitted) -	(omitted) -	(omitted) -
	Constante	6,8421*** (0,0060)	6,6793*** (0,0131)	6,7493*** (0,0066)	6,9791*** (0,0091)
	R ²	0,2225	0,1046'	0,1006'	0,1199'

Fonte: elaboração própria a partir das estimações.
Notas: erro-padrão entre parênteses. * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$. 'Pseudo R²

Figura 1.5 – Comportamento dos coeficientes estimados das variáveis por quantil (Brasil, 2017)



Fonte: elaboração própria a partir das estimações.

Considerando o impacto das características individuais sobre o desempenho educacional, destaca-se inicialmente que a variável *Idade* apresenta efeito negativo sobre o desempenho educacional em todos os modelos, sendo os alunos do quantil 10 os que são mais afetados por essa variável.

Com relação à variável de gênero, o fato de o indivíduo ser do sexo feminino apresenta impacto positivo sobre a nota de redação. Esse comportamento é observado em todos os quantis como também no modelo OLS. Essa característica é observada nos dados através da nota média de participantes femininas no Enem é 4,44% maior que a média dos participantes masculinos.

Esse resultado é explicado por Reilly, Neumans e Andrews (2019), que afirmam que desde a infância meninas apresentam melhor capacidade de leitura e escrita em comparação aos meninos. Isso é devido a habilidades cognitivas naturais que interferem no nível de aprendizagem. Essas diferenças levam as meninas estarem à frente no desempenho educacional em todas as faixas etárias e em todos os países analisados. Considerando a amostra, apenas 17 participantes obtiveram nota máxima na redação, desses 70,58% eram do gênero feminino.

Ao analisar a variável de raça que considera o impacto do indivíduo “ser branco”, percebe-se que, no modelo OLS, há um efeito negativo, porém, não estatisticamente significativo, enquanto, nas estimações por quantil, todos os modelos foram significativos e apresentam impactos particulares. No quantil 10 o impacto é positivo sobre o desempenho na prova de redação, porém em magnitude menor do que o coeficiente estimado para o quantil 90, cujo efeito é negativo sobre o resultado do aluno. Em média, a diferença entre os grupos de indivíduos brancos e não-brancos é de 6,52%. O comportamento das variáveis que representam as características individuais (*idade*, *mulher* e *branco*) confirma os resultados observados por Nogueira e Figueiredo (2019), Carvalho (2017) e Paschoal (2005).

Analisando às duas variáveis que tratam do processo migração ressaltam-se efeitos opostos de maior intensidade na extremidade inferior da distribuição. É importante fazer a ressalva que a decisão de migrar nesse estudo compete a família do aluno. A primeira “*Migração*” trata dos alunos que residem em cidades diferentes daqueles em que nasceram, em todos os quantis os efeitos dessa decisão apresentam significância estatística e sinal negativo. Sobre a “*Migração escolar*”, onde o aluno estuda em cidade diferente da sua residência, em todos os pontos da distribuição o efeito é positivo e também significativo. Compreende-se, com base em Aguiar

(2019) e Rodrigues (2017), que a decisão migrar pode exigir dos indivíduos a um nível de esforço adicional dada a necessidade de adaptação a uma realidade diferente. Assim, os resultados podem apontar que alunos que precisam se deslocar para outras cidades para estudar assumem um custo de oportunidade maior exigindo a um esforço educacional maior o que pode resultar em um melhor desempenho.

A variável *background* familiar que representa um conjunto de características dos pais (escolaridade e ocupação) e a variável *Renda* contribuem positivamente para o desempenho educacional, conforme também observado em Annegues e Figueiredo (2016), Travitzki *et al.* (2016) e Riani e Rios-Neto (2008), Vernier, Bagolin e Jacinto (2015) e Mediavilla e Duque (2012). Esse efeito deve ser interpretado da mesma forma em todos os cenários, dado que, quanto maior o nível educacional e a atividade ocupacional do pai ou da mãe bem como o nível de renda familiar o desempenho educacional do filho tende a ser melhor. No presente estudo observou-se que a magnitude dos coeficientes aponta um peso maior no quantil inferior da distribuição.

A variável *Tamanho da família* representa o impacto do número de pessoas que moram com o aluno sobre o seu desempenho. Essa apresenta um coeficiente negativo em todos os modelos indicando que famílias com maior número de membros tendem a impactar negativamente o resultado no Enem. É possível deduzir que um número maior de pessoas ou de filhos na família pode reduzir o auxílio dos pais nas atividades educacionais dos filhos, bem como restringir as condições físicas, materiais e de espaço que um estudante necessita para estudar. Esse resultado corrobora com o estudo de Marteleto (2002) de que jovens em famílias grandes são mais penalizados em indicadores educacionais, além de destacar que famílias menores contribuem para o aumento dos anos de escolaridade.

As últimas características familiares condicionantes para o desempenho educacional estão relacionadas ao acesso tecnológico na residência do aluno através de *computador* e *internet*. Nesses casos, em todos os modelos estimados essas variáveis foram estatisticamente significativas e apresentaram efeitos positivos sobre o resultado da prova de redação do Enem. O acesso à informação e conteúdo educacional fora do ambiente escolar pode significar um reforço nas atividades de estudo, essa oportunidade pode excluir o suporte de outros familiares ou

orientações advindas da sala de aula. Dessa forma, expandir o acesso tecnológico a todas as camadas da população, em especial, às de rendas mais baixas, pode significar uma nova forma de acesso à educação ampliando suas oportunidades.

Das 8 variáveis que representam as características escolares, apenas duas, *Federal* e *Internet escolar*, não apresentaram efeito negativo sobre o desempenho no Enem. A variável *escola pública* é o principal condicionante do modelo para o resultado do aluno, porém destaca-se que esse tipo de ensino penaliza quem faz uso dele. Os alunos com melhor desempenho também sofrem esse impacto negativo, porém, menor em comparação aos demais quantis. Esse efeito da escola pública é ressaltado por De Sousa *et al.* (2020), Lobo, Cassuce e Cirino (2017) e Travitzki, Calero e Boto (2014).

Entende-se que o aluno no quantil 10 sofre um prejuízo maior pelo fato de ser oriundo de escola pública em comparação aos demais quantis da distribuição. Esse pior desempenho pode ser reflexo, conforme o relatório Interdisciplinaridade e Evidências no Debate Educacional/IEDE (2019), das limitações educacionais que o aluno de escola pública enfrenta em comparação ao aluno de escola privada. Apenas 34% dos estudantes da escola pública ingressam no ensino médio com o nível adequado de aprendizagem em português e apenas 15% em matemática. Em Feijó e França (2021) os resultados também apontam para um significativo diferencial entre as escolas privada e pública, que cresce à medida que se passa da extremidade inferior para a superior da distribuição quantílica.

O relatório destaca também a formação inadequada dos docentes: apenas 61,9% dos professores do ensino médio das escolas públicas apresentam formação superior nas áreas que lecionam. Além desses fatores podem ser ressaltadas as limitações de infraestrutura, recursos didáticos e recursos financeiros enfrentados pelas escolas de ensino médio no Brasil. Soma-se a esse contexto as desigualdades regionais existentes no país, cabe destacar que as unidades de ensino médio são de responsabilidade dos governos estaduais, o que torna ainda mais heterogênea a escola pública brasileira e dificulta a articulação de ações nacionais.

Com relação à variável *Federal*, que representa o efeito de escolas públicas federais de ensino médio, observa-se que esse tipo de escola gera um efeito positivo em todos os quantis da distribuição. Esse efeito também é destacado em Dutra *et al.* (2019) e Dutra *et al.* (2019) e cabe a ressalva de que essa é a variável que apresenta o coeficiente estimado de maior magnitude no presente estudo. Ou seja, estudar em

uma escola federal de ensino médio aumenta a chance de um melhor desempenho educacional, isso pode ser justificado por algumas características dessas unidades de ensino. Em sua maioria essas escolas federais exigem para seu ingresso um processo seletivo, selecionando assim alunos que já apresentam um desempenho melhor no ensino fundamental. Outra justificativa pode estar no fato de apresentarem um melhor quadro docente e uma maior capacidade orçamentária conforme destacado em Sousa *et al.* (2019).

A escola *rural* segue a mesma tendência de impacto negativo, mesmo não sendo significativa para os alunos de melhor desempenho, sendo estatisticamente significativa para o quantil 50 e principalmente para o quantil 10. É perceptível que o meio rural brasileiro ainda apresenta limitações de acesso a serviços, o que podem contribuir para esse efeito. Apesar de uma amostra pequena de alunos de escolas rurais, esse resultado destaca uma realidade recorrente no Brasil e também é destacado em Rodrigues (2017). É comum que muitos adolescentes que vivem no meio rural se desloquem para o meio urbano de modo a cursar o ensino médio. Esse contexto pode ser um desincentivo as atividades educacionais.

Outro condicionante significativo estatisticamente e, com efeito, positivo sobre o desempenho educacional foi observada na variável *Bolsista*. Nesse caso o aluno recebe um auxílio financeiro que lhe permite custear as despesas de uma escola privada ou uma remuneração que lhe permite se dedicar aos estudos na escola pública. É possível destacar que a adoção de bolsas de estudos permite ao aluno beneficiado a oportunidade de direcionar uma parcela maior de tempo para a vida estudantil. Além disso, a continuidade do benefício quando atrelada a manutenção ou melhoria do desempenho educacional pode criar uma motivação extra para o aluno conforme Harkreader (2008) e Figlio e Karbownik (2016).

Outras oportunidades educacionais também foram estimadas. A primeira foi o caso de alunos que estudaram uma parte do ensino médio em escola pública, parte em escola privada. A segunda possibilidade observada foi a existência de *atividades complementares* para os alunos de ensino médio. Por último, a possibilidade de o aluno cursar o *ensino técnico integrado* ao ensino médio regular. Em todos os quantis observados essas possibilidades tiveram efeitos negativos sobre o desempenho educacional no exame de redação no Enem e foram estatisticamente significativos. Esses resultados vão ao encontro do estudo de Arraes e Mariano (2019) que destaca o impacto positivo do ensino profissionalizante sobre o

desempenho educacional para o caso de escolas do estado do Ceará. Cabe a ressalva que devido à heterogeneidade das escolas brasileiras, a ampliação da amostra pode alterar a análise.

O *ensino médio híbrido*, pode ser entendido por duas óticas, no caso do aluno que troca a escola privada pela pública ou no caso da troca inversa. No primeiro caso pode ocorrer de a unidade educacional pública anular o efeito positivo trazido pelo aluno da escola privada e no caso contrário o estabelecimento privado reduz o efeito do tipo de escola. O efeito negativo das outras possibilidades pode estar relacionado a redução da dedicação dos alunos aos conteúdos exigidos no exame nacional, dado que agora eles estão inseridos em outras atividades.

Relativamente à disponibilidade de acesso à *internet* para os alunos foi observado o efeito dessa variável se altera conforme caminha na distribuição, saindo de um coeficiente negativo no quantil 10 para um coeficiente positivo no quantil 90. Porém apenas o quantil da extremidade superior da distribuição apresentou significância estatística.

Por último foi estimado os efeitos das regiões como condicionantes para o desempenho educacional aplicado nesse estudo foram as características regionais, identificando a região do país da escola. As estimações apresentaram coeficientes positivos e estatisticamente significativos para todas as regiões, exceto a região sul omitida nas estimações.

Para comprovar que o modelo de regressão quantílica permite avaliar o comportamento dos coeficientes estimados em momentos diferentes da distribuição a Tabela 1.3 apresenta a aplicação o Teste de Wald proposto por Koenker e Machado (1999). O poder do teste está em aferir se os efeitos das variáveis diferem nos quantis. Para isso, utiliza-se a hipótese nula de proposição de que efeitos dos coeficientes estimados são iguais para todos os quantis. Conforme a Tabela 3, a hipótese nula foi fortemente rejeitada para todas variáveis testadas, exceções feitas as variáveis “*Migração*” e “*Ensino Técnico Integrado*”. Isso significa que existe diferença estatisticamente significativa entre os coeficientes nos quantis Q10, Q50 e Q90.

Tabela 1.3 – Teste de Wald

Grupo	Variável	Teste de Wald	Hipótese nula ($\beta_{1q10} = \beta_{1q50} = \beta_{1q90}$)
Características Individuais	Idade	395,69	Rejeita
	Sexo	245,71	
	Raça	8,46	
	Migração	0,66*	
	Migração escolar	3,62	
Características Familiares	<i>Background</i> familiar	161,60	
	Tamanho da família	12,96	
	Renda familiar	30,75	
	Computador residencial	109,13	
	Internet residencial	28,84	
Oportunidades Educacionais	Escola pública	245,47	
	Escola rural	6,03	
	Escola federal	436,47	
	Internet escolar	6,88	
	Bolsista	5,42	
	Ensino médio híbrido	4,80	
	Atividades complementares	3,15	
	Ensino técnico integrado	1,23*	
Região	Norte	146,18	
	Nordeste	199,37	
	Centro-oeste	49,55	
	Sudeste	100,13	
	Sul	(omitted)	

Fonte: elaboração própria a partir das estimações.

Nota: estatística para todas as variáveis = F(2,912753).

*Não rejeita a hipótese nula.

Apesar da validade do teste de Wald, o modelo de regressão quantílica apresentou problema de endogeneidade para a variável *escola pública*, conforme o teste J de restrição de sobre-identificação¹¹. O teste avalia a exogeneidade da variável em questão e utiliza a hipótese nula de que os estimadores são exógenos. Ante o exposto, não é possível assumir que a variável é exógena, o que implica em endogeneidade.

Considerando que esse estudo se baseia em dados individuais dos participantes do Enem, há um limitado número de variáveis o que torna inviável o uso de modelos tradicionais de regressão quantílica com variável instrumental. Essa

¹¹ Teste J (*p-valor*): 711,301

condição sugere o uso do modelo de seleção de amostra com base em regressões quantílicas extremas proposta por D'Haultoeuille, Maurel e Zhang (2014) para solução do problema de endogeneidade.

A nova estimação utilizou o mesmo banco de dados e as mesmas variáveis adotadas anteriormente. O diferencial está em utilizar um processo de *bootstrap* para determinar os intervalos de confiança para as estimativas dos parâmetros através de um processo de reamostragem do banco de dados original. Isso resulta em uma estimação mais robusta tornando mais preciso o coeficiente da variável de interesse. A Tabela 1.4 apresenta os resultados da estimação para os quantis 10 e 90.

Os resultados indicam que o fato de o aluno ser oriundo de escola pública resulta em penalização de -13,19% no desempenho na prova de redação do Enem, sendo esse valor estatisticamente significativo para o quantil 10. O *p-valor* do Teste J implica que não se pode rejeitar a especificação em nenhum nível estatístico, sendo assim, considera-se a hipótese nula de que as variáveis são exógenas. É pertinente comparar os resultados obtidos na seção anterior via regressão quantílica e o mesmo conjunto de variáveis. A regressão de quantis extremos estimou um impacto negativo e levemente menor da escola pública sobre o desempenho aluno superior em comparação aos -13,46% do modelo de regressão quantílica tradicional. Isso indica sobrestimação de 2,01% entre os modelos.

Tabela 1.4 – Estimação via Regressão de quantis extremos

Informações	Quantil 10	Quantil 90
Escola pública	-0,1319*** (0,0017)	-0,1379*** (0,0025)
Teste J	0,0157	0
Intervalo de confiança (95%)	(-0,1353; -0,1285)	(-0,1428; -0,1330)

Fonte: elaboração própria a partir das estimações.
Nota: erro-padrão entre parênteses. * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

Para o grupo de alunos do quantil 90, a estimação pela regressão de quantis extremos apresentou efeito negativo da escola pública de -13,79% conforme a Tabela 1.4, o que representa que a estimação quantílica tradicional sobestimou a variável em 13,3%. A estimação anterior, presente na Tabela 1.4, expôs que a escola pública reduz a nota dos alunos de melhor desempenho em 11,64% e, assim como no modelo anterior, a hipótese nula não pode ser rejeitada, garantindo a validade

do modelo. O novo modelo penaliza ainda mais o aluno de melhor de desempenho quando esse é oriundo de escola pública. Em outras palavras, exige um esforço mais expressivo desse aluno para melhorar seu desempenho.

De maneira geral, é possível destacar que, pela regressão quantílica tradicional, o quantil 90 apresenta coeficientes estimados menores, em todas as variáveis, em comparação com o quantil 50, a média e, principalmente, em relação ao quantil 10. Isso sugere que características individuais, o *background* familiar e as características da escola têm implicação menor para os alunos de melhores notas. Em sentido oposto os coeficientes do quantil 10, com os alunos de pior desempenho, são superiores às estimações pela média, mediana e pelo quantil 90.

O fato de os alunos de melhor desempenho apresentarem menores coeficientes pode ser um indicativo de que o esforço educacional desses pode compensar essas limitações impostas pelo contexto familiar, social e educacional, ou seja, o componente de esforço individual se sobressai aos demais condicionantes. No caso dos alunos de pior desempenho, esses são fortemente impactados pelos condicionantes, o que exigiria um esforço individual muito expressivo para superar esses limitadores do seu desempenho ou então que necessitam mais de fatores externos para conseguir progredir na questão educacional. Esses fatores externos poderiam ser representados pela escola, ambiente que deveria compensar a realidade social e o *background* familiar e potencializar o aprendizado.

Os resultados evidenciam que os coeficientes estimados do impacto da escola pública entre o grupo de melhor e pior desempenho não apresentam disparidade expressiva, porém, o que é mais importante, nessa análise, é identificar que os alunos da escola pública têm seus desempenhos penalizados devido à origem educacional. Em outras palavras, é possível indicar que a escola pública no Brasil apresenta uma contribuição negativa sobre o desempenho educacional. Convém destacar que esse efeito se torna mais expressivo para os alunos que obtêm melhores notas no Enem em comparação aos alunos de menores notas no exame.

A regressão quantílica tradicional conseguiu identificar o sentido do impacto da escola pública sobre o desempenho educacional dos alunos no Enem, porém, considerando o problema da endogeneidade da variável “escola pública”, é a regressão via quantis extremos que torna robusta a estimação para os quantis 10 e 90.

5. Considerações finais

Um país com fortes investimentos públicos educacionais, mas que apresenta um desempenho educacional desigual tende a não permitir a população, em especial, às crianças e jovens, um ambiente de igualdade em oportunidades. Entretanto, a indicação de caminhos possíveis e soluções plausíveis deve se basear em evidências sólidas que levem a construção de políticas educacionais efetivas.

O presente trabalho avaliou o impacto da escola pública brasileira nos diferentes intervalos de desempenho educacional dos alunos no Enem. Buscando oferecer uma contribuição a literatura foi construído um banco de dados que agregava informações dos alunos, suas famílias e escolas a partir dos microdados de Enem e do Censo Escolar. Em especial podem ser destacadas variáveis que captaram questões relativas à migração escolar, a proposição de um indicador para mensurar o *background* familiar. Do mesmo modo, está presente a análise da disponibilidade de internet residencial e escolar, o efeito de incentivos através de bolsas, atividades complementares e o acesso a escolas federais.

Considerando-se as estimações realizadas evidenciou-se que a escola pública afeta de maneira semelhante os alunos de melhor e pior desempenho. Do mesmo modo, há um significativo diferencial de desempenho em desfavor dos participantes do Enem oriundos de escolas públicas quando comparados aos participantes de escolas privadas. Essas diferenças identificadas se expandem quando se observa que o peso das variáveis, as quais os alunos não controlam, se altera na distribuição por quantil desfavoravelmente aos alunos de pior desempenho.

Especificamente, indivíduos mais velhos, não-brancos, homens, de escola pública e rural, sem acesso à internet tendem a apresentar desempenho insatisfatório na avaliação de redação do Enem. Os alunos do quantil 90, com melhor desempenho, são menos vulneráveis às condições relacionadas a características individuais, *background* familiar e escolar, o que pode sugerir que características de esforço individual sejam relevantes nesse caso.

Para superar o problema da engoneidade foi utilizado um modelo de seleção de amostra via regressão quantílica extrema (D'haultoeuille; Maurel; Zhang, 2014) que possibilita manter a análise por quantil. Outra vantagem dessa proposta está em apresentar uma estimação robusta sem o uso de variáveis instrumentais, o que se tornou adequada a esse estudo devido à limitação de opções na base de dados original. O uso dessa estratégia tornou possível identificar que o coeficiente da

variável “escola pública” apresentava erros de mensuração, mais expressivo no quantil superior da amostra. Observou-se que havia sobrestimação do coeficiente estimado tanto para o grupo de alunos do quantil quanto para o grupo de alunos do quantil 90.

A escola pública deveria ser um espaço que pudesse minimizar o peso de condicionantes, tais como o *background* familiar baixo, especialmente, para alunos de desempenho insatisfatório de aprendizagem. Porém, o que foi observado é que a escola pública contribui negativamente para o desempenho do aluno, ou seja, ao invés de potencializar o desempenho do aluno, a escola se apresenta como obstáculo a se superar para obtenção de melhores resultados. Assim, é possível afirmar que alunos de escola pública e privadas encontram-se em condições diferentes, levando à desigualdade de oportunidades, sendo isso potencializado por fatores socioeconômicos e pelo contexto familiar.

Por fim, percebe-se que a escola pública de ensino médio no Brasil, com base nos dados analisados, não contribui de forma efetiva para o melhorar desempenho educacional de seus alunos. Dessa forma, torna-se necessária a adoção de políticas que permitam potencializar o processo educacional para essa tornar-se capaz de contribuir efetivamente no processo de aprendizagem.

Com base nos resultados algumas políticas podem contribuir para a melhoria do desempenho dos alunos. No curto prazo é possível indicar a oferta de bolsas de estudo que permitam a dedicação exclusiva aos estudos ou ao acesso a escolas privadas e a ampliação do acesso a computadores e internet, tanto no ambiente escolar como nas residências. No longo prazo é possível sugerir expandir o acesso de escolas públicas federais para um maior número de alunos.

Dentre as limitações desse trabalho cabe descartar a indisponibilidade no banco de dados do Enem de variáveis de esforço individual ou estratégias educacionais como ensino integral, reforço escolar ou curso preparatório que possam compensar desigualdade de oportunidades. Como propostas para continuidade da pesquisa, podem ser sugeridas duas vertentes. A primeira consiste na inclusão de variáveis que diferenciem os tipos de escolas públicas, de modo a mensurar seus impactos sobre o desempenho educacional e a segunda consiste em captar como a escola pública impacta na mobilidade intergeracional.

É importante destacar que no ensino público existem experiências que se apresentam como casos reconhecidos de bom desempenho em exame de

aprendizagem, como as escolas militares e os institutos federais. Essas poderiam servir de *benchmark* para o sistema educacional brasileiro, indicando experiências que poderiam ser copiadas para potencializar o desempenho educacional.

Referências

- AGUIAR, M. A. S. de. Ensaios sobre migração e deslocamento pendular. 2021. 131f. Tese (Doutorado em Economia). Faculdade de Economia, Administração, Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Universidade Federal do Ceará. Fortaleza, 2019.
- ANDERSON, G.; LEO, T. W.; MUELHAUPT, R. *Qualified equal opportunity and conditional mobility: gender equity and educational attainment in Canada. Working Papers Tecipa*, n. 368. University of Toronto, Department of Economics, 2009.
- ANDERSON, G.; LEO, T. W. Quantifying the progress of economic and social justice: charting changes in equality of opportunity in the USA, 1960-2000. *Journal of Human Development and Capabilities*, v. 18, n. 1, p. 17-45, 2017.
- ANNEGUES, A. C.; FIGUEIREDO, E. Mobilidade intergeracional qualificada: uma abordagem de mensuração utilizando regressões quantílicas. *Economia Aplicada*, v. 20, n. 1, 2016.
- ARRAES, R. A.; MARIANO, F. Z. Decomposição quantílica incondicional dos diferenciais de desempenho entre alunos de escolas privadas e públicas profissionalizantes. *Pesquisa e Planejamento Econômico*, v. 49, n. 3, p. 29-80, 2019.
- BARRO, R. J. Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, v. 106, n. 2, p. 407-443, 1991.
- BENEVIDES, A. A.; SOARES, R. B. Diferencial de desempenho de alunos das escolas militares: o caso das escolas públicas do Ceará. *Nova Economia*, v. 30, n. 1, p. 317-343, 2020.
- BRASIL. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. Brasília: Censo Escolar, 2017.
- CARVALHO, F. E. D. Fatores socioeconômicos associados ao desempenho dos estudantes na prova de redação do Exame Nacional do Ensino Médio (ENEM). 2017. 179f. Tese (Doutorado em Economia). Programa de Pós-graduação em Educação Brasileira, Universidade Federal do Ceará. Fortaleza, 2017.

CHERNOZHUKOV, V. *et al.* Extremal quantile regression. *The Annals of Statistics*, v. 33, n. 2, p. 806-839, 2005.

CHERNOZHUKOV, Victor; FERNÁNDEZ-VAL, Iván; KAJI, Tetsuya. Extremal quantile regression. *Handbook of Quantile Regression*, p. 333-362, 2017.

CHETTY, R. *et al.* Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, v. 129, n. 4, 2014.

COUTO, A. P. Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: uma visão a partir da análise de dados do Enem. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, v. 24, p. 1-32, 2016.

DAOUIA, A. *et al.* On kernel smoothing for extremal quantile regression. *Bernoulli*, v. 19, n. 5B, p. 2557-2589, 2013.

DELGADO, V. M. S.; RIBEIRO, A. M.; SOARES, J. F. Desigualdade escolar e desempenho. In: FAHEL, M.; RAMBLA, X.; COSTA, B. L. D.; BRONZO, C. *Desigualdades educacionais & pobreza*. Belo Horizonte: PUC Minas, 2012.

DE SOUSA, E. C. *et al.* Desigualdades em discussão: o rendimento das escolas públicas e privadas no Exame Nacional do Ensino Médio (ENEM) no ano de 2017/Disequalities in discussion: the performance of public and private schools in the national high school (ENEM) exam in 2017. *Brazilian Journal of Development*, v. 6, n. 5, p. 26450-26458, 2020.

DIAZ, M. D. M. *et al.* (Des)igualdades de oportunidades no ensino médio brasileiro: escolas públicas e privadas. *Economía*, v. 13, n. 3a, p. 553-568, 2012.

D'HAULTFŒUILLE, X. *et al.* Estimating selection models without an instrument with Stata. *The Stata Journal*, v. 20, n. 2, p. 297-308, 2020.

D'HAULTFŒUILLE, X.; MAUREL, A.; ZHANG, Y. Extremal quantile regressions for selection models and the black-white wage gap. *Journal of Econometrics*, v. 203, n. 1, p. 129-142, 2018.

D'HAULTFŒUILLE, X.; MAUREL, A.; ZHANG, Y. Extremal quantile regressions for selection models and the black-white wage gap. *National Bureau of Economic Research*, 2014.

DUTRA, R. S. *et al.* Determinantes do desempenho educacional dos Institutos Federais do Brasil no Exame Nacional do Ensino Médio. *Educação e Pesquisa*, v. 45, 2019.

DUTRA, R. S. *et al.* Indicadores educacionais e proficiência no ENEM: um estudo nos Institutos Federais do Brasil. *Revista Meta: avaliação*, v. 11, n. 31, p. 124-153, 2019.

DUTRA, R. S. *et al.* O que mudou no desempenho educacional dos Institutos Federais do Brasil? *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 27, p. 631-653, 2019.

FEIJÓ, J. R.; FRANÇA, J. M. S. de. Diferencial de desempenho entre jovens das escolas públicas e privadas. *Estudos Econômicos*, v. 51, p. 373-408, 2021.

FERNANDES, I. F. A. L. Apêndice B - regressão quantílica. In: FERNANDES, I. F. A. L. A democracia reduz a desigualdade econômica? Um estudo sobre as possibilidades de construção de uma sociedade mais igual por meio da democracia. São Bernardo do Campo: Editora UFABC, 2017.

FERTIG, M. Who's to blame? The determinants of German students' achievement in the PISA 2000 study. Discussion paper, n. 4, *RWI*, 2003.

FIGLIO, D.; KARBOWNIK, K. Evaluation of Ohio's EdChoice scholarship program: selection, competition, and performance effects. Thomas B. Fordham Institute, 2016.

FIGUEIREDO, E.; LIMA, L. R.; SCHAUR, G. Robust estimation of gravity equations and the WTO impact on trade inequality. Knoxville: University of Tennessee, 2014. Working paper.

FIGUEIRÊDO, Erik; NOGUEIRA, Lauro; SANTANA, Fernanda Leite. Igualdade de oportunidades: Analisando o papel das circunstâncias no desempenho do ENEM. *Revista Brasileira de Economia*, v. 68, n. 3, 2014.

FINN, A.; LEIBBRANDT, M.; RANCHHOD, V. *Patterns of persistence*: intergenerational mobility and education in South Africa. Working paper, n. 175. Cape Town: SALDRU, 2016.

GRAWE, N. *Primary and secondary school class size and intergenerational earnings mobility*. Working paper, n. 2007-03. Carleton College Department of Economics, p. 1-32, 2007.

HANUSHEK, E. A. Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, p. 351-388, 1979.

HANUSHEK, E. A. *School resources and student performance*. Does money matter? The effect of school resources on student achievement and adult success. Massachusetts: Brookings Institution Press, 1996. p. 43-73.

HANUSHEK, E. *The production of education, teacher quality and efficiency*. Washington: U. S. Office of Education, 1970.

HANUSHEK, E. A.; WÖßMANN, L. The Role of School Improvement in Economic Development. NBER Working Paper No. 12832. Cambridge (MA): National Bureau of Economic Research, 2007.

HARKREADER, S. *et al.* The impact of Florida's bright futures scholarship program on high school performance and college enrollment. *Journal of Student Financial Aid*, v. 38, n. 1, p. 1, 2008.

HECKMAN, J. J.; MASTEROV, D. V. Productivity argument for investing in young children. Working paper, n. 5. *Invest in Kids Working Group Committee for Economic Development*, 2004.

HILD, M.; VOORHOEVE, A. Roemer on equality of opportunity. California Institute of Technology, Social Science Working Paper, v. 1128, 2001.

IEDE - Interdisciplinaridade e Evidências no Debate Educacional. Excelência com equidade no ensino médio: a dificuldade das redes de ensino para dar um suporte efetivo às escolas. Portaliede, 2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Microdados do Exame Nacional do Ensino Médio - ENEM 2017 [recurso eletrônico]. Brasília, DF: INEP, 2018.

JIMÉNEZ, M.; JIMÉNEZ, M. Intergenerational educational mobility in Latin America: an analysis from the equal opportunity approach. *Cuadernos de Economía*, v. 38, n. 76, p. 289-329, 2019.

KAN, K.; LI, I.-H.; WANG, R. Intergenerational income mobility in Taiwan: evidence from TS2SLS and structural quantile regression. *The BE Journal of Economic Analysis & Policy*, v. 15, n. 1, p. 257-284, 2015.

KIM, H. Intergenerational mobility and the role of education in Korea. *Korean Development Institute Workshop on Human Capital Policy*. Seoul, 2014.

KOENKER, R.; BASSETT, G. Quantile regression. *Econometrica*, v. 46, p. 33-50, 1978.

KOENKER, R.; HALLOCK, K. Quantile regression: an introduction. *Journal of Economic Perspectives*, v. 15, n. 4, p. 43-56, 2001.

- KOENKER, R.; MACHADO, J. AF. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, v. 94, n. 448, p. 1296-1310, 1999.
- LOBO, G. D.; CASSUCE, F. C. C.; CIRINO, J. F. Avaliação do desempenho escolar dos estudantes da região nordeste que realizaram o Enem: uma análise com modelos hierárquicos. *Revista Espacios*, v. 6, 2017.
- LÖBLER, M. L. *et al.* Acesso e uso da tecnologia da informação em escolas públicas e privadas de ensino médio: o impacto nos resultados do ENEM. *Sistemas & Gestão*, v. 5, n. 2, p. 67-84, 2010.
- MANKIW, N. G.; ROMER, D.; WEIL, D. N. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, v. 107, n. 2, p. 407-437, 1992.
- MARCONI, G. Give it Time: Education Affects Economic Growth in the Long Term. SSRN Electronic Journal, 2016.
- MARTELETO, L. J. O papel do tamanho da família na escolaridade dos jovens. *Revista Brasileira de Estudos de População*, v. 19, n. 2, p. 159-177, 2002.
- MEDIAVILLA, M.; DUQUE, L. G. Un análisis de los condicionantes del rendimiento académico en Brasil a partir del Saeb-2005. In: FAHEL, M. *et al.* (Orgs.). *Desigualdades educacionais & pobreza*. Belo Horizonte: PUC Minas, 2012.
- NEIDHÖFER, G.; SERRANO, J.; GASPARINI, L. Educational inequality and intergenerational mobility in Latin America: a new database. *Journal of Development Economics*, v. 134, p. 329-349, 2018.
- NOGUEIRA, L. C. B.; FIGUEIREDO, E. Tal pai, tal filho? Uma análise dos efeitos fatores de circunstâncias sobre o desempenho dos alunos na avaliação do PISA 2012. *Pesquisa e Planejamento Econômico*, v. 49, 2019.
- PARMAN, J. American mobility and the expansion of public education. *The Journal of Economic History*, v. 71, n. 1, p. 105-132, 2011.
- PASCHOAL, I. P. *Mobilidade intergeracional de educação no Brasil*. 2005. Tese (Doutorado em Economia). Programa de Pós-Graduação em Economia, Universidade de São Paulo. São Paulo, 2005.
- REEVES, E. B.; LOWE, J. Quantile regression: an education policy research tool. *Journal of Rural Social Sciences*, v. 24, n. 1, p. 10, 2009.

REILLY, D.; NEUMANN, D. L.; ANDREWS, G. Gender differences in reading and writing achievement: evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, v. 74, n. 4, p. 445, 2019.

RIANI, J. L. R; RIOS-NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros? *Revista Brasileira Estudos Populacionais*, v. 25, n. 2, p. 251-269, 2008.

ROCHA, L. A. *et al.* O impacto dos investimentos em P&D no desempenho das empresas: aplicações no uso de regressão quantílica com variáveis instrumentais. *Revista de Economia Contemporânea*, v. 22, n. 3, e182235, 2018.

ROEMER, J. E. *Egalitarian perspectives: essays in philosophical economics*. Cambridge: Cambridge University Press, 1996.

RODRIGUES, L. O. *Ensaio sobre diferencial de desempenho escolar entre alunos de escolas rurais e urbanas no Brasil*. 2017. 99 f. Dissertação (Mestrado em Economia Rural). Centro de Ciências Agrárias, Universidade Federal do Ceará. Fortaleza, 2017.

ROMER, P. M. The origins of endogenous growth. *Journal of Economic perspectives*, v. 8, n. 1, 1994.

ROTHBAUM, J. Parent characteristics and the geography of mobility. In: *Biennial Federal reserve system community development research conference*, 9. Washington. *Anais [...]* Washington, 2015.

SANTOS SILVA, J.; TENREYRO, S. The log gravity. *The Review of Economics and Statistics*, v. 88, n. 1, p. 641-658, 2006.

SOUSA, L. A. D. *et al.* Desempenho das Instituições Federais de Educação Profissional, Científica e Tecnológica Brasileiras no ENEM. *Tendências pedagógicas*, 2019.

TAKEUCHI, I. *et al.* Nonparametric quantile estimation. *Journal of Machine Learning Research*, v. 7, p. 1231-1264, 2006.

TRAVITZKI, Rodrigo; CALERO, Jorge; BOTO, Carlota. What does the national high school exam (ENEM) tell brazilian society?. *Cepal Review*, v. 2014, n. 113, p. 157-174, 2014.

TRAVITZKI, Rodrigo; FERRÃO, Maria Eugénia; COUTO, Alcino Pinto. Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: uma visão a partir da análise de dados do ENEM. *Education Policy Analysis Archives*, v. 24, 2016.

VERNIER, L. D. S.; BAGOLIN, I. P.; JACINTO, P. Fatores que influenciam o desempenho escolar no Estado do Rio Grande do Sul: uma análise com regressões quantílicas. *Análise Econômica*, v. 33, n. 64. 2015.