

# Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM

Erik Figueirêdo\*, Lauro Nogueira†, Fernanda Leite Santana‡

**Conteúdo:** 1. Introdução; 2. Igualdade de Oportunidades: Uma Breve Revisão; 3. Procedimentos Teóricos e Metodológicos; 4. Resultados; 5. Considerações Finais.

**Palavras-chave:** Desigualdade de Oportunidades; ENEM.

**Códigos JEL:** J31;C31;J62.

Este estudo avalia como as circunstâncias sociais influenciam o desempenho educacional dos alunos que prestam o Exame Nacional do Ensino Médio (ENEM). Serão utilizadas duas estruturas teóricas. Inicialmente, adota-se uma abordagem não paramétrica assumindo o Roemer Identification Axiom, isto é, assumimos a independência entre circunstâncias e esforço. Em seguida, relaxa-se esta hipótese buscando mensurar o viés da omissão do esforço e do talento individual. O primeiro conjunto de resultados aponta forte diferencial no esforço empregado entre os tipos para obter um bom desempenho no ENEM. O resultado mais surpreendente é que o efeito indireto da educação da mãe é 2,36 vezes maior do que os efeitos diretos.

*This study examines how social circumstances influence the educational performance of students taking the National Secondary Education Examination (ENEM). We will use two theoretical frameworks. Initially, we adopt the Roemer Identification Axiom, i.e., we assume independence between background and effort. After that, we relax this hypothesis seeking to measure the bias of omission of individual effort and talent. The first set of results shows strong difference in effort between types to obtain a good ENEM performance. The most striking result is that the indirect effect of mother's education is 2.36 times larger than the direct effects.*

\*Professor do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba (UFPB) e Bolsista de Produtividade em Pesquisa do CNPq.

†Professor Assistente da Universidade Federal do Semi-Árido e Doutorando do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba. E-mail: lauro.nogueira@ufersa.edu.br

‡Analista de Controle Externo – TCE-AC e Doutoranda em Economia pelo Programa de Pós-Graduação da Universidade Federal da Paraíba.



## 1. INTRODUÇÃO

Este estudo pretende investigar o quanto a origem individual influencia no desempenho dos alunos que prestaram o Exame Nacional do Ensino Médio (ENEM). Assume-se, *a priori*, que o resultado econômico, sintetizado pelo desempenho no teste, é fruto de dois fatores: um está relacionado à busca pela maximização do auto interesse, i.e., esforço e; o outro foge do controle individual – daqui por diante, denominado de conjunto de circunstâncias ou tipo – podendo ser sintetizado por variáveis como raça e região de nascimento, além do *background* familiar do estudante.

Serão consideradas duas abordagens. Na primeira, o esforço será tomado como uma variável latente e independente das circunstâncias. Neste caso, vigorará o *Roemer Identification Axiom* (RIA), estabelecendo que: duas pessoas com diferentes níveis de oportunidade, mas no mesmo percentil da sua distribuição condicionada ao seu fator de circunstâncias, empreendem o mesmo nível de esforço (Roemer, 1998). Essa etapa seguirá o referencial teórico de O'Neill et alii (2001), permitindo observar a diferença do desempenho dos alunos pertencentes a diferentes conjuntos de circunstâncias (tipos).

A segunda abordagem visa relaxar a hipótese simplificadora de Roemer (1998), fornecendo estimativas que consideram os diferentes níveis de esforço e talento individuais. Para tanto, o ponto de partida será o modelo de função de produção educacional de Hanushek (1970, 1979, 2007), que demonstra como os diversos insumos influenciam os resultados educacionais e, com isso, proporcionam uma orientação para a aplicação de políticas públicas alocativas. A magnitude da influência do esforço/talento individual será mensurada a partir do instrumental desenvolvido por Klein e Vella (2010), o qual utiliza os segundos momentos condicionados para identificar os efeitos diretos e indiretos dessas variáveis – sintetizados por um mecanismo de transmissão intergeracional representado pela educação dos pais – no desempenho educacional dos filhos. Tal procedimento é adotado devido a inexistência de variáveis instrumentais apropriadas no banco de dados.

Dito isso, o trabalho está organizado em cinco seções além da Introdução, que são: a Seção 2, onde apresenta-se uma breve revisão de literatura relacionada à desigualdade de oportunidades e desempenho educacional; a Seção 3, que é destinada aos modelos teóricos, métodos de estimação e descrição do banco de dados; a Seção 4, que avalia os principais resultados; e a Seção 5, que apresenta a conclusão do estudo.

## 2. IGUALDADE DE OPORTUNIDADES: UMA BREVE REVISÃO

Como já destacado, a abordagem predominante sobre a (des)igualdade de oportunidades, considera que as diferenças entre os resultados econômicos individuais são fruto de fatores de responsabilidade e não responsabilidade (Roemer, 1998). Em outras palavras, uma parte deles é determinada por variáveis de esforço como nível educacional, decisão de migrar, horas trabalhadas por ano, entre outras; e a outra, por fatores que fogem do controle dos agentes econômicos, variáveis de circunstâncias, ou seja, *background* familiar (nível educacional e ocupação dos pais), atributos individuais como raça, gênero, idade ou região de nascimento, entre outras.

Para tornar o conceito mais claro, considerem uma pequena ilustração: Antônio (A) e Bruno (B) possuem um conjunto de oportunidades idêntico, ou seja, pertencem a famílias de mesma classe social, frequentaram as mesmas escolas, são da mesma raça, sexo, etc. Após a conclusão do ensino médio, ambos decidem trabalhar na mesma profissão. Todavia, (A) trabalha duro e destina poucas horas ao lazer, enquanto que (B) destina maior parte do tempo ao lazer. De tais escolhas imerge uma desigualdade de renda, visto que: (A) obtém um maior nível de renda proveniente do trabalho do que (B). Nesta situação genuína de pura escolha dos agentes, a desigualdade existente é um problema social? Para os igualitários de resultado, a resposta é sim, pois qualquer desigualdade social é indesejável. Por outro lado, os igualitários de oportunidades, consideram a resposta não, dado que, a diferença nos resultados é nitidamente devida a uma escolha ótima de preferência individual.

Agora, assumam o caso de duas jovens: Carol (C), pertencente a uma família de classe alta; e Daniela (D), uma jovem com origem familiar humilde. Ambas desejam cursar medicina, entretanto, os gastos com a formação impedem que (D) siga seu objetivo. Dada a falta de recursos, (D) se torna enfermeira. Convém ressaltar o diferencial de rendimentos das duas profissões. Nesse contexto, dado que as duas jovens, ao trabalharem, exercem um nível de esforço semelhante, (C) torna-se rica rapidamente, enquanto que (D) apresenta um padrão de vida apenas considerável. Neste caso, assim como os igualitários de resultados, os igualitários de oportunidades consideram a desigualdade observada injusta ou indesejável.

Os casos referidos acima nos ilustram duas importantes questões. A primeira diz respeito às diferenças do conjunto de oportunidades enfrentadas pelos indivíduos, enquanto que a segunda corrobora sobre a influência das variáveis no resultado individual. As variáveis relacionadas às responsabilidades como, por exemplo, a quantidade de horas trabalhadas são denominadas de esforço e são estas determinantes para o diferencial nos resultados econômicos individuais. Contudo, existe uma gama de variáveis que fogem ao escopo da responsabilidade do indivíduo, como por exemplo, a origem, raça, sexo, *background* familiar, etc., variáveis denominadas de não responsabilidade ou circunstâncias.

No caso de (A) e (B), temos circunstâncias semelhantes, com esforço diferenciado. Por sua vez, o caso de (C) e (D), embora (D) se esforce da mesma forma que (C), as circunstâncias a impedem de obter melhores resultados. A literatura que trata igualdade de oportunidades segue dois princípios:

- i) Princípio de Compensação; e
- ii) Princípio de Recompensa. Em termos práticos, temos que (A) deve ser recompensado pelo esforço e (D) deve ser compensada devido à circunstância.

Inicialmente, as regras de mensuração e compensação baseadas na desigualdade de oportunidades, foram empregadas na análise da desigualdade de renda (ver Bourguignon et alii, 2007). A adaptação dessa abordagem ao estudo de desempenho educacional requereu um cuidado adicional: os estudos deveriam se concentrar na avaliação condicionada do desempenho em exames – *testes scores*. A razão para isso é que os estudos que contemplam a realização educacional, ou anos de estudo, não costumam ponderar a existência da heterogeneidade entre escolas, regiões e outros fatores. Em outras palavras, deve-se vislumbrar o fato de que a realização educacional não caracteriza a qualidade da educação (ver Ferreira e Gignoux, 2008).

Sob esse enfoque e usando dados do PISA 2000,<sup>1</sup> Ferreira e Gignoux (2008), mensuraram qual parcela da desigualdade observada no desempenho escolar é devida a fatores de circunstância. Em resumo, constatou-se que os maiores níveis de desigualdade de oportunidades educacionais são registrados em países em desenvolvimento – América Latina – embora, haja considerável heterogeneidade entre os países desenvolvidos. O *peso* das variáveis de circunstância, na desigualdade educacional total, varia entre 9% a 30% nos testes de matemática e entre 14% a 33% no teste de leitura.

Em outro estudo, Ferreira e Gignoux (2011), propõem duas formas de mensurar a desigualdade educacional. Uma voltada para o desempenho educacional – variância ou desvio padrão – e outra para a oportunidade educacional – parcela da variância que explica a influência das circunstâncias. Os resultados apontam que a desigualdade de oportunidades é responsável por 35% de todas as disparidades no desempenho educacional dos 57 participantes do PISA 2006.

De forma semelhante, Gamboa e Waltenberg (2012), analisam a desigualdade de oportunidade no desempenho educacional em seis países da América Latina.<sup>2</sup> Utilizando a metodologia desenvolvida por Checchi et alii (2010), os autores constatam que o grau de desigualdade de oportunidade educacional, varia entre 1% a 25%, o que denota uma considerável heterogeneidade entre os seis países. Nessa linha de atuação Aguirreche (2012), investiga como nível de desigualdade de oportunidades de um país

<sup>1</sup>Foram analisados 14 países, sendo 9 da OCDE e 5 da América Latina.

<sup>2</sup>Os países analisados foram Argentina, Brasil, Chile, Colômbia, México e Uruguai.



afeta o desempenho médio dos estudantes. Adotando a moderna teoria de justiça, a desigualdade foi decomposta em fatores de circunstâncias, esforço e sorte. Entretanto, na ausência de variáveis que representassem a sorte procurou-se explicar somente as diferenças propiciadas pelos determinantes do esforço – desigualdade justa – e das circunstâncias – desigualdade injusta. Utilizou-se um modelo estrutural baseado em Fleurbaey e Schokkaert (2009). Os resultados sinalizam elevado grau de desigualdade de oportunidades, isto é, mais de 30% de desigualdade injusta. Além disso, apura-se uma relação negativa de  $(-0,69)$  entre desigualdade de oportunidades e desempenho educacional.

Para o Brasil, em específico, Diaz (2012), mensura a trajetória temporal do nível de desigualdade de oportunidades existentes em relação ao conhecimento adquirido no Ensino Médio do Brasil. Além disso, analisa como a desagregação por tipo de escola – públicas e privadas – contribui nessa trajetória. Os resultados sugerem um maior nível de desigualdade em Matemática do que em Língua Portuguesa. Os mesmos se repetem quando se decompõem por tipo de escola. Entretanto, nas escolas privadas nota-se uma tendência de queda nos níveis de desigualdades. Por outro lado, nas escolas públicas, tais índices permanecem relativamente estáveis.

Em resumo, constata-se que o papel das circunstâncias é decisivo para o desempenho educacional dos indivíduos. Essa constatação é mais forte em países em desenvolvimento, em especial, em países da América do Sul, Europa Oriental e Ásia. Ao contrário, os menores índices de desigualdade de oportunidades são constatados, na sua grande maioria, em países da América do Norte, Europa Ocidental e Oceania. No caso específico do Brasil, vê-se que as circunstâncias observadas representam cerca de um quarto das disparidades constatadas (ver Bourguignon et alii, 2007). Onde a educação dos pais é de longe a circunstância mais importante que afeta os resultados. No que se refere a heterogeneidade regional, observa-se que respectivamente as regiões Norte e Nordeste oferecem as menores oportunidades de desempenho educacional.

### 3. PROCEDIMENTOS TEÓRICOS E METODOLÓGICOS

Para atingir os objetivos propostos, este estudo adotará duas estratégias teóricas e empíricas. A primeira baseada no modelo intergeracional bivariado para a desigualdade de oportunidades sugerido por O'Neill et alii (2001), onde os parâmetros de interesse são obtidos a partir de uma abordagem não-paramétrica e; a segunda estratégia, que visa o relaxamento da suposição relativa ao comportamento do esforço no resultado individual do procedimento anterior, adota um modelo para a função de produção educacional de Hanushek (1970, 1979, 2007), em conjunção com o instrumental desenvolvido por Klein e Vella (2010). Esta seção discutirá o passo a passo das duas abordagens, subseções 3.1 e 3.2, respectivamente. Por fim, na subseção 3.3 será apresentado o conjunto de dados.

#### 3.1. Modelo bivariado

De acordo com O'Neill et alii (2001), e assumindo que não há políticas compensatórias, admite-se que o conjunto de oportunidades de um indivíduo é representado por  $S_x$ , sendo o mesmo determinado por um vetor de características de não responsabilidade,  $x$ . Tal fato faz com que os resultados do indivíduo dependa do nível de esforço empregado, ou seja, das características de responsabilidade individual condicionado ao conjunto de oportunidade que o indivíduo possui.

Assim, os indivíduos podem optar por diferentes escolhas e diferentes resultados, estes podem ser resumidos por  $z = y[e, x]$ , em que  $z$  representa a utilidade ou renda (no caso da pesquisa, o desempenho no ENEM) ao longo do tempo; e  $e$  representa o esforço individual empregado. Assume-se, também, que a função de distribuição de  $e$  é contínua e postula-se duas suposições:

- SINC (*Strictly Increasing*):  $z = y[e, x]$  é estritamente crescente em  $e$ . Tal pressuposto é bastante plausível, dado que, quanto maior o nível de esforço maior a utilidade resultante. Definindo  $F_z^*(z|x)$

e  $F_e^*(e|x)$  a função de distribuição acumulada (fda) de  $z$  e  $e$ , ambos condicionadas a  $x$ , respectivamente. Logo, podemos expressar SINC por:

$$F_z^*(y[e,x]|x) = F_e^*(e|x) \quad (1)$$

A equação 1 nos indica que o nível de esforço de um indivíduo ( $e$ ) condicionado ao seu tipo – conjunto de oportunidades – será menor que  $\alpha$ -th percentil da distribuição do seu esforço se, e somente se, o resultado for inferior a  $\alpha$ -th percentil.

- IND (*Independente*):  $F_e^*(e|x)$  é independente de  $x$ . Esta suposição significa que não são assumidas diferenças na fda do esforço entre tipos diferentes, ou seja, entre indivíduos com diferentes características de não responsabilidade pertencentes ao mesmo percentil. Tal pressuposto é igualmente plausível, visto que os indivíduos não poderiam ser responsáveis pelo nível de esforço empregado, caso estes dependessem das circunstâncias aos quais estivessem inseridos. A equação 1 adicionada a suposição IND implica no Axioma de Identificação de Roemer (RIA).
- RIA (*Roemer's Identification Axion*):  $F_z^*(y[e,x]|x) = F_e^*(e|x) \Rightarrow e' = e''$ . A RIA implica que dois indivíduos com diferentes conjuntos de oportunidades, mas com o mesmo percentil da distribuição dentro do seu tipo, exercem o mesmo nível de esforço.

Com base nestas suposições, define-se  $\pi = F_z^*(z|x)$  como a fda do resultado de  $z$  condicionado as características de não responsabilidade,  $x$ . Analogamente, assume-se que esta função é estritamente crescente em  $z$ .  $F_z^{-1}(\pi|x)$  expressa o resultado obtido – desempenho dos alunos no ENEM – pelo indivíduo do tipo  $x$  e que estava no  $100^*\pi$ -th percentil da fda do resultado dentro do seu tipo. De acordo com a RIA,  $F_z^{-1}(\pi|x)$  equivale a observar  $y[p,x]$ . Entretanto,  $F_z^{-1}(\pi|x)$  fornecerá informações das características de responsabilidade e não responsabilidade dos indivíduos. De acordo com O'Neill et alii (2001), isto possibilita desenhar o resultado – desempenho no ENEM com uma função de  $\pi \in [0,1]$  para diferentes valores de  $x$ . Sendo o conjuntos de oportunidades para um particular tipo de  $x$  determinado por alguns resultados do tipo  $x$ , e podendo ser obtido ao variar suas características de responsabilidade  $e$  ou  $\pi$ . Desse modo, o conjunto de oportunidade do indivíduo do tipo  $x$  será:

$$S_x = (z,\pi) \in (\mathbb{R}^+ \times [0,1]) \quad z = F_z^{*-1}(\pi|x) \quad (2)$$

em que  $\mathbb{R}^+$  representa o conjunto dos números reais não negativos. Se  $F_z^{*-1}$  estiver disponível, pode-se descrever o conjunto de oportunidades para diferentes tipos de indivíduos e, também, qual a extensão das opções diferentes ou níveis de esforço que produzem diferentes resultados.

Usualmente, O'Neill et alii (2001) assumem que a sociedade determina de algum modo os elementos de  $x$ . Considera-se, no entanto, que  $x$  é uma variável multidimensional composta por elementos como raça, sexo, *background* familiar e habilidade inata. Os dados, assim exigidos, são complexos e difíceis de serem delimitados, em que pese a estrutura de mobilidade intergeracional que leva em conta o desempenho dos filhos ( $z$ ) em função das características do pais ( $x$ ).

Segundo O'Neill et alii (2001), as curvas de densidade acumuladas do resultado de  $z$  condicionada as características de não responsabilidade de  $x$ , serão inferidas a partir da modelagem kernel bivariada. Conforme Cameron e Trivedi (2005), o estimador de densidade de Kernel, introduzido por Rosenblatt (1956) é uma generalização do histograma e que faz uso de uma função de ponderação alternativa expressa por:

$$\hat{f}_{(x_0)} = \frac{1}{Nh} \sum_{i=1}^n K\left(\frac{x_i - x_0}{h}\right) \quad (3)$$



A ponderação da função  $K(\cdot)$  é denominada função de kernel. O parâmetro  $h$  é um parâmetro de alisamento chamado de largura da banda e a densidade é estimada através da avaliação de  $\hat{f}_{(x_0)}$  sob uma vasta gama de valores de  $x_0$  utilizado na formação de um histograma. Usualmente a avaliação se dá nos valores amostrais de  $x(x_1, \dots, x_N)$ , ou seja, a estimativa de densidade de Kernel torna-se mais suave do que a de um histograma.

Assim, o vetor de características de não responsabilidade  $x$  – variáveis de circunstância – e o resultado obtido pelo indivíduo ( $z$ ), a distribuição de  $z$  pode ser expressa como:

$$f[z|x] = \frac{f[z|x]}{f_x[x]} \quad (4)$$

em que  $f_x[x]$  representa a distribuição marginal das circunstâncias e  $f[z|x]$  a distribuição conjunta de  $z$  e  $x$ . Para estimar (4), substitui-se o numerador e o denominador por estimativas. A distribuição marginal das circunstâncias é estimada utilizando a técnica adaptada de densidade de kernel para distribuições univariadas:

$$\hat{f}_{xA}[x] = \frac{1}{nh_x} \sum_{i=1}^n \frac{1}{w_i} k \left[ \frac{x - x_i}{w_i h_x} \right] \quad (5)$$

A distribuição conjunta das circunstâncias e resultados obtidos (numerado da equação (4)) é obtida por:

$$\hat{f}_A[z|x] = \frac{1}{nh_z h_x} \sum_{i=1}^n \frac{1}{w_i^2} k \left[ \frac{z - z_i}{w_i h_z} \right] \left[ \frac{x - x_i}{w_i h_x} \right] \quad (6)$$

O estimador de kernel adaptativo ajusta a largura da janela tornando-a mais estreita quando a densidade é maior, e amplia quando a densidade é menor, preservando os detalhes em que os dados são abundantes e reduzindo o ruído quando os dados são esparsos. A janela local dos fatores usados são determinados por:

$$w_i = \left[ \frac{\hat{f}_g}{\hat{f}_k[z_i, x_i]} \right]^{1/2} \quad (7)$$

em que  $\hat{f}_k[z_i, x_i]$  é o estimador kernel de janela fixa de  $f[z, x]$ ; e  $\hat{f}_g$  é a média geométrica de  $\hat{f}_k[z, x]$ . Conforme destaca O'Neill et alii (2001), a operacionalização deste procedimento envolve uma estratégia de estimação em dois passos. Primeiro,  $\hat{f}_k[z, x]$  é estimado utilizando uma janela de largura fixa, obtida com uma largura inicial através de *Scott's optimal bandwidth*.

O segundo passo, destina-se a obter a densidade utilizando pesos envolvidos na construção da densidade final das equações (5) e (6). Assume-se que a kernel é multiplicativa, conforme Tiede (1988), tornando-a em uma simples expressão de *fda*. Dessa forma, substitui os termos da equação (4) pelas estimativas das equações (5) e (7), obtendo-se a distribuição condicional:

$$\hat{F}_{zA}^*[Z|x] = \frac{\sum_{i=1}^n \frac{1}{w_i} K \left[ \frac{x - x_i}{w_i h_x} \right] G \left[ \frac{z - z_i}{w_i h_z} \right]}{\sum_{i=1}^n \frac{1}{w_i} K \left[ \frac{x - x_i}{w_i h_x} \right]} \quad (8)$$

em que  $G(z) = \int_{-\infty}^z K(t) dt$  representa a fda da função de kernel. O conjunto de oportunidades para o indivíduo dado o seu tipo  $x_i$ , pode assim ser estimado por:

$$\hat{S}_{x_i} = \left\{ (z, \pi) \in (\mathbb{R}^+ \times [0, 1]) \mid z = \hat{F}_{zA}^{*-1}[\pi|x] \right\} \quad (9)$$

### 3.2. Função de produção educacional

A Função de Produção de Educação, denominada de FPE, analisa a forma como os diversos insumos do processo educacional podem afetar os resultados educacionais dos indivíduos, servindo de respaldo para a escolha dos principais determinantes educacionais e, com isso, proporcionando um guia para a aplicação de políticas públicas. De acordo com Hanushek (1970, 1979, 2007), a FPE é dada por:

$$A_{it} = g(F_i^t, P_i^t, I_i, S_i^t) \quad (10)$$

em que  $A_{it}$  é o vetor educacional realizado pelo estudante  $i$ -ésimo no tempo  $t$ ;  $F_i^t$  é o vetor de características individuais e familiares do  $i$ -ésimo estudante acumulado no período  $t$ ;  $P_i^t$  é o vetor do corpo discente (influências de pares), ou seja, variáveis socioeconômicas e *background* familiar de outros estudantes na escola acumulados no período  $t$ ;  $I_i$  é o vetor de dotações iniciais do  $i$ -ésimo indivíduo; e  $S_i^t$  é o vetor de insumos escolares relevantes para o  $i$ -ésimo estudante acumulados no período  $t$ .

A partir desse modelo, Albernaz et alii (2002) estimaram a função de produção educacional para o ensino fundamental através de um modelo linear hierárquico, utilizando como base de dados o SAEB para as 8ª Série de 1999. A conclusão dos autores é de que cerca de 80% da variância do desempenho médio entre as escolas deve-se à diferenças na composição socioeconômica de seus alunos.

Cabe aqui ressaltar um aspecto relevante para a estimação de (10). Refere-se omissão de variáveis como, por exemplo, a habilidade do indivíduo. Infelizmente, esta é uma variável de difícil mensuração e a sua negligência na FPE torna os regressores do *Ordinary least square* (OLS) endógenos. Além disso, temos problemas de simultaneidade entre renda dos pais e habilidade dos filhos. Pois, quanto mais educado o indivíduo, em média, maior o nível de habilidade dos filhos. De modo semelhante que, quanto maior a habilidade, maior o nível de renda. Assim, a fim de superar tais problemas, adotou-se as estratégias de estimação e identificação utilizadas em Klein e Vella (2009, 2010) tratadas a seguir.

Considere os seguintes modelos simultâneos de transferência no desempenho educacional:

$$E_i^f = \alpha_0 + \beta_0 X_i + \beta_m E_i^m + \beta_p E_i^p + u_i, \quad i = 1, \dots, N \quad (11)$$

$$E_i^j = \alpha_0 + \gamma_j X_i + v_i^j, \quad j = m, p \quad (12)$$

onde  $E_i^f$  denota o desempenho educacional do filho;  $E_i^j$  denota a educação dos pais, isto é, da mãe quando considera-se  $j = m$  ou, caso contrário,  $j = p$  a educação do pai;  $X_i$  representa o vetor de variáveis exógenas, o qual assume-se, em geral, serem idênticos em ambas as equações dos desempenhos educacionais dos filhos e dos pais. Os  $\beta'_s$  e os  $\gamma_j$  são parâmetros desconhecidos;  $u_i$  e  $v_j$  são os termos de erros com covariância não nula no quais refletem a endogeneidade dos termos  $\beta_m E_i^m$  e  $\beta_p E_i^p$ , de modo que, esta covariância diferente de zero torna as estimações dos  $\beta'_s$  por OLS inconsistentes.

Considere a função controle como uma versão de estimação por variáveis instrumentais para este modelo. Entretanto, tal procedimento requer purificar o termo de erro da equação (11) em razão do mesmo está correlacionado com o termo de erro da equação (12) em ambas as estimações de  $E_i^j$ , ou seja, quando estimamos respectivamente para  $j = m, p$ . De forma que, o erro da equação principal pode ser reescrito da seguinte forma:

$$u_i = \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i \quad (13)$$

Sendo  $\lambda^j = \frac{cov(v^j u)}{var(v^j)}$  quando não existir dependência entre a distribuição dos erros e os  $X$ 's. Para tal, necessita-se estimar (12) os resíduos da forma reduzida duas vezes, isto é, tanto para a mãe quanto para o pai. Assim, estima-se:

$$E_i^f = \alpha_0 + \beta_0 X_i + \beta_m E_i^m + \beta_p E_i^p + \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i, \quad i = 1, \dots, N \quad (14)$$



onde  $\varepsilon_i$  denota o termo de erro de média zero o qual não é correlacionado com as variáveis explicativas inclusas. Contudo, a estimação de (14) não é possível com a ausência de restrições na forma reduzida apresentada em (12) que garantam que a matriz  $M^1 = [X, E, E^m, E^p, \hat{v}^m, \hat{v}^p]$  tenha rank máximo. Ao assumir que os erros da distribuição dependem dos  $X$ 's, ou seja, que em geral são heterocedásticos, o coeficiente em (13) torna-se:

$$\lambda^j = \frac{cov(v^j u | X)}{var(v^j | X)} = A^j(X) \tag{15}$$

De modo que, o impacto de  $v_i^j$  em  $u_i$  depende do valor de  $X_i$ . Sob o pressuposto de correlação condicional, o metodo KV mostra:

$$\lambda^j = \frac{cov(v^j u | X)}{var(v^j | X)} = \rho^j \frac{\sqrt{var(u_i | X_i)}}{\sqrt{var(v_i^j | X_i)}} \tag{16}$$

Ao qual, dada as suposições mostradas abaixo resulta em:

$$u_i = \rho^m \frac{H_{ui}}{\hat{H}_{vi}^m} \hat{v}_i^m + \rho^p \frac{H_{ui}}{\hat{H}_{vi}^p} \hat{v}_i^p + \varepsilon_i \tag{17}$$

No entanto, Farré et alii (2013), impõem uma estrutura paramétrica através das seguintes suposições:

$$H_{ui}^2 = \exp(\alpha_1 (Z_{ui} \theta_1)) \tag{18}$$

$$H_{vi}^{j2} = \exp(\alpha_2^j (Z_{vi} \theta_{2j})), \quad j = m, p \tag{19}$$

A estimação é agora possível com a matriz  $M^1 = [X, E, E^m, E^p, \hat{v}^m, \hat{v}^p, \frac{H_{ui}}{\hat{H}_{vi}^m} \hat{v}_i^m, \frac{H_{ui}}{\hat{H}_{vi}^p} \hat{v}_i^p]$  que apresenta rank máximo devido a não linearidade induzida pelo papel multiplicativo dos  $X$ 's. Todavia, KV mostra que os parâmetros do modelo são identificados mesmo sem suposições paramétricas. Os autores mostram que é possível estimar o seguinte modelo de função controle consistentemente.

$$E_i^f = \alpha_0 + \beta_0 X_i + \beta_m E_i^m + \beta_p E_i^p + \rho^m \frac{H_{ui}}{\hat{H}_{vi}^m} \hat{v}_i^m + \rho^p \frac{H_{ui}}{\hat{H}_{vi}^p} \hat{v}_i^p + \varepsilon_i, \quad i = 1, \dots, N \tag{20}$$

onde  $\hat{v}_i^m$  e  $\hat{v}_i^p$  representam respectivamente os resíduos das equações da educação da mãe e do pai;  $H_{ui}$  representa o desconhecido  $H_u(X_i)$ , ao passo que  $\hat{H}_{vi}^j$  são as estimações de  $H_v^j(X_i)$ . No que se refere à heterocedasticidade, assume-se que:

$$E[u_i | X_i] = E[v_i^j | X_i] = 0 \tag{21}$$

Tal formulação constitui uma importante restrição paramétrica do modelo. Contudo, alguns autores, entre eles, Card (2001); Farré et alii (2013), argumentam que essa forma funcional possui, pelo menos, duas justificativas práticas:

- i) é provável que fatores regionais como “a distância para a escola” seja determinante para a realização educacional, onde uma distribuição geográfica desigual do número e qualidade das instituições de ensino pode produzir importantes diferenciais tanto na média como na variância da escolaridade em todas as regiões. Ademais, a heterocedasticidade também pode surgir de vários determinantes da educação, devido ao efeito heterogêneo individual e, portanto, mesmo após a inclusão de uma função indicadora que capta o efeito individual, é provável que o efeito varie no tempo e entre indivíduos de mesmas características;



- ii) em segundo lugar, a restrição apresenta uma correlação constante. Em termos econômicos, isso significa que as transferências da habilidade não observadas, mensuradas pela correlação dos coeficientes entre  $u_i^*$  e  $u_i^{j'}$ s, são independentes do ambiente socioeconômico individual.

### 3.3. Descrição dos dados

As estimativas da pesquisa serão possibilitadas pelos microdados do Exame Nacional do Ensino Médio (ENEM, 2010), captados no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O ENEM foi criado em 1998 com a finalidade de avaliar o desempenho dos estudantes brasileiros ao final do Ensino Médio. A partir de 2009, foram implementadas algumas mudanças para contribuir com a democratização das oportunidades de acesso à universidade, visto que a partir desse ano, o exame passou a ser utilizado como mecanismo de seleção para o ingresso no ensino superior.

Os microdados são divididos em:

- i) variáveis relacionadas aos inscritos no exame; e
- ii) variáveis do questionário socioeconômico (características dos indivíduos e de sua família).

A junção dessas bases possibilita compilar informações a respeito das variáveis de circunstâncias e resultados individuais. Adicionalmente, serão utilizados os dados do SAEB, a fim de obter um indicativo de qualidade da educação. A escolha pelo SAEB se deve ao fato de incorporar-se uma *proxy* – qualidade das escolas – nas estimações. Em outras palavras, possibilita mensurar qual a influência de frequentar-se uma escola de boa qualidade, pois, embora, as repartições de ensino privado ofereçam em média educação de maior qualidade, isso por si só não garante tal fato.

Quadro 1 – Descrição das variáveis

Variável	Descrição
Resultados individuais	Somatório dos pontos obtidos nas provas que contemplam quatro áreas de conhecimento: Linguagens, códigos e suas tecnologias; Matemática e suas tecnologias; Ciências da Natureza e suas tecnologias; e Ciências Humanas e suas tecnologias.
Renda familiar	Somatório da renda familiar que são distribuídas em cinco classes de renda: i) Renda familiar mínima - Até um salário mínimo; ii) Renda familiar baixa – 1 a 3 salários mínimos; iii) Renda familiar média – 3 a 6 salários mínimos; iv) Renda familiar Alta – 6 a 9 salários mínimos; v) Renda familiar Máxima - acima de 9 salários mínimos.
Escolaridades dos pais	Agregadas em três estratos: i) Pais com ensino fundamental incompleto e analfabetos – menos de 4 anos de estudos; ii) Pais com ensino fundamental e/ou médio – de cinco a onze anos de estudos; iii) Pais com ensino superior ou pós-graduação – mais de onze anos de estudos.
Tipo de Escola	Se o aluno cursou predominantemente em escola pública ou privada. Cabe ressaltar aqui, que optou-se por excluir as escolas indígenas e escolas quilombolas pela falta de dados.
Sexo	Feminino ou masculino.
Raça	Divididos em brancos e não brancos (aqui abrangendo pretos, pardos, amarelos e indígenas).
Moradia	Zona rural ou zona urbana.
Qualidade escolar	Resultado do SAEB por unidade escolar.

Fonte: Elaboração dos autores, a partir de dados do ENEM.



Com vistas a investigar se os fatores de não responsabilidade do indivíduo afetam o seu desempenho no ENEM, a estratégia utilizada se divide em dois blocos de resultados. O primeiro, visa estimar a densidade de Kernel, com as seguintes variáveis selecionadas:

- i) resultados individuais (somatório das notas);
- ii) renda familiar;
- iii) escolaridade dos pais; e
- iv) tipo de escola – pública ou privada.

Após as tabulações, excluindo-se os *missing*, a amostra totaliza 2.937.253 que corresponde a 63,49% do total de estudantes de todas as regiões do Brasil.

Por sua vez, o segundo bloco, sob uma perspectiva paramétrica, realizar-se-á regressões tendo como variável dependente os resultados individuais, e como variáveis explicativas:

- i) escolaridade dos pais;
- ii) renda familiar;
- iii) tipo de escola;
- iv) sexo;
- v) raça;
- vi) localização da moradia; e
- vii) qualidade da escola.

Cabe aqui destacar, que esta última variável denota o resultado do SAEB por escola como uma *proxy* para a qualidade escolar, sendo que após a junção dos dados com o ENEM, a amostra restringe-se a 502.116 observações. No entanto, a queda no tamanho da amostra dar-se em função de muitas escolas compostas presentes na amostra principal não comporem a amostra do SAEB.

Em adição, a fim de capturar informações a respeito da amostra do ENEM, a Tabela 1 sumariza as principais informações das variáveis utilizadas, bem como as estatísticas descritivas. Ao observar a Tabela, verifica-se que, aproximadamente 60% dos que prestaram o exame são mulheres; mais de 80% vêm de escola pública; aproximadamente 71% dos candidatos possuem renda familiar de no máximo três salários mínimos, em que 33% destes têm renda familiar de até 1 salário mínimo. Além do mais, constata-se que 31% dos candidatos são filhos de mães que estudaram no máximo quatro anos, ou seja, possuem, no máximo, o primeiro ciclo do Ensino Fundamental.

Em termos de nota média, constata-se que a Região Nordeste está aquém das Regiões Sul, Sudeste e Centro-Oeste e um pouco acima da Região Norte. Além disso, averigua-se que 38% dos candidatos moram na região sudeste e outros 30% moram no nordeste do país, seguidos de 13%, 8,83% e 8,43% das Região Sul, Norte e Centro-Oeste, respectivamente.

De um modo geral, observa-se que boa parte da amostra é composta por indivíduos de origem pobre e de baixo *background* familiar, além de que a maioria dos estudantes são oriundos de escola pública, ainda vista com os piores resultados em termos de qualidade educacional quando comparada a escola privada.

Tabela 1: Análise descritiva

Estudantes por variáveis de circunstâncias				
Sexo		Raça		
Masculino	Feminino	Branco	Não Branco	
1.187.420	1.749.833	1.326.571	1.610.682	
-40,42%	-59,58%	-45,16%	-54,84%	
Alunos conforme renda familiar				
Mínima	Baixa	Média	Alta	Máxima
709.298	1.439.917	492.999	129.143	165.896
-24,14%	-49,02%	-16,78%	-4,40%	-5,64%
Tipo de escola		Escolaridade da mãe		
Pública	Privada	Analfabeta	Nível médio	Nível superior
2.374.253	563.000	911.794	1.538.368	487.091
-80,83%	-19,17%	-31,04%	-52,38%	-16,58%
Regiões dos alunos e nota média				
Centro Oeste	Nordeste	Norte	Sul	Sudeste
247.592	895.005	260.453	402.020	1.132.183
2590*	2560*	2532*	2678*	2720*
-8,43%	-30%	-8,83%	-13%	-38%

Fonte: Elaboração dos autores, a partir de dados do ENEM.

Os valores em % representam a participação percentual do total da amostra.

Os valores (\*) representam o somatório das notas das provas de ciências naturais, ciências humanas, língua portuguesa, matemática e redação.

## 4. RESULTADOS

Nesta seção, serão apresentados os resultados da pesquisa. Primeiramente, na subseção 4.1, controem-se as densidades de Kernel sob perspectiva nacional e regional através de interações de circunstâncias. Nessa etapa admite-se a RIA. Por fim, subseção 4.2, infere-se sobre as relações das desigualdades de oportunidades no Brasil, considerando-se o esforço individual como uma variável observável, porém, omitida na inferência.

### 4.1. Primeiro bloco

Neste primeiro momento, estima-se as densidades conjuntas de Kernel, descrita pela equação 9. De acordo com O'Neill et alii (2001), assumindo a RIA, o conjunto de oportunidades de um indivíduo pode ser determinado por um vetor de características de não responsabilidade. Assim, o desempenho no ENEM depende do nível de esforço empregado condicionado ao conjunto de oportunidades que o indivíduo possui. Nesse contexto, as densidades de Kernel permite inferir sobre os diferenciais de esforço dos estudantes que prestaram o exame. No entanto, isso somente é possível se todas as circunstâncias relevantes para determinação do sucesso individual estiverem sendo consideradas na estimação. Consequentemente, tal fato denota uma limitação empírica.



Os resultados encontrados, expressam, a nível nacional, um elevado nível de desigualdade de oportunidades, apreendido no esforço – em termos de probabilidade – necessários para que o indivíduo com baixo *background* social atinja a média nacional. Assumi-se que baixo *background* social são aqueles indivíduos que detêm as seguintes características de não responsabilidade: filhos de mães analfabetas, com renda familiar menor do que 1 (um) salário mínimo e que estudaram predominantemente em instituição pública. O inverso se aplica aos indivíduos com alto *background* social, ou seja, estudantes filhos de mães com nível superior, renda familiar acima de 9 salários mínimos e que estudaram predominantemente em escola privada. A Figura 1 ilustra os resultados para o Brasil e suas regiões, conforme descrito acima. Todavia, tais resultados são válidos para os estudantes que prestaram o exame nacional e não entre todos os alunos do Ensino Médio.

A curva à direita representa os alunos que estão associadas às circunstâncias favoráveis e a curva à esquerda, por sua vez, os alunos associados as situações desfavoráveis. Ao observar a Figura 1, verifica-se que embora haja disparidades regionais, o padrão da desigualdade de oportunidades educacionais configura-se de forma semelhante em que os indivíduos com baixo *background* social encontram-se aquém dos demais.

De acordo com os resultados apresentados na Tabela 2, em termos nacionais, a probabilidade de um indivíduo com baixo *background* social obter a nota acima da média do ENEM (média de 2.638) é quatro vezes menor<sup>3</sup> do que um indivíduo com alto *background* social. Além do mais, a probabilidade de indivíduos com baixo *background* estar entre os 5% que obtêm as melhores notas é de 0,27%, enquanto que os indivíduos com alto *background* é de 35%. De um modo geral, o conjunto de oportunidades do primeiro grupo (baixo) é cerca de 130 vezes menor do que o segundo grupo (alto), o que, *a priori*, indica existir um abismo social expresso na desigualdade educacional do país.

Por sua vez, em termos regionais, os resultados sugerem que as menores oportunidades de desempenho educacional estão na Região Norte e Nordeste. Por exemplo, na Região Nordeste, para que o aluno de baixo *background* social atinja uma nota acima da média deverá se esforçar (assumindo a RIA e considerando que todas as circunstâncias influentes no resultado estão sendo consideradas nas estimativas) 4,6 mais do que o aluno de alto *background*, ou seja, representando um diferencial de 15% em relação ao cenário nacional. Entretanto, quando compara-se a Região Sudeste, tem-se que o conjunto de oportunidades educacionais dessa é cerca de 24% maior do que o cenário nacional e 33% maior do que na Região Nordeste.

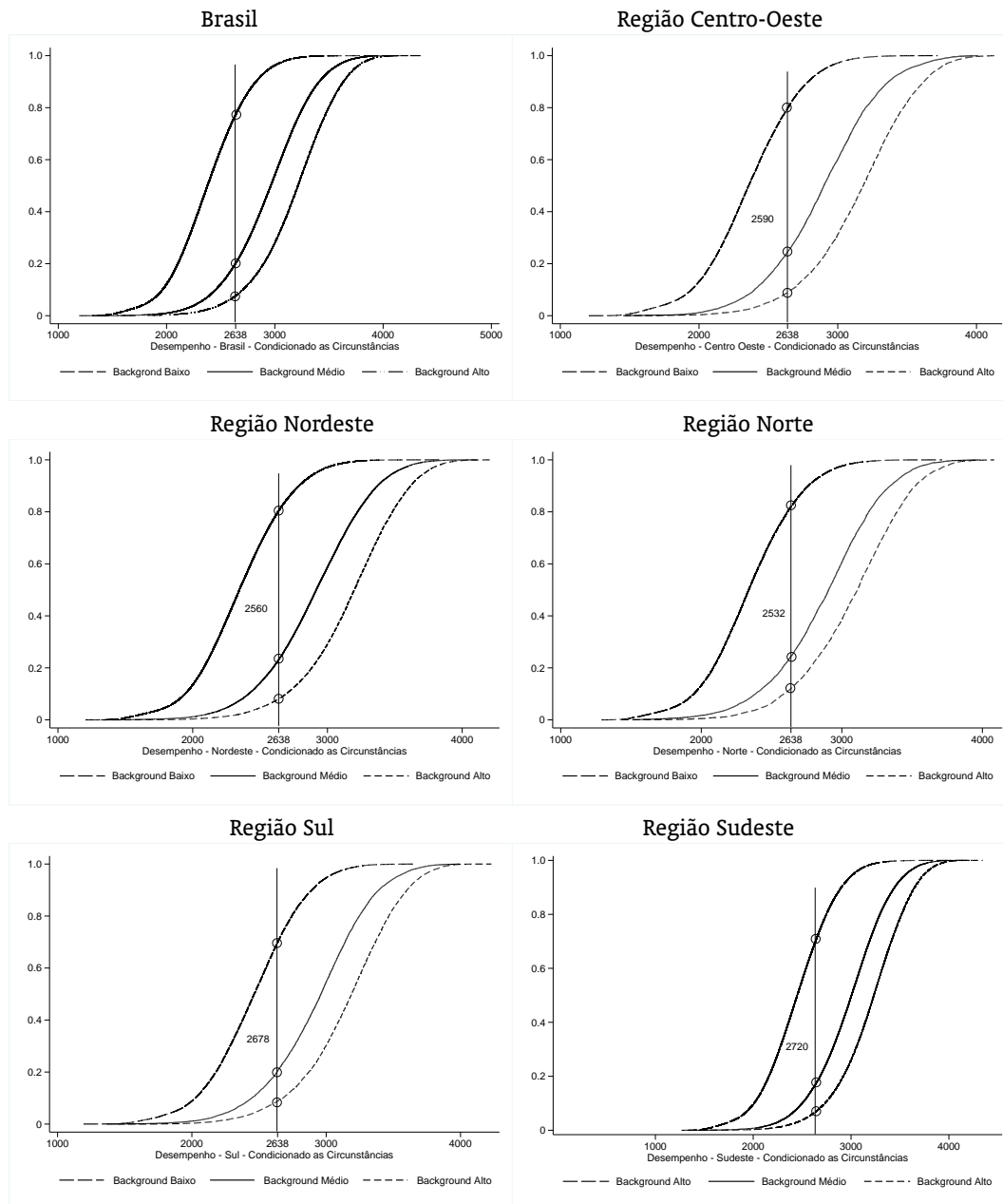
Em outras palavras, a probabilidade – a chance – de estudantes nordestinos, filhos de mãe analfabeta, oriundos de escolas públicas e com renda familiar menor do que 1 (um) salário mínimo de atingir uma nota acima da média nacional é de apenas 20%. Enquanto que entre um estudante com situação social oposta é de aproximadamente 92%. Por outro lado, quando analisamos as probabilidades de um aluno das Regiões Norte e Nordeste estarem entre os 5% que obtiveram os melhores desempenhos no ENEM, são semelhantes ao cenário nacional. Respectivamente, a chance de um indivíduo pertencente ao percentil social inferior estar nesse seletor grupo é de apenas 0,15% e 0,18%.

Além disso, quando considera-se os estudantes destas regiões do percentil social superior, verifica-se que a possibilidade dos mesmos estarem entre os 5% melhores é de aproximadamente 25% e 34%, respectivamente. Em síntese, o conjunto de oportunidades de um nordestino com baixo *background* social é aproximadamente 167/189 vezes menor do que indivíduos com alto *background*.

E se compararmos os resultados da Região Nordeste com os da Região Sudeste? Bem, quando se analisa os estudantes sudestinos, constata-se que a possibilidade do estudante de baixo *background* social estar entre os 10% que obtêm melhores resultados é de aproximadamente 1,74% contra 0,74%,

<sup>3</sup>O cálculo dessa probabilidade é bastante simples. Considerando a análise acima (Brasil), a probabilidade de um aluno com *background* social baixo tirar uma nota acima da média – percentil 50% – é igual a um menos a probabilidade de se estar exatamente naquele ponto. Ou seja:  $P_B = 1 - P_{50\%} = 1 - 0,7694 = 0,2306$ . Enquanto que para um aluno com *background* social alto é:  $P_A = 1 - P_{50\%} = 1 - 0,0755 = 0,9245$ . Em síntese  $P_A = 0,9245$  corresponde a quatro vezes o valor de  $P_B = 0,2306$ .

Figura 1: Conjunto de Oportunidades – Brasil e Regiões



Fonte: Elaboração dos autores, a partir de dados do ENEM.



Tabela 2: Probabilidade condicional acumulada – Brasil e regiões

Background social	Brasil					Centro Oeste				
	Percentil					Percentil				
	25%	50%	75%	90%	95%	25%	50%	75%	90%	95%
Baixo	0.4755	0.7694	0.9401	0.9899	0.9973	0.4977	0.7915	0.9529	0.9924	0.9981
Médio	0.0652	0.1983	0.4599	0.7323	0.8605	0.0814	0.2425	0.5184	0.7838	0.8854
Alto	0.0218	0.0755	0.2173	0.4637	0.6458	0.0252	0.0862	0.2416	0.4934	0.6736
Background social	Norte					Nordeste				
	Percentil					Percentil				
	25%	50%	75%	90%	95%	25%	50%	75%	90%	95%
Baixo	0.5230	0.8157	0.9593	0.9934	0.9985	0.5118	0.8000	0.9517	0.9926	0.9982
Médio	0.0855	0.2390	0.5139	0.7897	0.8987	0.0765	0.2269	0.5027	0.7654	0.8828
Alto	0.0338	0.1166	0.3109	0.5795	0.7488	0.0233	0.0783	0.2255	0.4794	0.6575
Background social	Sul					Sudeste				
	Percentil					Percentil				
	25%	50%	75%	90%	95%	25%	50%	75%	90%	95%
Baixo	0.3753	0.6918	0.9137	0.9847	0.9968	0.3950	0.6969	0.9098	0.9826	0.9945
Médio	0.0684	0.1973	0.4511	0.7400	0.8698	0.0516	0.1684	0.4098	0.6939	0.8356
Alto	0.0258	0.0855	0.2314	0.4871	0.6702	0.0192	0.0677	0.2008	0.4401	0.6238

Fonte: Elaboração dos Autores, a partir de dados do ENEM.

se o mesmo for aluno nordestino. Em termos de conjunto de oportunidades entre grupos, averiguasse que o estudante sudestino pertencente ao estrato social inferior tem aproximadamente 32,17 menor possibilidade de estar entre os 10% melhores, contra 70,35 menor, caso o estudante seja nordestino. O fato de o estudante morar nos nordeste reduz em mais de 100% as chances do mesmo atingir os 10% melhores desempenhos nacionais, comparados aos alunos da Região Sudeste.

Outra questão importante a ser destacada, diz respeito às variáveis utilizadas nesta pesquisa. A Tabela 3 traz os principais determinantes da desigualdade de oportunidades. Conforme exposto, a renda familiar apresentou-se como o fator mais incisivo na desigualdade de oportunidade educacional, seguido da escolaridade da mãe, escolaridade do pai, tipo de escola frequentada pelo aluno, raça, localidade onde reside e sexo.

Observa-se que a influência da mãe é ligeiramente superior a do pai como determinante. No entanto, a possibilidade de um estudante filho de mãe com nível superior ou pós-graduação – mais de onze anos de estudo – obter um resultado acima da média é mais do que o dobro comparado com o estudante filho de mãe analfabeta. Dentre as variáveis, a renda é sem dúvida crucial na determinação do resultado, já que alunos que pertencem a faixa de renda mínima têm que se esforçar praticamente três vezes mais do que os indivíduos pertencentes a faixa máxima de renda para obterem resultados acima da média.

Em linhas gerais, os resultados mostram-se ainda mais expressivos quando avaliamos percentis superiores. Por exemplo, a possibilidade de um aluno pertencer ao nível mínimo de renda estar entre os 5% melhores no exame é de 45 vezes menos que os alunos que fazem parte da faixa superior da renda. No entanto, deve-se ressaltar que a renda, em geral, carrega em si outras características.

## 4.2. Segundo bloco

Com a finalidade de iniciar a busca por respostas para questões como: quais fatores estão associados a essa desigualdade de oportunidades? Ou, o que poderia ser feito para atenuar tais desigualdades? Esta seção estimará a função de produção da educação expressa pela Equação (10). Todavia, destaca-se

Tabela 3: Probabilidade condicional acumulada – Brasil

Variáveis	Tipo				% Probabilidades			
	Analfabeto	Fundamental	Médio	Superior	Analfabeto	Fundamental	Médio	Superior
Pai	0,63	0,54	0,42	0,22	0,37	0,46	0,58	0,78
Mãe	0,64	0,56	0,44	0,26	0,36	0,44	0,56	0,74
Renda	Mínima	Média	Máxima		Mínima	Média	Máxima	
	0,71	0,33	0,12		0,29	0,67	0,88	
Escola	Pública		Privada		Pública		Privada	
	0,57		0,20		0,43		1,00	
Raça	N Brancos		Brancos		N Brancos		Brancos	
	0,57		0,42		0,43		0,58	
Sexo	Feminino		Masculino		Feminino		Masculino	
	0,54		0,45		0,46		0,55	
Moradia	Rural		Urbana		Rural		Urbana	
	0,64		0,48		0,36		0,52	

Fonte: Elaboração dos Autores, a partir de dados do ENEM.

Nota: A renda corresponde: Renda familiar mínima – até um salário mínimo; Renda familiar média – três a seis salários mínimos; e Renda familiar Máxima – acima de nove salários mínimos.

que a fim de tornar os resultados paramétricos robustos será adotada a metodologia desenvolvida por Klein e Vella (2009, 2010) e Farré et alii (2013) – daqui por diante, KV.

Os resultados da estimação OLS e KV são apresentados na Tabela 4. Eles assumem que resultado educacional do estudante (logaritmo da nota no ENEM), é uma função dos seguintes fatores:

- a) escolaridade da mãe;
- b) escolaridade do pai;
- c) tipo de escola, se particular ou pública;
- d) localização da escola, se localizada em zonas rurais ou urbanas;
- e) raça;
- f) gênero, sexo do candidato;
- g) qualidade da escola sintetizada pelos dados do SAEB por unidade escolar e;
- h) a renda familiar.

Os resultados obtidos por OLS coincidem com a estimação não paramétrica, bem como o esperado na literatura. Ou seja, quanto maior o conjunto de oportunidades maior o nível educacional do indivíduo. Ao observar a Tabela 4, notem que não há evidências de que as variáveis têm influência indireta da educação dos pais sobre os resultados individuais quando a endogeneidade não é levada em conta. No entanto, prevalece a relação positiva entre os resultados individuais e suas características, tais como: o aluno ter estudado em escola particular e de boa qualidade, morar em zona urbana, ser do sexo feminino, ser branco, e possuir renda alta, além de que ser filhos de pais com elevado grau de instrução.

Para corrigir possíveis vieses decorrentes da situação descrita anteriormente, o estudo adota a metodologia de KV. De acordo com Klein e Vella (2010), o uso dessa abordagem evita o uso de variáveis instrumentais em cenários em que os instrumentos não estão facilmente disponíveis,<sup>4</sup> como é o caso

<sup>4</sup>Ver, por exemplo, Angrist e Keueger (1991), Farré et alii (2013).



aqui em questão. A diferença entre o coeficiente KV e OLS obtido, representa a medida do viés da variável educação obtido por OLS. Cabe destacar que, uma das principais vantagens da adoção do método KV é a identificação do efeito direto da educação dos pais filtradas pelas funções controle ( $\rho$  mãe e  $\rho$  pai) que refletem o efeito indireto das variáveis não observadas.

Contudo, a omissão das variáveis de esforço e talento individuais, aliada a sua relação com variáveis incluídas na estimação, podem gerar vieses na estimação OLS. Poder-se-ia inferir, por exemplo, que parte do talento é transmitida geneticamente. Neste sentido, pais mais talentosos gerariam filhos mais hábeis. A maior habilidade dos pais é, em média, refletida por um maior nível educacional e a dos filhos por um melhor desempenho no ENEM. Logo, parte da relação entre o Log da nota e a educação dos pais, pode ser fruto do talento individual. Nesse sentido, os parâmetros associados a estas variáveis podem estar superestimados.

Tabela 4: Estimação KV e OLS

MODELO	OLS	KV	KV (Bootstrap)
Lognota	$\beta_{ols}$	$\beta_{kv}$	$\beta_{ols}$
Anos de Estudo da Mãe	0.01250*	0.00968*	0.00967*
Anos de Estudo do Pai	0.01240*	0.01170*	0.01171*
Tipo de Escola	0.03000*	0.03190*	0.03187*
Localização	0.00182*	0.00325*	0.00325*
Raça	0.02160*	0.02190*	0.02191*
Sexo	-0.00479*	-0.00481*	-0.00481*
Qualidade	0.00137*	0.00138*	0.00138*
Renda Familiar	0.02270*	0.02390*	0.02387*
$\rho$ mãe	-	0.02283*	0.02283*
$\rho$ pai	-	0.01122	0.01122
Constante	7.372*	7.37409*	7.374*
TOTAL: 502.116 Observações			

Fonte: Elaboração própria a partir de dados do ENEM/SAEB.

\*1%; \*\*5% significância.

Sobre tais pressupostos, o termo de erro da equação (12) não somente captura os fatores não observáveis como, por exemplo, a habilidade. Entretanto,  $\rho^j$  positivo, sugere que o retorno da habilidade individual não observada é positivo, embora, isso pode incluir outras características como motivação, sorte. Esses resultados indicam que a educação dos pais é um fator não exógeno a educação dos filhos e que fatores não observáveis que influenciam o desempenho educacional são positivamente correlacionados entre as gerações. Uma interessante consideração dos resultados é que eles refletem a transferência de habilidade não observada dos pais para os filhos. Os principais resultados são:

O impacto direto da educação da mãe é 2,36 vezes inferior aos indiretos capturados por  $\rho^m$ . Ou seja, esforço e habilidade são mais importantes para o desempenho educacional. Por outro lado, observa-se que a educação do pai somente tem efeito direto no resultado educacional do filho, embora, tal efeito direto supera a influência direta da educação da mãe. Tais resultados diferem dos encontrados por Farré et alii (2013), no qual realizaram o estudo para os Estados Unidos concluindo que o efeito direto e indireto da educação da mãe é igual, além de que, não há, segundo eles, impactos diretos da educação do pai sobre a do filho. Todavia, destaca-se que tal influência resguarda fatores como motivação, apoio, entre outros, fatores esses mais suscetíveis a serem transmitidos pela mãe.

Na estimação não paramétrica há uma indefinição entre a dominância da variável sexo. Ou seja, parece não haver diferença significativa no desempenho médio entre os alunos do sexo masculino e feminino. No entanto, os resultados paramétricos, tanto em OLS como em KV, sugerem que os estudantes do



sexo feminino levam ligeira vantagem no desempenho educacional. Por outro lado, quando se analisa a raça a qual o estudante se declara, os estudantes considerados não negros apresentam uma considerável vantagem no desempenho educacional sobre estudantes negros. Tais resultados apresentam-se mais robustos que em outros trabalhos, por exemplo, Farré et alii (2013), pois eles, somente encontram efeitos significativos de sexo e raça após estimar com a função controle,<sup>5</sup> embora, a direção dos resultados sejam os mesmos.

A maior parte dos resultados encontrados apresenta-se consistentes com a teoria, uma vez que reflete a correlação positiva apontada na literatura entre a educação dos pais e dos filhos, impactadas tanto pelas características observáveis como pelas não observáveis como: habilidade e motivação. Isto é, vai ao encontro a textos seminais como Hanushek (1979), como também trabalhos recentes como Lefranc et alii (2009) e Barros (2009). Além do mais, a estimação KV confirma que os coeficientes OLS da educação dos pais são superestimados, embora, difira no aspecto comumente encontrado na literatura de que a educação da mãe impacta mais fortemente do que a educação do pai no desempenho educacional do filho. Outro importante resultado é que somente observa-se, quanto ao pai, a influência direta estatisticamente significativa da escolaridade sobre o desempenho escolar do filho. Ou seja, não há indícios de efeitos indiretos da educação do pai.

Os resultados apontam a qualidade das escolas brasileiras como um dos principais fatores para desigualdade apontada. Não obstante, quando se olha para a variável "tipo de escola" verifica-se que essa tem influência significativa sobre o desempenho no exame. Em resumo, estudar em escola pública é uma grande desvantagem. Ao assumir esse fato como verdadeiro, diante dos resultados empíricos apresentados destaca-se "o porquê disso". Assim, parece plausível aceitar que a qualidade de nossas escolas pode ser a grande vilã da desigualdade de oportunidade educacional do Brasil.

Estes resultados possuem grandes implicações políticas. Por exemplo, pode-se vislumbrar um cenário em que a desigualdade de oportunidades, resultante da origem familiar, pode ser suavizada com políticas que procurem minimizar a papel da família sobre o resultado educacional do indivíduo, melhorando a qualidade das escolas públicas. Além disso, políticas voltadas a reduzir a desigualdade de renda, também tem impacto positivo para atenuar tais desigualdades de oportunidades educacionais.

## 5. CONSIDERAÇÕES FINAIS

Nesse trabalho, analisou-se o grau de igualdade de oportunidades no desempenho educacional advindos dos fatores sociais dos estudantes brasileiros que prestaram o exame do ENEM. Procurou-se investigar se a origem familiar tem papel relevante no desempenho dos alunos que concluíram o Ensino Médio no Brasil e no Nordeste. A partir dos microdados do ENEM e do SAEB (2010), lançou-se mão de duas estratégias empíricas a fim de obter indicativos do papel das circunstâncias sociais no resultado econômico.

A amostra foi composta por 2.937.253 observações, dos quais 60% são mulheres (40% são do sexo masculino); 80% vêm de escola pública e, aproximadamente, 71% dos candidatos possuem renda familiar de no máximo três salários mínimos. Para a primeira estratégia, fez-se o recorte dos indivíduos em baixo *background* social, sendo esses filhos de mães analfabetas, com renda familiar menor do que 1 (um) salário mínimo e que estudou predominantemente em instituição pública; e alto *background* social, ou seja, estudantes filhos de mães com nível superior, renda familiar acima de 9 salários mínimos e que estudou predominantemente em escola particular.

Os resultados indicaram que variáveis como: renda familiar, escolaridade dos pais, tipo de escola, etc., são fatores essenciais na determinação da desigualdade de oportunidades. Pelos resultados (em termos de probabilidade) parecem inviável aqueles penalizados pelas circunstâncias obterem melhores resultados. Por exemplo, um indivíduo com baixo *background* precisa esforçar em torno de 99,38% a

<sup>5</sup>As funções controle ou impacto de controle são dadas por  $A^j(X)$ , em que se concentram em covariâncias condicionais. Ver equação (15) da metodologia.



mais do que um indivíduo com alto *background* para estar entre os 5% com melhores notas. Destaca-se, também, que tais resultados tornam-se mais expressivos quando analisados para a Região Nordeste. Sabendo que a maior parte dos estudantes possuem características que os classificam em baixo *background* social, conclui-se, *a priori*, que há uma baixa mobilidade intergeracional educacional. Os resultados nos levam a crer que alguns estudantes brasileiros são sumariamente excluídos do estrato social superior.

Buscando investigar o efeito indireto da educação dos pais, além de corrigir o viés ao estimar por OLS, aplicou-se um instrumental capaz de mensurar a omissão de variáveis de esforço individual no primeiro conjunto de resultados. Nessa nova estratégia, verificou-se que quanto melhores o conjunto de oportunidade (expressos no *background* familiar – escolaridade e renda dos pais), maiores são os desempenhos no ENEM. Observou-se um impacto positivo para aqueles indivíduos que moram em área urbana, do sexo feminino, declarados não negros e que estudaram em escolas privadas e com melhores resultados no SAEB. Os resultados encontrados abrem caminho para outras questões relacionadas à igualdade de oportunidades, como o papel do sistema de cotas segundo os princípios de recompensa e compensação. Adicionalmente, estimula as discussões a respeito das políticas voltadas à melhoria da qualidade das escolas como alternativa para suavizar o papel da família sobre resultado educacional, visto que o Brasil ainda apresenta cenário com baixo nível de renda e educação.

**BIBLIOGRAFIA**

- Aguirreche, A. L. (2012). Inequality of opportunity in education. Empirical applications and policy, Universidade Del Pais Vasco, Biscay, Espanha.
- Albernaz, A., Ferreira, F. H. G., & Franco, C. (2002). Qualidade e equidade na educação fundamental brasileira. *Pesquisa e Planejamento Econômico*, 33:453–476.
- Angrist, J. D. & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106:979–1014.
- Barros, R. P. (2009). *Measuring inequality of opportunities in Latin America and the Caribbean*. World Bank Publications.
- Bourguignon, F., Ferreira, F. H. G., & Menendez, M. (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth*, 53:585–618.
- Cameron, A. C. & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69:1127–1160.
- Cecchi, D., Peragine, V., & Serlenga, L. (2010). Fair and unfair income inequalities in Europe. Discussion paper, Forschungsinstitut zur Zukunft der Arbeit.
- Diaz, M. D. M. (2012). (des)igualdades de oportunidades no ensino médio brasileiro: Escolas públicas e privadas. *Economia*, 13:553–568.
- ENEM (2010). Microdados. Acesso em 2012, disponível em INEP: <http://portal.inep.gov.br/>.
- Farré, L., Klein, R., & Vella, F. (2013). A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY. *Empirical Economics*, 44:111–133.
- Ferreira, F. H. G. & Gignoux, J. (2008). Towards an understanding of socially-inherited inequalities in educational achievement: Evidence from Latin America and the OECD. Background paper, World Bank, Washington DC.
- Ferreira, F. H. G. & Gignoux, J. (2011). The measurement of educational inequality. *Review of Income and Wealth*, 57:622–657.
- Fleurbaey, M. & Schokkaert, E. (2009). Unfair inequalities in health and health care. *Journal of Health Economics*, 28:73–90.
- Gamboa, L. F. & Waltenberg, F. D. (2012). Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006-2009. *Economics of Education Review*, 31:694–708.
- Hanushek, E. A. (1970). The production of education, teacher quality and efficiency. In *U.S. Office of Education, Do Teachers Make a Difference?*, pages 79–99. Government Printing Office, Washington, DC.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14:351–388.
- Hanushek, E. A. (2007). Schools and the equal opportunity problem. In Woessmann, L. & Peterson, P. E., editors, *Schools and the Equal Opportunity Problem*, pages 159–190. MIT Press, Cambridge.



- Klein, R. & Vella, F. (2009). Estimating the return to endogenous schooling decisions via conditional second moments. *Journal of Human Resources*, 44:1047–1065.
- Klein, R. & Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154:154–164.
- Lefranc, A., Pistoiesi, N., & Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics*, 93:1189–1207.
- O'Neill, B., Sweetman, D., & van de Gaer, D. (2001). Equality of opportunity and Kernel density estimation: An application to intergenerational mobility. In fombo, T. & Hill, C., editors, *Applying Kernel and nonparametric estimation to economic topics*, volume 14 of *Advances in Econometrics*. JAI Press, Stanford, Conn.
- Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press, Cambridge.
- SAEB (2010). Microdados. Acesso em 2012, disponível em INEP: <http://portal.inep.gov.br/>.
- Trede, M. (1988). Making mobility visible: A graphical device. *Economic Letters*, 59:77–82.