

UNIVERSIDADE FEDERAL DA PARAÍBA

Centro de Ciências Sociais Aplicadas

Departamento de Economia

Programa de Pós-Graduação em Economia

LAURO CÉSAR BEZERRA NOGUEIRA

**ENSAIOS SOBRE DESIGUALDADE DE OPORTUNIDADES
EDUCACIONAIS E DE RENDA**

JOÃO PESSOA – PB

2015

LAURO CÉSAR BEZERRA NOGUEIRA

**ENSAIOS SOBRE DESIGUALDADE DE OPORTUNIDADES
EDUCACIONAIS E DE RENDA**

JOÃO PESSOA – PB

2015

LAURO CÉSAR BEZERRA NOGUEIRA

**ENSAIOS SOBRE DESIGUALDADE DE OPORTUNIDADES
EDUCACIONAIS E RENDA**

Tese apresentada ao programa de Pós-Graduação em Economia, Centro de Ciências Sociais Aplicadas, Universidade Federal da Paraíba, em cumprimento aos requisitos necessários para obtenção do título de Doutor em Ciências Econômicas.

Orientador: Erik Figueiredo

JOÃO PESSOA – PB

2015

N778e Nogueira, Lauro César Bezerra.

Ensaio sobre desigualdade de oportunidades educacionais e renda / Lauro César Bezerra Nogueira.- João Pessoa, 2015.

108f. : il.

Orientador: Erik Figueiredo

Tese (Doutorado) - UFPB/CCSA

1. Economia. 2. Transmissão intergeracional da educação. 3. Igualdade de oportunidades educacionais. 4. Habilidade cognitiva geral. 5. Ambiente compartilhado.

UFPB/BC

CDU: 33(043)

LAURO CÉSAR BEZERRA NOGUEIRA

ENSAIOS SOBRE DESIGUALDADE DE OPORTUNIDADES EDUCACIONAIS E RENDA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba como parte dos requisitos necessários à obtenção do título de Doutor em Economia. Submetida e _____ pela Comissão Examinadora abaixo assinada.

Defesa realizada no Campus I da UFPB em João Pessoa, Paraíba, no dia 06 de fevereiro de 2015.

Dr. Erik Alencar de Figueiredo
Orientador

Dr. Cláudio Djissey Shikida (IBMEC)
Membro Externo

Dr. Francisco Soares de Lima – (UERN)
Membro Externo

Dr. Magno Vamberto Batista da Silva
Membro Interno

Dr. José Luis da Silva Netto Júnior
Membro Interno

DEDICATÓRIA

Dedico a minha mãe Maria Aldivaniza e aos
meus filhos João Victor e Maria Clara.

AGRADECIMENTOS

Primeiramente agradeço a Deus por me encorajar nos momentos mais difíceis que enfrentei nessa caminhada.

À minha família, em especial, à minha mãe Maria Aldivaniza por todo apoio oferecido durante toda a jornada acadêmica.

Ao meu orientador Erik Figueiredo por ter acreditado no meu trabalho desde o primeiro momento. E ao qual esse trabalho não seria possível sem a sua valiosa contribuição.

À secretaria do Programa de Pós-Graduação em Economia, em especial, à Risomar Farias por ser essa pessoa maravilhosa.

Ao professor Ivan Targino Moreira pela força ofertada no início dessa jornada. Jamais esquecerei.

Aos amigos do Núcleo de Estudos em Economia Social, em especial, a minha querida amiga Fernanda Leite Santana, uma vez que, esse trabalho contém sua contribuição essencial.

A todos os colegas e professores do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba, em especial, aos professores José Luiz, Hilton Brito, Ignácio Tavares e Magno Vamberto.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela bolsa de pós-graduação concedida.

Quando muito ainda é pouco. Você quer
infantil e louco. Um sol acima do Sol.
“Chico Amaral / Samuel Rosa”.

RESUMO

Este estudo propõe uma investigação sobre a clássica teoria de igualdade de oportunidades. Consideram-se informações referentes às oportunidades educacionais e de renda. Para tanto, utilizam-se quatro bases de dados distintas associadas a algumas estratégias empíricas. Diante disso, este estudo foi dividido em três etapas: i) na primeira, verifica-se o nível de transmissão intergeracional da educação em diversas economias. Adicionalmente, mensura-se a influência indireta de fatores circunstanciais sobre o esforço individual despendido. Além do mais, apura-se qual o efeito de um incremento socioeconômico no resultado do teste; ii) na segunda etapa, destaca-se o papel das características sociais e econômicas no desempenho educacional. Em especial, verifica-se a influência indireta da educação parental sobre o resultado individual. E, por fim, na última etapa, investiga-se se a denominada sorte opção tem influência determinística no resultado individual. Os resultados encontrados apontam diferentes padrões de transmissão intergeracional da educação. Em síntese, apuram-se baixos níveis de transmissão educacional nos países sul-americanos. Por outro lado, observa-se um efeito significativo das circunstâncias sobre o esforço individual empregado e uma forte participação das características sociais e econômicas no desempenho educacional. Por último, constatou-se uma parcela substancial da sorte opção na determinação da renda, como também, uma importante contribuição das características não observáveis no resultado econômico individual.

Palavras-chave: Transmissão Intergeracional da Educação. Igualdade de Oportunidades Educacionais. Sorte Opção. Habilidade Cognitiva Geral. Ambiente Compartilhado.

ABSTRACT

This study proposes a research on classical theory of equal opportunities. The expression information related to educational opportunities and income. For this purpose, are used four different databases associated with some empirical strategies. Thus, this study was divided into three stages: i) in the first, it appears the intergenerational transmission level of education in several economies. In addition, measures to indirect influence of environmental factors on individual effort expended. Moreover, clears up what the effect of a socio-economic growth in the test result; ii) in the second stage, we highlight the role of social and economic characteristics in educational performance. In particular, it appears the indirect influence of parental education on individual results. And finally, in the last step, it investigates the option called luck has deterministic influence on individual results. The results show different patterns of intergenerational transmission of education. In short, shall discharge to low levels of educational transmission in South American countries. On the other hand, there is a significant effect of the circumstances of the individual employee effort and a strong involvement of social and economic characteristics in educational performance. Finally, there was a substantial portion of luck option in determining the income, but also an important contribution of unobservable characteristics in individual economic results.

Keywords: Education Intergenerational Transmission. Equal Educational Opportunities. Option luck. General Cognitive Ability. Shared environment.

LISTA DE FIGURAS

Figura 2.1	Influência Educacional Parental no Teste PISA – Média Geral	33
Figura 2.2	Influência Educacional Parental.....	35
Figura 3.1	Conjunto de Oportunidades – Brasil e Regiões	59

LISTA DE TABELAS

Tabela 2.1	Desempenho Médio – Teste Scores – PISA 2012.	28
Tabela 2.2	Análise Descritiva.....	31
Tabela 2.3	Determinantes do Resultado Individual – Média Geral – Circunstancial.	32
Tabela 2.4	Determinantes do Resultado Individual – Média Geral.	36
Tabela 2.5	Decomposição da Participação do Conjunto de Oportunidades	38
Tabela 2.6	Equalização das Oportunidades.....	39
Tabela 2.7	Decomposição do Efeito das Circunstâncias por Países no PISA.....	41
Tabela 2.8	Efeito Tratamento Médio Quantílico – Circunstâncias – ETMQ.....	42
Tabela 3.1	Análise Descritiva.....	57
Tabela 3.2	Probabilidade Condicional Acumulada – Brasil e Regiões.....	61
Tabela 3.3	Probabilidade Condicional Acumulada – Brasil.....	62
Tabela 3.4	Estimação KV e OLS	64
Tabela 4.1	Testes de Especificação do Modelo.....	86
Tabela 4.2	Análise Descritiva.....	86
Tabela 4.3	Diferença nas Médias	87
Tabela 4.4	Distribuição do Propensity Score	87
Tabela 4.5	Impacto da Sindicalização	88
Tabela 4.6	Propensity Score Matching – Sindicalização	88
Tabela 4.7	Efeito Tratamento Sobre o Tratado – ATT – Sorte Opção.....	89
Tabela 4.8	Influência da Sorte Opção no Salário Hora	90

LISTA DE QUADROS

Quadro 2.1 – Descrição das Variáveis.....	30
Quadro 3.1 – Conjunto de Informações.....	56
Quadro 4.1 – Descrição das Variáveis.....	85

LISTA DE ABREVIATURAS E SIGLAS

ATET	Efeito Tratamento sobre o Tratado
ENEM	Exame Nacional do Ensino Médio
ETMQ	Efeito de Tratamento Médio Quantílico
FPE	Função de Produção de Educação
IDC	Inferências Contrafatuais
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IOPs	Igualdade de Oportunidades
KV	Klein e Vella
OCDE	Organização para Cooperação e Desenvolvimento Econômico
OLS	Ordinary Least Square
PISA	Programa Internacional de Avaliação de Estudantes
PSM	Propensity Score Matching
RIA	Roemer Identification Axiom
SAEB	Sistema de Avaliação da Educação Básica
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

SUMÁRIO

1	INTRODUÇÃO	16
2	TAL PAI, TAL FILHO? UMA ANÁLISE DOS EFEITOS FATORES DE CIRCUNSTÂNCIAS SOBRE O DESEMPENHO DOS ALUNOS NA AVALIAÇÃO DO PISA 2012.	18
	2.1 INTRODUÇÃO.....	18
	2.2 REVISÃO DA LITERATURA.....	20
	2.3 ESTRATÉGIA EMPÍRICA	22
	2.3.1 Descrição e Análise dos Dados	27
	2.4 PRINCIPAIS RESULTADOS	31
	2.5 ALGUMAS CONSIDERAÇÕES	43
3	IGUALDADE DE OPORTUNIDADES: ANALISANDO O PAPEL DAS CIRCUNSTÂNCIAS NO DESEMPENHO DO ENEM.	45
	3.1 INTRODUÇÃO.....	45
	3.2 IGUALDADE DE OPORTUNIDADES: UMA BREVE REVISÃO.....	46
	3.3 PROCEDIMENTOS TEÓRICOS E METODOLÓGICOS	49
	3.3.1 Modelo Bivariado	49
	3.3.2 Função de Produção Educacional.....	53
	3.3.3 Descrição dos Dados	55
	3.4 PRINCIPAIS RESULTADOS	58
	3.4.1 Primeiro Bloco	58
	3.4.2 Segundo Bloco	63
	3.5 CONSIDERAÇÕES FINAIS	66
4	A SORTE SORRI PARA TODOS? UMA NOVA ABORDAGEM EMPÍRICA DE DESIGUALDADE DE OPORTUNIDADES.....	68
	4.1 INTRODUÇÃO.....	68
	4.2 REVISÃO DE LITERATURA	71
	4.2.1 Igualdade de Oportunidades	71
	4.2.2 Sorte.....	73
	4.2.3 Habilidade Cognitiva Geral, Herdabilidade Genética e Ambiente Compartilhado	77

4.3 ESTRATÉGIA EMPÍRICA	79
4.4 Descrição e Análise dos Dados	85
4.5 PRINCIPAIS RESULTADOS	86
4.6 ALGUMAS CONSIDERAÇÕES	90
5 CONCLUSÃO	92
REFERÊNCIAS	94
APÊNDICE A: Resultados Adicionais	99
APÊNDICE B: Resultados Adicionais.....	108

1 INTRODUÇÃO

Este trabalho de tese é constituído por três ensaios relacionados à moderna teoria de desigualdade de oportunidades educacionais, de renda e transmissão intergeracional da educação. Nessa linha de pensamento, destacam-se algumas particularidades fundamentais em cada um deles. No primeiro ensaio, a transmissão intergeracional da educação e desigualdade de oportunidade educacional será avaliada tanto sob a ótica da influência direta das características individuais observáveis, como também, estima-se a influência indireta desses predicados sobre o esforço observado. Adicionalmente, verifica-se qual a consequência de um incremento no *background* socioeconômico na desigualdade observada no exame do Programa Internacional de Avaliação de Estudantes – PISA – versão 2012. No ensaio seguinte, investiga-se qual a participação das características socioeconômicas no desempenho obtido no Exame Nacional do Ensino Médio – ENEM – versão 2010. Além do mais, busca-se mensurar a participação indireta da educação parental no desempenho alcançado no teste. E, por fim, no terceiro ensaio, observa-se a desigualdade de oportunidades analisando parte da sorte – *sorte opção* – como fator determinístico e não aleatório. Além disso, neste último, edifica-se um cenário empírico que contempla fatores não observáveis como, habilidade cognitiva geral, herdabilidade genética, ambiente compartilhado, entre outros.

O primeiro ensaio é motivado por dois problemas fundamentais. *A priori*, tem como finalidade responder o que determina o alto diferencial de desempenho educacional observado entre os jovens dos países participantes do PISA. Nessa perspectiva, procura-se verificar se a transmissão intergeracional da educação é um fator decisivo na desigualdade de oportunidades educacionais. Se sim, encontrar quais são os principais fatores que contribuem para isto. Além disso, objetiva-se examinar se as circunstâncias exercem efeitos indiretos sobre o resultado do exame. Isto é, se tais fatores influenciam de modo significativo o esforço empregado na obtenção do resultado.

A fim de responder tais questionamentos, o primeiro ensaio, investiga a transmissão intergeracional da educação de 58 economias distintas. Nele, a desigualdade de oportunidades educacional será avaliada considerando os efeitos diretos e indiretos das circunstâncias. Destaca-se que um amplo conjunto de variáveis que expressam as circunstâncias e esforço individual será utilizado neste estudo. Em seguida, verifica-se qual o efeito de um incremento no *background* socioeconômico individual na redução da desigualdade observada. Contudo, é plausível admitir que haja fatores observáveis e não disponíveis na análise, como também,

características individuais não observáveis, que possam estar limitando os resultados encontrados. Diante disso, o segundo ensaio busca também preencher parte dessa limitação.

Nesse contexto, no segundo ensaio, pretende-se incorporar na análise fatores como habilidade, ambiente compartilhado, entre outros. Em linhas gerais, avalia-se o quanto o ambiente individual é responsável pela desigualdade de oportunidade educacional no Brasil, levando em conta os efeitos diretos e indiretos do background socioeconômico individual, em especial, nesse último, o resultante da educação parental. Em outras palavras, busca-se mensurar o quanto as características socioeconômicas respondem pelo resultado educacional individual. Além do mais, estima-se a influência indireta da educação parental no resultado educacional obtido no exame do ENEM.

Quanto ao terceiro ensaio, tem-se como principal motivação a controvérsia existente entre a literatura filosófica e econômica referente às componentes do termo de erro. Ou seja, procura-se verificar se há variações da sorte que exercem efeito determinístico no resultado individual? Se sim, qual a magnitude desse efeito.

Nesse sentido, as principais integrantes do fator sorte destacadas na literatura econômica e filosófica, que em geral, compõem o termo de erro, nos tradicionais modelos de Igualdade de Oportunidades – IOPs – serão desagregadas em fatores aleatórios e determinísticos. Nesse contexto, espera-se obter um considerável avanço nesse estudo ao trabalhar-se com uma base de indivíduos composta por gêmeos monozigóticos. Ou seja, fatores não observáveis, como também, fatores observáveis, mas não disponíveis na maior parte das bases de dados, serão considerados. Dito de outro modo, fatores até então ignorados, ou no mínimo, negligenciados como, por exemplo, habilidade cognitiva geral, herdabilidade genética, ambiente compartilhado, entre outros serão incorporados na análise.

Contudo, destaca-se que nos dois primeiros ensaios – oportunidades educacionais – utiliza-se como variável de resultado o desempenho educacional e não a realização educacional como visualizado na maior parte dos estudos voltados para o problema em questão. A principal razão para isso dar-se ao fato que a realização educacional, em muitos casos, não necessariamente reflete a qualidade da educação, e, portanto, a obtenção de conhecimento e habilidade individual. Dado que, esses desígnios são incisivos na aquisição de renda e avanço do bem estar social.

2 TAL PAI, TAL FILHO? UMA ANÁLISE DOS EFEITOS FATORES DE CIRCUNSTÂNCIAS SOBRE O DESEMPENHO DOS ALUNOS NA AVALIAÇÃO DO PISA 2012.

2.1 INTRODUÇÃO

Normalmente economistas e cientistas sociais diferem da maior parte das teorias econômicas, porém, educação de qualidade é consensualmente vista como forte indicador de bem estar. Diante disso, a desigualdade educacional entre as regiões, países e continentes determina o objetivo e caráter de diversas políticas públicas.

Nessa linha de pensamento, Daude (2011), advoga que o acesso à educação de qualidade é uma ferramenta poderosa para incrementar o bem estar individual. E a provisão social desta para todos é uma necessidade insubstituível para equalização de oportunidades, notadamente, dos menos favorecidos.¹ Nesse sentido, devem-se garantir determinadas condições para que isso ocorra. Por exemplo, todos os indivíduos precisam ter acesso a uma educação de qualidade homogênea, independentemente do seu conjunto de oportunidades. Além disso, a sociedade deve avalizar a supremacia do mérito, de forma que, o talento e as habilidades individuais prevaleçam. Pois, sem essas garantias, os retornos dos investimentos educacionais são insatisfatórios, especialmente, para aqueles mais vulneráveis da sociedade, e, portanto, reduz-se a mobilidade intergeracional.

Em outras palavras, a educação é vista como elemento chave para o desenvolvimento econômico e social de qualquer sociedade. Uma vez que, uma força de trabalho qualificada eleva a produtividade das atividades econômicas. Possibilitando assim, alto crescimento dos setores e facilitando a absorção de tecnologia e inovação. Além desses aspectos, a educação também é característica essencial para o bom funcionamento da democracia, permitindo às pessoas exercerem plenamente os seus direitos e responsabilidades como cidadãos.

Entretanto, o relatório da Organização das Nações Unidas para a Educação, a Ciência e a Cultura – UNESCO – versão 2005, enfatiza que, apesar do crescente consenso em torno da importância da qualidade educacional, o nível de ações quanto a esse conceito é bem

¹ Por exemplo, em Anshenfelter e Rouse (1998), os autores avaliam qual o papel da escolaridade como potencial equalizador da correlação intergeracional da renda.

menor na realidade. Segundo o relatório, dois princípios distinguem geralmente as formas de definir a qualidade da educação. O primeiro estabelece o desenvolvimento cognitivo dos alunos como principal objetivo explícito de qualquer sistema educacional. Reconhecendo o sucesso alcançado por estes últimos como indicativo de sua qualidade. O segundo destaca o papel da educação na promoção de valores compartilhados e no desenvolvimento criativo e emocional. Nesse caso, a consecução desses objetivos é bem mais complexa de analisar.

Focando nesses conceitos, recentemente, cientistas sociais, tem estudado a transmissão intergeracional da educação. Por exemplo, Niknami (2010), Ferreira e Veloso (2003), Black, Devereux e Salvanes (2005). Entretanto, os estudos ainda são bastante esparsos e inconclusivos. Além disso, a maior parte desses estudos utiliza realização educacional – anos de estudo – ao invés de desempenho educacional. Porém, os resultados sugerem baixa transmissão educacional, em especial, nos países em desenvolvimento. Esses resultados contribuem para persistência intergeracional da educação, e, portanto, para elevar a desigualdade de oportunidades educacionais.

Contudo, podem-se considerar diversos fatores responsáveis pela baixa transmissão educacional, em especial, supor baixa qualidade educacional nesses países. Pois, embora nos últimos anos tenha crescido o grau de escolaridade nessas economias, a realização educacional não necessariamente garante equivalência de oportunidades. Visto que, anos de estudo não reflete a qualidade educacional de um sistema, essencialmente, devido à heterogeneidade existente entre as escolas em suas respectivas regiões, países e continentes.²

Diante disso, compreender os mecanismos subjacentes por trás dessa relação – transmissão intergeracional da educação – é de fundamental importância. Nesse sentido, o presente estudo pretende investigar três importantes questões. Primeiro, o processo de transmissão intergeracional da educação no exame PISA 2012 associado à teoria de igualdade de oportunidades educacionais.³ Para tanto, considera-se a função de produção educacional proposta por Hanushek (1970,1979), com algumas peculiaridades.⁴ Segundo, apurar se as circunstâncias exercem efeitos sobre as variáveis de esforço no desempenho da avaliação PISA. Nesse contexto, adota-se uma proposta idêntica à desenvolvida por Bourguignon, Ferreira e Menéndez (2007), adaptado aqui ao contexto de oportunidades educacionais. Terceiro, construir uma análise contrafactual proveniente de um aumento no nível

² Por exemplo, Ferreira e Gignoux, (2008).

³ Por exemplo, Ferreira e Gignoux, (2011).

⁴ Foram utilizadas diversas variáveis de controle, em especial, que representam o esforço individual.

socioeconômico parental dos estudantes que prestaram o exame do PISA 2012. Para isso, aplica-se a metodologia de inferência contrafatuais desenvolvida por Chernozhukov, Fernandez-Val e Melly (2013).

Destaca-se haver importantes diferenciais nesse estudo. Primeiro, verificar o padrão de transferência intergeracional da educação entre as economias participantes do PISA 2012. Segundo, decompor os efeitos das circunstâncias – diretos e indiretos – sobre o desempenho no exame. Terceiro, simular o efeito contrafactual de possíveis políticas públicas educacionais. Por fim, o ensaio está organizado como segue. Além desta introdução, na segunda seção faz-se uma breve revisão sobre transmissão e desigualdade de oportunidades educacionais, com atenção especial, às pesquisas que utilizam dados do PISA. A terceira seção é reservada a detalhar a metodologia utilizada, incluindo a descrição e análise dos dados. A quarta é destinada a apresentação dos resultados. E a última seção é dedicada a algumas considerações extraídas dos resultados obtidos.

2.2 REVISÃO DA LITERATURA

Nesta seção serão apresentados alguns trabalhos voltados para o tema. Inicia-se destacando, o estudo de Black, Devereux e Salvanes (2005). Os autores investigam por que pais com níveis educacionais mais elevados têm crianças mais educadas. Segundo o estudo, há várias explicações possíveis. Entretanto, destacam-se duas. Primeiro, seria o caso de seleção pura ou efeitos indiretos. Isto é, pais mais educados, ganham salários mais elevados, e, portanto, reflete em variáveis que irão contribuir substancialmente para a educação dos filhos. Por exemplo, matricular os filhos nas melhores escolas, comprar os melhores livros, como também, investir em mecanismos que auxiliem o aprendizado. Segundo, o denominado efeito causalidade ou direto. Nesse caso, alcançar mais educação faz de você um tipo diferente de pai, e, assim, induz seus filhos a resultados educacionais mais elevados. Essa relação direta de causalidade ocorre através de fatores potencialmente não observáveis, como ambientes compartilhados e herdabilidade genética.

Por outro lado, segundo Roemer (1998), atualmente, dois conceitos de igualdade de oportunidades prevalecem nas democracias ocidentais. O primeiro afirma que todo indivíduo com potencial relevante no seu período de formação deve ser admitido como possível candidato a competir por posições na sociedade. O segundo, denominado princípio não

discriminatório, estabelece que na competição por posições na sociedade todo indivíduo que possui atributos relevantes para exercer uma determinada função deve ser incluído como candidato elegível e que somente será julgado por características relevantes.

Guiado por esta linha teórica, Lefranc, Pistolesi e Trannoy (2009), defendem que a igualdade de oportunidades constitui um princípio básico para atenuar as desigualdades entre indivíduos. Em que os resultados dependem de um conjunto de fatores determinísticos e aleatórios relevantes para o sucesso ou fracasso do agente econômico. Para analisar a igualdade de oportunidades é preciso identificar as variáveis que estão sob a responsabilidade dos indivíduos, e as que não estão.

Todavia, destaca-se a existência de diversos estudos sobre os determinantes no desempenho educacional, inclusive utilizando dados de logro educacional, porém, a maior parte destes não passa de meros registros descritivos. Por exemplo, em Aguirreche (2012), investiga-se o modo como o grau de desigualdade de oportunidades de uma nação afeta o desempenho médio dos seus estudantes. 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.

Em outro estudo, envolvendo dados do PISA 2006-2009, Gamboa e Waltenberg (2011), avaliam a desigualdade de oportunidades educacional em seis países da América Latina. Para tanto, utilizam uma abordagem não paramétrica proposta por Checci e Peragine (2010), para decompor um índice de desigualdade. O estudo verifica que o grau de desigualdade de oportunidade educacional, varia entre 1% a 25%, o que denota uma considerável heterogeneidade entre os seis países.

De modo semelhante, Ferreira e Gignoux (2008), empregando dados do PISA 2000, calculam a parcela da desigualdade observada no desempenho escolar que é determinada pelo conjunto de oportunidades. Em suma, constata-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 da OCDE. 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 dois métodos de inferir a desigualdade educacional. O primeiro é voltado para o desempenho educacional – variância ou desvio padrão – e, o segundo para a oportunidade educacional – parcela da variância – que explica a influência das circunstâncias. Os resultados reportam que 35% de todas as

disparidades no desempenho educacional dos 57 países participantes do PISA 2006 ocorrem em função da desigualdade de oportunidades.

Em linhas gerais, verifica-se que o conjunto de oportunidades é decisivo para o desempenho educacional dos indivíduos. Tais aspectos são mais incisivos 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. Em que, a educação dos pais é a circunstância mais importante que afeta os resultados.

2.3 ESTRATÉGIA EMPÍRICA

A estrutura teórica tradicional da desigualdade de oportunidades, sugerida por Roemer (1998), considera que o resultado individual, w_i , é explicado por um conjunto de variáveis: i) de circunstância, C_i , por exemplo, background familiar, gênero, região de nascimento, etc.; ii) por fatores de esforço, E_i , como, horas trabalhadas, dispêndio em tempo de leitura, etc. e; iii) por um termo de sorte bruta, u_i . Onde $f(.)$ é uma função desconhecida.

$$w_i = f(C_i, E_i, u_i) \quad (2.1)$$

Entretanto, associando esses conceitos à teoria de transmissão intergeracional da educação. E adaptando-os a Função de Produção de Educação, doravante FPE, proposta Hanushek (1970, 1979), com algumas especificidades, tem-se:

$$\ln(W_{it}) = g(F_i^t, P_i^t, I_i, S_i^t) \quad (2.2)$$

Em que W_{it} é o vetor educacional realizado pelo estudante $i - th$ no tempo t ; F_i^t é o vetor de características individuais e familiares do $i - th$ estudante acumulado no período t ; P_i^t é o vetor de influências 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 $i - th$ estudante acumulados no período t . Na verdade, a FPE analisa a forma como os diversos insumos do processo educacional podem afetar os resultados educacionais dos indivíduos.

Todavia, ao acrescentar variáveis de controle que expressem as circunstâncias e esforços individuais torna-se idêntica a (2.3).

$$\ln(W_{f,i}^*) = \alpha + \beta_i W_{p,i}^* + \varepsilon_i, \quad (2.3)$$

Onde $\ln(W_{f,i}^*)$ representa o log do desempenho do filho ao longo do tempo da família i ; $(W_{p,i}^*)$ um vetor de características socioeconômico do filho, como, o log da educação parental, tipo de escola que frequenta, atendimento pré-escolar, etc. e ε_i um termo ortogonal a $W_{f,i}^*$. Além disso, supõe-se:

$$E(\varepsilon_i) = 0, E(\varepsilon_i, W_{p,i}) = 0 \text{ e } E(\varepsilon_i^2) = \sigma_\varepsilon^2, \quad (2.4)$$

De forma que, β_1 reflete o grau de transmissão intergeracional da educação. No entanto, ciente de possíveis fragilidades dessa abordagem, principalmente, devido à omissão de variáveis, como também, possível endogeneidade presente no vetor de covariadas, emprega-se um método baseado em Bourguignon, Ferreira e Menéndez (2007). Isto é, admite-se uma relação de dependência entre as variáveis de circunstância e de esforço. Portanto, relaxa-se a hipótese assumida em (2.4). Matematicamente tem-se:

$$(w_i) = f(C_i, E_i(C_i, v_i), u_i), \quad (2.5)$$

Em seguida, faz-se um link à abordagem de transmissão intergeracional e toma uma especificação log-linear como proposta em (2.3), implicando em:

$$\ln(\hat{w}_{f,i}) = \beta_1 W_{p,i}^* + \beta_2 E_{f,i}^* + u_{f,i}, \quad (2.6)$$

Adicionalmente, assume-se a endogeneidade imposta a determinadas variáveis que representam as circunstâncias individuais resultando em:

$$E_{f,i}^* = \gamma W_{p,i}^* + v_i, \quad (2.7)$$

Onde β_1 e β_2 denota dois vetores de coeficientes. Além disso, u_i e v_i são termos aleatórios com propriedades *white-noise* que denotam variáveis não observáveis de circunstâncias e esforço, como também, o fator sorte. E γ é uma matriz de coeficientes que liga variáveis específicas de circunstâncias as de esforço. Em resumo, assume-se, que os coeficientes desta matriz exercem influência sobre as variáveis de esforço.

Esta estrutura possibilita que algumas variáveis de esforço sejam nitidamente impactadas pelas circunstâncias, tornando assim, viável captar efeitos diretos (β_1) e indiretos ($\beta_2\gamma$) das variáveis de circunstância sobre o desempenho educacional individual. Agora, permita $M(w)$ representar a distribuição marginal do desempenho obtida a partir de (2.5). A seguir, constroem-se dois contrafatuais: i) no primeiro, anulam-se os efeitos totais das circunstâncias, isto é, $\tilde{w}_i = f(\bar{C}_i, E_i(\bar{C}_i, v_i), u_i)$; e ii) no segundo, anulam-se somente os efeitos diretos das circunstâncias, $\tilde{w}_i^d = f(\bar{C}_i, E_i(C_i, v_i), u_i)$. E, portanto, se as expressões

(2.6) – (2.7) poderem ser estimadas de forma prudente, a parcela da desigualdade – persistência das circunstâncias – sobre o desempenho individual é dada por:

$$\aleph_I = \frac{I(M(w)) - I(M(\tilde{w}))}{I(M(w))} \quad (2.8)$$

Em palavras, o vetor do desempenho resultante em (2.8) conterà a desigualdade total resultante das variáveis de circunstâncias. Por outro lado, a equação (2.9) revela a parcela da desigualdade – persistência – proveniente dos efeitos diretos das circunstâncias.

$$\aleph_I^d = \frac{I(M(w)) - I(M(w)^d)}{I(M(w))} \quad (2.9)$$

Então, o efeito indireto das circunstâncias é simplesmente determinado por:

$$\aleph_I^i = \aleph_I - \aleph_I^d \quad (2.10)$$

Felizmente, para realizar as decomposições propostas em (2.8) e (2.9), não é necessário calcular os sistemas estruturais (2.6) - (2.7). Pois, substituindo (2.7) em (2.6) temos a forma reduzida:

$$\ln(\hat{w}_i) = (\beta_1 + \gamma\beta_2)W_{p,i}^* + v_i\beta_2 + u_{f,i} \quad (2.11)$$

Entretanto, conforme destacam os autores, a omissão de variáveis relevantes na determinação do resultado e possível endogeneidade presente nas circunstâncias, tornam os parâmetros obtidos em (2.5) e (2.11) viesados, e, conseqüentemente, a apuração dos contrafatuais também. Posto isso, e na falta de um adequado conjunto de variáveis instrumentais, os autores, sugerem uma simulação de Monte Carlo para criar intervalos de variação dos coeficientes estimados. A fim de superar tais problemas, este estudo, adota o método de pós-estimações – Bootstrap – replicando mil – 1000 – amostras aleatórias a partir da amostra original, e em seguida, construindo um intervalo de confiança para os parâmetros estimados. Segundo, buscou-se superar possíveis vieses na construção dos contrafatuais a partir de uma abordagem quantílica. Isto é, não apenas baseada em estimações dos efeitos médios.

Em seguida, a fim de apurar em que medida um incremento no background social – circunstâncias – afeta o desempenho no teste, visto que, tais fatores mostram-se decisivos nos resultados individuais,⁵ emprega-se a metodologia de inferências contrafatuais desenvolvida por Chernozhukov, Fernandez-Val e Melly (2013), doravante, denominado IDC. A escolha pelo IDC deve-se em função do mesmo se basear em diversas abordagens principais para estimar funções quantílicas condicionais e funções de distribuições condicional. Outra

⁵ Ver por exemplo, Barros *et. al.* (2009).

vantagem é seu emprego tanto para analisar o efeito de simples intervenções – mudança unitária em uma característica – como também, em alterações complexas – mudanças gerais na distribuição das características.

O IDC é usado em especial, em casos onde uma política intervencionista origina uma modificação alterando parte da distribuição do conjunto de variáveis explicativas X – covariadas – que determinam a resposta na variável de interesse Y . Em outras palavras, o IDC consiste em estimar o efeito na distribuição de Y dada uma modificação na distribuição de X .

Os resultados observados são extraídos da amostra antes da intervenção política e, portanto, observável, enquanto que os resultados contrafatuais surgem da amostra após a intervenção política e, portanto, não observável. Em seguida, admite-se que as covariadas são observáveis antes e após a intervenção política. Isto é, os resultados observados são usados para estabelecer a relação entre a variável de interesse e as covariadas, que, juntamente com a distribuição contrafactual observada das covariadas, determinam a distribuição do resultado após a intervenção sob algumas condições impostas.

Segundo os autores, a fim de especificar um modelo que permita verificar como o resultado contrafactual é gerado, torna-se conveniente olhar a relação entre o resultado observado e covariadas utilizando uma representação quantílica condicional. O IDC assume que W^0 representa o resultado observado, e X^0 o $(p \times 1)$ vetor das covariadas com função de distribuição F_X^0 antes da política intervencionista. Onde, $Q_Y(u|X)$ denota o u – quantil condicional de W^0 dado X^0 . De modo que, o resultado W^0 pode ser ligado à função quantil condicional através da representação de Skorohod, isto é:

$$W^0 = Q_W(U^0|X^0), \text{ Onde } U^0 \sim U(1,0), \text{ independente de } X^0 \sim F_X^0 \quad (2.12)$$

Em que (2.12) destaca que o resultado é função das covariadas e do termo de erro U^0 . Sobretudo, em modelos de regressões clássicos, o termo de erro é separado das variáveis independentes como nos modelos de regressão pontual, mas em geral, isso não é necessário. Este método abrange ambos os casos.

Na verdade, o processo de inferências contrafactual consiste em desenhar o vetor de covariadas para uma distribuição diferente, isto é, $X^C \sim F_X^C$, onde F_X^C é uma função de distribuição conhecida das covariadas após a política de intervenção. Assim, sob a suposição que a função quantil condicional não é modificada pela política, o resultado contrafactual Y^C é gerado por:

$$W^C = Q_W(U^C|X^C), \quad (2.13)$$

Onde $U^C \sim U(1,0)$ é independente de $X^C \sim F_X^C$. Adicionalmente, o IDC faz a suposição que a função quantil $Q_Y(u|x)$ possa ser avaliada em cada ponto de x na base da distribuição das covariadas de F_X^C . Esta suposição requer que a base da F_X^C seja subconjunto da base de F_X^0 , ou então, que a função quantil possa ser adequadamente extrapolada. Tais suposições são formalizadas a seguir.

- **S1** – a distribuição condicional do resultado dado as covariadas é a mesma antes e após a política de intervenção;
- **S2** – o modelo condicional é válido para todo $x \in X$, onde X é um subconjunto compacto do $\mathbb{R}^p$ que contem as bases de F_X^0 e F_X^C .

Além disso, destaca-se que o IDC considera dois tipos distintos de mudanças na distribuição das covariadas. Primeiro, as covariadas são desenhadas para uma subamostra diferente antes e após a intervenção. Esta subamostra pode corresponder a diferentes grupos – tipos – demográficos, períodos de tempo ou localizações geográficas. Por exemplo, características de trabalhadores em anos diferentes, distribuição socioeconômicas parental para indivíduos de pele branca e negra, ou mais casualmente, para distribuições das covariadas em grupos de tratamento *versus* grupo de controle. Segundo, a intervenção pode ser empregada como uma transformação conhecida da distribuição de covariadas observadas. Em síntese, $X^C = g(X^0)$, onde $g(\cdot)$ é uma função conhecida. Por exemplo, mudanças unitárias na locação de uma das covariadas, $X^C = X + e_j$, onde e_j é um vetor unitário ($p \times 1$) com um na posição j ; ou ainda, preservar a redistribuição das covariadas implementadas como $X^C = (1 - \alpha)E[X^0] + \alpha X^0$.

Este tipo de intervenção pode ser utilizado para estimar, por exemplo, qual o efeito sobre as despesas de alimentos resultantes de uma alteração no imposto de renda? Qual a influência sobre os preços dos imóveis resultantes da limpeza de resíduos perigosos no entorno? E assim por diante. Note que, os dois casos acima correspondem a cenários experimentais diferentes. A principal diferença é que o segundo caso corresponde a uma experiência quase perfeitamente controlada, que fornece informações adicionais para identificar mais características da distribuição conjunta dos resultados antes e depois da intervenção. Além do mais, para fazer inferências do efeito total no resultado decorrente da intervenção, é preciso identificar as funções distribuição e quantílica do resultado antes e após a política. A função distribuição condicional associada à função quantílica $Q_W(u|x)$ é expressa por:

$$F_W(w|x) = \int_0^1 1\{Q_W(u|x) \leq w\} du, \quad (2.14)$$

Dado as suposições sobre como o resultado contrafactual é gerado, a distribuição marginal é expressa por:

$$F_{W^j}(w) := Pr\{W^j \leq w\} = \int_x F_W(w|x) dF_X^j(x), \quad (2.15)$$

Com funções u – *quantil* marginal correspondente.

$$Q_{W^j}(u) = \inf\{w: F_{W^j}(w) \geq u\}, \quad (2.16)$$

Onde o índice j corresponde ao status antes ou após a intervenção, $j \in \{0, C\}$. O u – *quantil* efeito tratamento da política de intervenção é determinado por:

$$QTE_W(u) = Q_{W_C}(u) - Q_{W_0}(u), \quad (2.17)$$

De forma análoga, o efeito na u – *distribuição* da política é expresso por:

$$DE_W(w) = F_{W_C}(w) - F_{W_0}(w), \quad (2.18)$$

O IDC possibilita apurar outras funcionais de interesse, por exemplo, a curva de Lorenz do resultado contrafactual observado.

$$L_W^j(w) := \frac{\int_{-\infty}^w t dF_W^j(t)}{\int_{-\infty}^{\infty} t dF_W^j(t)}, \quad (2.19)$$

Contando que as integrais existam e $\int_{-\infty}^{\infty} t dF_W^j(t) \neq 0$, $j \in \{0, C\}$. No entanto, geralmente pode ser mais interessante as funções da distribuição marginal do resultado antes e após a intervenção.

2.3.1 Descrição e Análise dos Dados

Os resultados apresentados nesta pesquisa serão possibilitados pelos microdados do PISA, versão 2012. O programa é desenvolvido e coordenado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE). Em cada país participante há uma coordenação nacional. No Brasil, a avaliação é coordenada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). A adoção desta base ocorre em virtude da riqueza apresentada. Pois, o exame PISA é uma pesquisa internacional trienal que tem como objetivo avaliar os sistemas de educação em todo o mundo. Para tanto, busca testar as habilidades e conhecimento de estudantes de 15 anos de idade. A edição 2012 é a quinta edição do programa e avalia as competências dos alunos em leitura, matemática e ciências com ênfase em matemática.

Outro forte atrativo, é que cerca de 510 mil estudantes de 65 economias participaram da edição 2012 representando algo em torno de 28 milhões de jovens. Além disso, ele é considerado único, devido desenvolver atividades que não estão inteiramente ligados ao currículo escolar. Na verdade, os testes são projetados para avaliar em que medida os alunos no final do ensino fundamental podem aplicar seu conhecimento a situações cotidianas da vida real e como estão capacitados para a plena participação na sociedade. Além do mais, as informações coletadas por meio de questionários fornecem elementos auxiliares na interpretação de resultados.

Em relação às análises preliminares dos dados, reportados na tabela 2.1, verifica-se um elevado grau de desigualdade no desempenho médio em todas as linhas de conhecimento, especialmente, em relação a países da América do Sul.

Tabela 2.1 – Desempenho Médio – *Teste Scores* – PISA 2012

Lugar	Matemática	Língua	Ciências	Média Geral
América do Sul	394,39	410,73	408,51	404,54
Argentina	396,47	403,99	410,58	403,68
Brasil	383,42	400,98	395,90	393,43
Chile	445,88	461,01	465,61	457,50
Colômbia	384,98	413,28	407,99	402,08
Peru	367,19	383,70	372,82	374,57
Uruguai	411,96	414,53	419,14	415,58
Países da OCDE	488,30	490,81	494,86	491,33
Alemanha	515,48	509,28	535,37	516,71
Canadá	509,29	510,95	514,36	511,53
Estados Unidos	481,03	497,62	497,57	492,07
França	498,42	509,27	502,53	503,41
Reino Unido	499,18	498,09	509,71	499,18
Espanha	495,82	494,45	504,83	498,37
Japão	534,99	536,67	545,77	539,14
México	418,44	428,84	419,59	422,29
Países Asiáticos	540,20	515,12	525,58	526,97
Shangai	612,03	569,36	580,33	587,24
Singapura	566,89	536,55	545,66	549,70
Hong-Kong	559,83	543,35	554,28	552,49
Chinesa Taipei	557,84	521,22	521,78	533,61
Coréia do Sul	552,61	535,61	537,14	541,78

Fonte: Elaboração própria, a partir de dados do PISA 2012.

Por exemplo, o Brasil, que apesar de possuir uma das dez maiores economias do planeta, segundo o Fundo Monetário Internacional – FMI – apresenta um resultado extremamente modesto. No entanto, destaca-se que os resultados não se modificam

significativamente em relação ao método e relatórios apresentados pelo programa. Pois, de acordo com PISA, o Brasil ocupa a 58^a posição no ranking dos 65 países analisados.

Por outro lado, cabe destacar as excelentes performances da maioria dos países asiáticos que obtiveram médias acima dos países da OCDE. Para ter-se uma ideia, conforme denotam os resultados apresentados, os cinco melhores resultados expostos pertencem a esse grupo. Além do mais, impressiona a diferença de desempenho no exame de matemática, que é o foco do PISA 2012. Especificamente, Shanghai, obteve em torno de 90 pontos acima do Japão, considerado na análise membro da OCDE. Em relação, aos países da América do Sul, os asiáticos obtém um desempenho aproximadamente 30% superior. Por exemplo, comparando o Chile – melhor desempenho entre os sul-americanos – a Singapura verifica-se um diferencial positivo próximo a 28% em favor dos asiáticos. Embora, nesses últimos haja uma maior variância nas notas médias. Este fato ocorre em virtude do baixo desempenho apresentado notadamente por Malásia e Indonésia.

Posto isto, ressalta-se que os microdados utilizados são divididos em: *i*) variáveis relacionadas aos estudantes; *ii*) variáveis relacionadas aos pais (características dos indivíduos e de sua família) *iii*) variáveis relacionadas as escolas (características específicas). A adoção destes possibilita reunir informações a respeito das variáveis de circunstâncias, esforço e resultados individuais. Especificamente, as variáveis utilizadas neste estudo são sumarizadas no quadro 2.1.

Entretanto, cabe ressaltar que foram realizados quatro testes de especificação do modelo com a finalidade de elencar o melhor grupo de covariadas.⁶ Os testes realizados foram: *ovtest*, *stepwise backward*, *forward* e *hierarchical*. O primeiro teste objetiva detectar problemas de variáveis omitidas. Os outros três testes servem para identificar a inclusão ou remoção de variáveis no modelo.⁷ Além da observância dos resultados dos testes, procurou-se optar também por variáveis clássicas utilizadas na literatura.⁸ Guiado por isto, na primeira abordagem regrediu-se o log da média individual no exame PISA em função do conjunto de variáveis circunstâncias descritas a seguir. Na segunda abordagem, acrescentam-se variáveis que expressam o esforço individual.

Conforme reportado anteriormente, a primeira abordagem utiliza somente o vetor de variáveis circunstanciais e correspondem às informações de 336.286 estudantes de um total

⁶ Destaca-se que embora o *ovtest* tenha indicado a existência de variáveis omitidas – por exemplo, habilidade e ambiente compartilhado – ele também indica o modelo log linearizado como melhor.

⁷ Maiores detalhes em Chatterjee e Hadi (2013).

⁸ Por exemplo, Firmo e Soares (2008); Bauer e Riphahn (2007).

422.413 de 58 países. Por outro lado, na segunda etapa, o vetor de variáveis que expressam o esforço é incluído nas estimações. Nessa fase, são utilizados dados de 225.629 estudantes de 58 países. A razão pela qual os dados utilizados não correspondem ao total disponível, como também, a não inclusão de todos os países devem-se em função de informações faltantes. Adicionalmente, ressalta-se a existência de uma consensual limitação na natureza dos dados empregados, dado que, em avaliações que não há consequências diretas para alunos, professores ou escolas, é possível que a habilidade dos alunos esteja sendo subestimadas.⁹

Quadro 2.1 – Descrição das Variáveis

Variáveis	Descrição
<i>Resultados individuais</i>	Log da Média aritmética dos pontos obtidos nas provas do PISA 2012 que contemplam três áreas de conhecimento: Linguagens; Matemática e Ciências.
Variáveis de Circunstâncias	
<i>Educação Parental</i>	Log do maior grau de instrução parental – anos de estudo – do pai ou da mãe.
<i>Diferença Educacional</i>	Diferença em anos de estudo do pai e mãe.
<i>Tipo de Escola</i>	Uma dummy para o tipo de escola, que assume valor 0 para escola pública e 1 para privada.
<i>Sexo</i>	Uma dummy para sexo, que assume valor 0 para mulheres e 1 para os homens.
<i>Atendimento Pré-escolar</i>	Uma dummy para atendimento pré-escolar, que assume valor 0 para não atendidos e 1 para atendidos.
<i>Localização da Escola</i>	Uma dummy para localização da escola, que assume valor 0 para zona rural e 1 para zona urbana.
<i>Estrutura Familiar</i>	Uma dummy para estrutura familiar, que assume valor 0 para famílias monoparental e 1 para biparental.
<i>Siblings - Irmãos</i>	Uma dummy para irmãos, que assume valor zero caso o indivíduo tenha irmãos morando em casa e 1 caso contrário.
Variáveis de Esforço Individual	
<i>Repetência de Ano</i>	Uma dummy para repetência, que assume 0 caso o aluno já tenha repetido o ano e 1 caso nunca tenha repetido.
<i>Imigrante</i>	Uma dummy para migrante, que assume 0 caso seja migrante e 1 para nativo.
<i>Perseverança</i>	Índice construído com base nas respostas dos alunos sobre a sua vontade de trabalhar em problemas que são difíceis, mesmo quando se deparam com problemas.
<i>Esforço Real</i>	Índice de esforço de auto-relato.
<i>Esforço Potencial</i>	Taxa de não resposta do teste

Fonte: Elaboração própria, a partir de relatórios do PISA 2012.

Dito isto, faz-se agora um breve relato a respeito da amostra do PISA. A tabela 2.2 sumariza algumas características da base. Verifica-se que nos dados analisados, praticamente 50% dos que prestaram o exame em 2012 são do sexo feminino; 80% dos estudantes são oriundos de escola pública; aproximadamente 87% dos candidatos frequentaram no mínimo um ano de atendimento pré-escolar; 61,5% pertencem a países membros da OCDE, 8,9% a

⁹ Por exemplo, Quintano, Castellano e Longobardi (2009).

países da América do Sul. Além do mais, constata-se que em torno de 37% dos pais possui no mínimo educação superior. E somente, algo em torno, de 4% dos estudantes que prestaram o exame PISA em 2012 são filhos de indivíduos que no máximo concluíram o primeiro ciclo do ensino fundamental. Por sua vez, verifica-se que há uma ligeira superioridade educacional das mães, em especial, ao extrato superior da amostra.

Tabela 2.2 – Análise Descritiva

Estudantes por Variáveis de Não Responsabilidade				
Sexo		Membros – Excetos outros		
Masculino	Feminino	OECD	América do Sul	Asiáticos
49,42%	50,48%	61,50	8,90%	9,12%
Nível Educacional Pai - Mãe				
Primária	Básica	Secundária	Superior	
3,42%	21,28%	32,66%	35,94%	
3,89%	21,42%	33,48%	37,41%	
Tipo de Escola		Atendimento Pré-escolar		
Pública	Privada	Atendido	Não Atendido	
80,58%	19,42%	86,90%	13,10%	

Fonte: Elaboração própria, a partir de dados do PISA. % diferentes de 100% (missing).

2.4 PRINCIPAIS RESULTADOS

Nesta seção apresentam-se os principais resultados encontrados. A tabela 2.3 reporta as estimações da equação (2.3)¹⁰ para a amostra geral. Como se esperava os resultados sugere elevado papel das circunstâncias no resultado individual. Onde o conjunto de oportunidades indica responder por algo em torno de 35% de todas as disparidades existentes no exame.

Nessa mesma linha de raciocínio, percebe-se que, as características que mais influenciaram o resultado nos testes scores do PISA 2012, foram respectivamente, educação parental, atendimento pré-escolar, tipo de escola, localização da escola, diferença educacional dos pais e sexo.¹¹

Em síntese, como encontrado em outros trabalhos, por exemplo, Ferreira e Gignoux (2011), o background social – circunstâncias – impacta significativamente os resultados dos indivíduos. Outro aspecto interessante pode ser observado quanto ao efeito negativo da

¹⁰ Estimou-se (2.2) por OLS, Regressão Quantílica e Quantílica tratando a heterocedastidade.

¹¹ Os resultados se mantiveram, para média em ciência, línguas e matemática. Verificar tabelas no apêndice.

diferença educacional parental no teste. A influência negativa desta variável informa que quanto maior a assimetria – disparidade – educacional entre pai e mãe, menor a nota. Esse resultado é coerente com a literatura, por exemplo, ao encontrado em Bourguignon, Ferreira e Menendez (2007), embora, isso ocorra de forma bastante incipiente. Além do mais, acrescenta-se que em quantis superiores, a variável sexo, por vezes não seja estatisticamente diferente de zero.

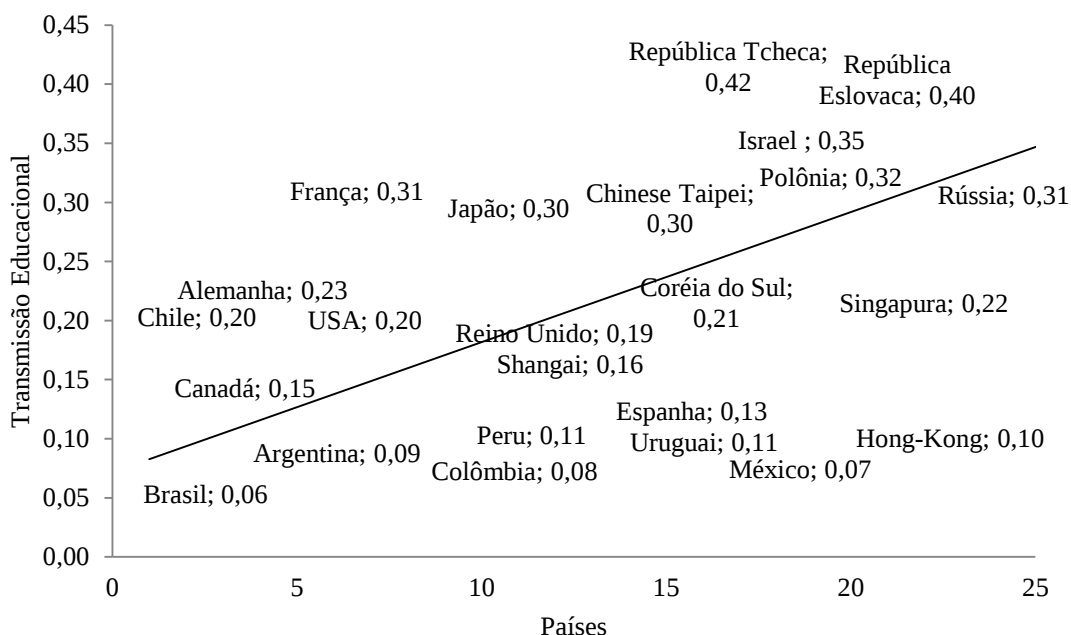
Tabela 2.3 – Determinantes do Resultado Individual – Média Geral – Circunstâncial

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1449*** (0.0010)	0.1610*** (0.0013)	0.1610*** (0.0015)
Dif. Educação	-0.0024*** (0.0002)	-0.0021*** (0.0003)	-0.0021*** (0.0003)
Tipo Escola	0.0468*** (0.0008)	0.0477*** (0.0011)	0.0477*** (0.0010)
Sexo	-0.0172*** (0.0006)	-0.0121*** (0.0008)	-0.0121*** (0.0008)
Pré-escolar	0.1150*** (0.0010)	0.1181*** (0.0013)	0.1181*** (0.0014)
Local Escola	0.0440*** (0.0012)	0.0416*** (0.0015)	0.0416*** (0.0014)
Estrutura Familiar	0.0543*** (0.0009)	0.0551*** (0.0012)	0.0551*** (0.0009)
Irmãos	0.0226*** (0.0009)	0.0261*** (0.0012)	0.0261*** (0.0011)
Intercepto	5.6000*** (0.0028)	5.5704*** (0.0037)	5.5704*** (0.0040)
N	336286	336286	336286
adj. R ²	0.138		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * p<0.10, ** p<0.05, *** p<0.01.

Adicionalmente, verifica-se que a influência da educação parental nos resultados do teste é mais incisiva em países desenvolvidos. Em síntese, a transmissão educacional entre gerações é claramente mais forte nesses países. Os resultados reproduzidos no gráfico 2.1 destacam tal situação. Para ter-se uma ideia, Argentina e Brasil juntos assinalam uma influência educacional – transmissão – parental abaixo da metade do efeito observado em relação a japoneses, franceses e chineses e quase 3 vezes inferiores aos tchecos e eslovacos. Especificamente, Argentina e Brasil ocupam respectivamente a 51^a e 56^a posição no ranking de transmissão educacional, num total de 58 países analisados. De acordo com as estimações gerais, o Brasil somente obtém resultados ligeiramente superiores a Macao e Qatar.

Gráfico 2.1 – Influência Educacional Parental no Teste PISA – Média Geral

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Esses resultados poderiam sugerir, por exemplo, na teoria de mobilidade educacional, que Brasil e Argentina têm baixa persistência educacional, e, portanto, alta mobilidade? E que República Tcheca, Eslováquia, Japão e França possuem alta persistência, e, portanto, baixa mobilidade? A resposta é não, pois, neste estudo faz-se uso de um sistema de avaliação educacional – PISA 2012 – e não renda ou realização educacional como a maior parte dos estudos voltados para esta área. Ratificando tal justificativa, há diversos trabalhos sobre mobilidade que apontam baixa mobilidade em países em desenvolvimento. Visto que, geralmente nesses países, os indivíduos que cursam ou tem nível superior são filhos de pais que possuem no mínimo nível superior.¹² Já nos países desenvolvidos essa relação não é tão predominante. Pois, há mais oportunidades e melhor qualidade educacional.

Corroborando a direção desses resultados, de acordo com Black e Devereux (2010), conceitualmente, pode-se admitir que a escolha educacional dos jovens é fruto de alguns fatores, em especial, custos educacionais, retornos da educação e renda familiar. Essa última, em especial, nos casos onde há restrições ao crédito. Por outro lado, parece haver consenso que os retornos da educação são maiores para indivíduos mais habilidosos e também para aqueles que têm pais mais educados. Estas hipóteses implica que indivíduos que tem pais mais educados tendem a buscar maior nível educacional devido a dois efeitos: i) efeitos diretos – pais mais educados – também interpretados como canal causal, e, ii) efeitos

¹² Por exemplo, Daude (2011); Hertz *et al.* (2007).

indiretos, ou seja, possuir maior habilidade – herdabilidade – evidenciada pela transmissão intergeracional da educação.

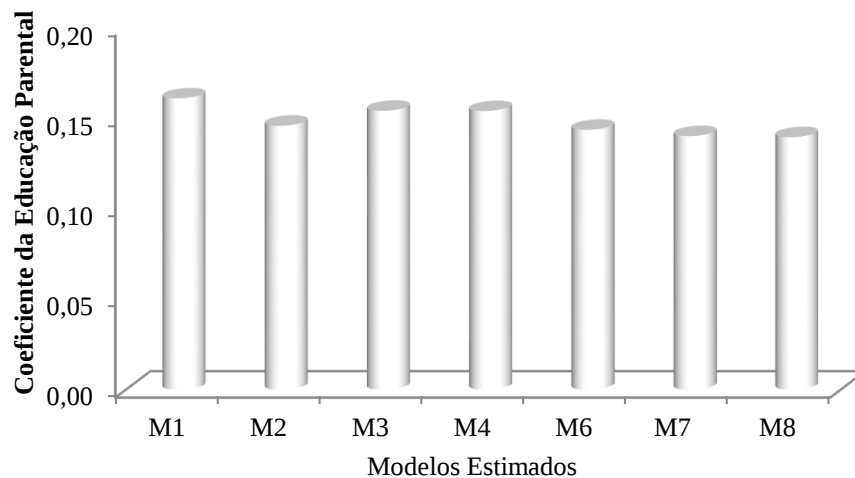
Além disso, segundo os autores, existe a possibilidade de mecanismos subjacentes transmitirem efeitos diretos da educação parental no desempenho dos indivíduos. Pois, quanto maior a educação parental, em geral, maior a renda familiar, e, conseqüentemente, pode impactar positivamente o desempenho dos indivíduos. Segundo, esta característica pode incrementar o tempo dedicado a desenvolver atividades de reforço com os seus filhos. Além disso, possibilita também, elevar o poder de barganha das famílias. Dado que, mães mais educadas podem ser mais susceptíveis a direcionar gastos em investimentos e atividades voltadas para os jovens ou crianças da família.

Outro importante fato a se verificar, seria a presença de mobilidade qualificada intergeracional da educação. Em suma, países desenvolvidos, comumente, tem um nível educacional mais elevado, e, portanto, origina filhos mais educados. Designadamente, no Brasil e em países em desenvolvimento, isso pode não valer. Ratificando tal hipótese, em média, a educação parental – anos de estudo dos pais – nesses países é mais baixa conforme reporta a Tabela A.4 no Apêndice.

Outro padrão interessante a ser observado, olhando para tabela A.4, é o valor encontrado em variáveis como: tipo de escola, atendimento pré-escolar e localização da escola. Esses fatores desempenham um papel crucial nos países em desenvolvimento, sugerindo haver heterogeneidade na qualidade da educação. Diferentemente do que ocorre na maior parte dos países desenvolvidos, dado que, essas variáveis não exercem influência demasiada no desempenho educacional. Nomeadamente, no Brasil e Argentina, o simples fato de se estudar em escola privada afeta em média a nota do PISA em 17% e 14% respectivamente. No entanto, em países como Alemanha e Japão tais efeitos são insipientes.

Neste rol de investigação, acrescentaram-se as estimações algumas variáveis que expressam o esforço individual. Embora, a princípio ainda assume-se a hipótese postulada na equação (2.4). Entretanto, antes de discutirmos os resultados reproduzidos na tabela 2.4, o gráfico 2.2, mostra oito especificações distintas que consideram *dummys* continentais. Claramente, os resultados estimados são bem semelhantes.¹³ Todavia, os testes de especificação informam que o modelo geral é mais ajustado sem as *dummys* continentais.

¹³ M1 não há dummy para continentes; M2 uma dummy para OCDE; M3 é igual a M2 mais uma dummy (países asiáticos); M4 é M3 acrescido da dummy(países nórdicos); M5 modelo geral incluindo o vetor de esforços individuais. Os demais modelos acrescentam-se uma a uma as *dummys* continentais.

Gráfico 2.2 – Influência Educacional Parental

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Em outras palavras, os resultados apresentados na tabela 2.4, não diferem significativamente dos resultados reportados na tabela 2.3.¹⁴ Nitidamente, os coeficientes estimados – efeitos das covariadas – apesar de menores, se situam bem próximos, sugerindo assim, o bom ajuste do modelo. Adicionalmente os testes de identificação anteriormente citados ratificam estes resultados.

Contudo, dois fatos merecem destaque entre as covariadas que representam o esforço individual nos resultados. Primeiro, a alta influência exercida pelo fato do aluno nunca ter repetido o ano. Isto é, essa característica influencia em média 11,2% no desempenho do teste. Segundo, o baixo efeito atribuído ao esforço real e esforço potencial, com destaque especial para o sinal desta última. Ou seja, evidenciando uma alta taxa de não resposta. Tal característica aponta para o problema de limitação em dados desta natureza. Em outras palavras, como o teste teoricamente não exerce nem um reflexo na vida de alunos, professores e escolas, há uma elevada taxa de não respostas, em especial, em questões mais complexas. Nessa mesma linha de raciocínio, percebe-se que o coeficiente da variável perseverança parece confirmar essas evidências.

Em relação às variáveis alusivas à escola, como tipo, localização e pré-escola elas sugerem responder por aproximadamente 20% das diferenças entre os desempenhos obtidos. Onde a pré-escola se apresenta como fator mais importante entre estes. Esta última característica é ainda mais expressiva quando consideramos os resultados dos países asiáticos. Isto é, em média nessas economias a influência da pré-escola é quase quatro vezes maior do

¹⁴ Os resultados se mantiveram, para média em ciência, línguas e matemática. Ver tabelas no apêndice.

que o observado nos países da OCDE e países sul-americanos. E aproximadamente três vezes mais do que nos países nórdicos. Vale ressaltar que, embora os resultados gerais referentes ao tipo de escola – pública ou privada – informam uma menor influência desta no resultado do exame. Quando avaliamos os resultados dos países da América do Sul, verifica-se que a influência chega a ser cinco vezes mais efetiva do que nos países membros da OCDE. Aproximadamente sete vezes ao efeito atribuído aos asiáticos. E infinitamente superior aos países nórdicos.

Tabela 2.4 – Determinantes do Resultado Individual – Média Geral

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1361*** (0.0012)	0.1489*** (0.0015)	0.1489*** (0.0012)
Dif. Educação	-0.0012*** (0.0002)	-0.0012*** (0.0003)	-0.0012*** (0.0004)
Tipo de Escola	0.0503*** (0.0009)	0.0515*** (0.0012)	0.0515*** (0.0010)
Sexo	-0.0056*** (0.0007)	-0.0019** (0.0009)	-0.0019*** (0.0007)
Pré-escolar	0.0996*** (0.0012)	0.1029*** (0.0016)	0.1029*** (0.0013)
Local Escolar	0.0430*** (0.0013)	0.0391*** (0.0017)	0.0391*** (0.0020)
Migrante	0.0204*** (0.0011)	0.0175*** (0.0015)	0.0175*** (0.0015)
Repetente	0.1124*** (0.0010)	0.1192*** (0.0013)	0.1192*** (0.0011)
Perseverança	0.0160*** (0.0004)	0.0145*** (0.0005)	0.0145*** (0.0004)
Esforço Real	0.0038*** (0.0002)	0.0047*** (0.0003)	0.0047*** (0.0003)
Esforço Potencial	-0.0009*** (0.0001)	-0.0010*** (0.0001)	-0.0010*** (0.0001)
Intercepto	5.5540*** (0.0038)	5.5228*** (0.0049)	5.5228*** (0.0039)
<i>N</i>	225629	225629	225629
adj. <i>R</i> ²	0.190		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Não obstante, há evidências empíricas que apontam haver possível endogeneidade entre as covariadas – variáveis de circunstâncias e de esforço – e, portanto, parece haver vieses nos coeficientes estimados. Adicionalmente, destaca-se que apesar da inclusão inédita de variáveis representativas do esforço individual, certamente, não se dispõem de todas as variáveis – omissão – que afetam o resultado. Nesse sentido, buscando averiguar e corrigir

estes detalhes emprega-se o método baseado em Bourguignon, Ferreira e Menéndez (2007). Em suma, admite-se a existência de dependência entre as variáveis de circunstância e de esforço. Isto é, quebra-se a suposição assumida em (2.4).

Diante desse cenário, primeiramente, calcula-se o índice de desigualdade para distribuição fatual da média geral dos resultados. Logo em seguida, estimam-se os coeficientes da forma reduzida – equação 2.10 – para os respectivos quantis (0.25; 0.50 e 0.75). Por sua vez, utilizando os coeficientes médios estimados em (2.10), como também os extremos inferiores e superiores, simulou-se a distribuição contrafactual $M(\tilde{w})$ proveniente da expressão: $\tilde{w} = \exp[(\beta_1 + \gamma\beta_2)\overline{W}_{p,l}^* + v_i\beta_2 + u_{f,i}] \rightarrow \tilde{w} = \exp[\beta_i\overline{W}_{p,l}^* + \varepsilon_i]$, onde $\beta_1 + \gamma\beta_2 = \beta_i$ e $\varepsilon_i = v_i\beta_2 + u_{f,i}$. Em síntese, calcula-se o índice de desigualdade para distribuição fatual w , como também para distribuição contrafactual $\tilde{w}$. Logo após, seguindo a equação (2.7) obtém-se a participação total do conjunto de oportunidades na desigualdade da nota. Destaca-se que o índice de desigualdade utilizado nesta pesquisa foi o desvio padrão dos logs. Encontrou-se resultados semelhantes ao utilizar o coeficiente de variação. No entanto, há outros índices bastante utilizados na literatura, por exemplo, Ferreira e Gignoux (2011); Lefranc, Pistolesi e Trannoy (2009).

De forma semelhante, obtém-se o efeito parcial ou direto da participação das oportunidades observadas na desigualdade total. Primeiro, estima-se (2.5) diretamente e utilizando os coeficientes estimados constrói a distribuição contrafactual $M(w_i^d)$ originada por: $w_i^d = \exp[\beta_4\overline{W}_{p,l}^* + E_i + u_i]$. Percebe-se que, no primeiro contrafactual $M(\tilde{w})$ anula-se completamente o efeito das circunstâncias, e no segundo, $M(w_i^d)$ remove-se somente o efeito parcial ou direto. Os resultados fatuais, contrafatuais e as decomposições propostas em (2.7) e (2.8) são reportados na tabela 2.5.

Os mesmos sugerem que a desigualdade no desempenho PISA 2012 se reduz em torno de 21% quando se iguala o conjunto de oportunidades individuais. Por sua vez, a participação do conjunto de oportunidades – expresso por oito variáveis – no desempenho do teste varia de forma estável entre os quantis. Em outras palavras, a desigualdade de oportunidades parece não alterar o padrão entre os quantis da distribuição observada.

Quanto aos efeitos diretos, situa-se em torno de 18%. Ou seja, ao anular apenas os efeitos diretos das circunstâncias, percebe-se que a desigualdade total da distribuição das notas se reduz em média 18,7%. Esses resultados indicam um importante aspecto, dado que, o efeito indireto das circunstâncias é dado por: $\aleph_I^i = \aleph_I - \aleph_I^d$. Isto é, há um efeito positivo das circunstâncias favoráveis sobre o esforço individual em torno de 2,3% a 2,8%. Diante desses

resultados, ao menos duas inferências parecem viáveis: i) há efeitos das circunstâncias sobre o esforço individual despendido; e ii) a magnitude desse efeito situa-se em torno de 15% do total da desigualdade injusta observada no exame.

Tabela 2.5 – Decomposição da Participação do Conjunto de Oportunidades

Desigualdade Total Observada - Desvio Padrão do log da média geral	0,21457		
Quantil	0.25	0.50	0.75
Painel A: Efeito Total das Circunstâncias			
Limite Superior	0,16991	0,16978	0,16996
Valor Médio	0,16991	0,16978	0,16996
Limite Inferior	0,16991	0,16978	0,16996
% Total das Circunstâncias na Desigualdade do Teste			
Limite Superior	0,20811	0,20874	0,20789
Participação Média ((1)-(2A)/1)	0,20811	0,20874	0,20789
Limite Inferior	0,20811	0,20874	0,20789
Painel B: Efeito Direto das Circunstâncias			
Limite Superior	0,17584	0,17571	0,17595
Valor Médio	0,17531	0,17526	0,17552
Limite Inferior	0,17483	0,17484	0,17513
% Direto das Circunstâncias na Desigualdade do Teste			
Limite Superior	0,18048	0,18108	0,17999
Participação Média ((1)-(2B)/1)	0,18294	0,18319	0,18196
Limite Inferior	0,18519	0,18515	0,18380
Painel C: Efeito Indireto Circunstância			
Limite Superior	0,02763	0,02766	0,02790
Participação Média ((1)-(2B)/1)	0,02517	0,02555	0,02593
Limite Inferior	0,02292	0,02359	0,02409
Painel D: Tratando o Esforço Observado como Circunstância			
Limite Superior	0,16991	0,16978	0,16996
Valor Médio	0,16991	0,16978	0,16996
Limite Inferior	0,16991	0,16978	0,16996
% Circunstâncias e Esforço Observados na Desigualdade do Teste			
Limite Superior	0,20811	0,20874	0,20789
Participação Média ((1)-(2C)/1)	0,20811	0,20874	0,20789
Limite Inferior	0,20811	0,20874	0,20789

Fonte: Elaboração própria, a partir dos dados do PISA 2012. Significativos a 95% .

Em outras palavras, o esforço individual exercido pelos tipos – diferentes conjuntos de oportunidades – é influenciado pelas circunstâncias as quais o indivíduo está inserido. Outro aspecto interessante, é que os efeitos observados são maiores no quantil inferior. Isto é, quanto pior o desempenho do aluno no teste, maior a influência exercida pelas circunstâncias sobre o esforço empregado.

Contudo, é possível que os efeitos das circunstâncias nas variáveis de esforço – repetência, perseverança, migrante, esforço real e potencial – sejam maiores. Dado que, é provável que as circunstâncias não observadas sejam responsáveis por uma parcela maior da

variância nos resíduos aleatórios v_i referente à equação (2.6). Além disso, é aceitável supor que a decisão de algumas variáveis caracterizadas aqui como esforço seja uma decisão dos pais. Ainda mais, considerando que os jovens que prestaram o *teste score* têm apenas 15 anos de idade. Em resumo, suponha que a decisão de migrar e repetir o ano sejam definidos pelos pais, e que as três outras variáveis sofram forte influência instrutiva dos pais. Ou seja, estas hipóteses – algumas vezes consideradas extremas – correspondem a considerar todas nossas variáveis como circunstâncias observáveis.

Posto isso, constrói-se um contrafactual $\tilde{w} = \exp[\beta_i \overline{W_{p,i}^*} + \bar{E}_i \beta_i + u_i]$. Os resultados reportados no painel (D) da tabela 2.5 indicam uma participação total de 21% no total da desigualdade observada na nota. Isto é, caso fossem equalizadas todas as variáveis consideradas – circunstâncias e esforço – a desigualdade total seria em torno de 16,9%. Perceba que os valores encontrados são exatamente iguais à participação total das circunstâncias no resultado do teste.

Em outro exercício, apresentado na tabela 2.6, buscou-se verificar qual o percentual da parcela das variáveis de circunstâncias – isoladamente – no total da desigualdade. Para tanto, simulou-se o seguinte contrafactual da forma reduzida: $w_i^j = \exp[\beta_i \overline{W_{p,i,j}} + W_{p,i,-j} \beta_{-i} + u_i]$, sendo $M(w_i^j)$ o contrafactual da distribuição condicionada das notas alcançado ao se manter uma das variáveis que representam as circunstâncias constante, enquanto as demais variam.

Tabela 2.6 – Equalização das Oportunidades

Equalização	Quantil		
	0.25	0.50	0.75
Desigualdade Total Observada	0.21457		
Educação Parental	0.17468	0.17453	0.17464
Estrutura Familiar	0.18067	0.18071	0.18090
Tipo de Escola	0.17932	0.17950	0.17994
Localização Escola	0.18035	0.18124	0.17786
Atendimento Pré-escolar	0.17748	0.17756	0.18062
Sexo	0.18125	0.18043	0.18149

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Os resultados convergem aos encontrados na literatura, por exemplo, ver Ferreira e Veloso (2006). Perceba que educação parental, tipo de escola e pré-escola são os três principais determinantes de desigualdade injusta.¹⁵ Especificamente, se equalizássemos a educação parental, tipo de escola e pré-escola à desigualdade total observada no desempenho

¹⁵ Ver, por exemplo, Lefranc, Pistolese e Tranoy, (2009).

do teste se reduziria em torno de 17% a 19%. Por outro lado, observa-se que a circunstância menos incisiva na desigualdade é a variável sexo. Inclusive, sendo insignificante estatisticamente em vários cenários. Em resumo, o sexo não é determinante na desigualdade do desempenho do exame.

Adicionalmente, a tabela 2.7 sintetiza os resultados da decomposição da influência das circunstâncias em termos de países e continentes. Notem que, os efeitos indiretos sobre o esforço individual são mais significativos nos países asiáticos e nórdicos respectivamente e menos contundentes nos membros da OCDE e América do Sul. Embora, a magnitude da desigualdade no desempenho do teste seja maior neste último. Por sua vez, é necessário compreender que fatores podem compor esse efeito indireto. Poder-se-ia supor que aspectos como habilidade e motivação através da transmissão intergeracional parental poderiam ser um fator decisivo. Ou seja, dado que esses países historicamente possui uma educação de melhor qualidade isso explicaria tal diferença? De acordo com a literatura sim. Pois, sabe-se que há três tipos de fatores que podem determinar a transmissão Intergeracional educacional: a genética, o comportamento dos pais e os fatores ambientais.¹⁶ Apesar disso, os resultados são inconclusivos, pois, embora a França apresente um efeito indireto de 0,17; a OCDE e alguns dos seus principais membros – Japão – apresentam valores menores aos encontrados nos países sul-americanos.

Posto isso, acredita-se que o efeito indireto das circunstâncias nos países da América do Sul é menos incisivos em razão de diversos fatores, por exemplo, a qualidade da educação parental e das escolas desta região. Uma vez que, essa característica supera os efeitos da educação parental, exceto para o Chile. Para ter-se uma ideia, no Brasil e Argentina, o efeito atribuído à escola no resultado do teste é respectivamente 30% a 60% superior a influência atribuída à educação. Estes resultados apontam para uma possível solução do problema, isto é, nitidamente precisa-se melhorar a qualidade das escolas desse continente, em especial, as públicas, uma vez que, 73,5% dos jovens desses países que participaram do exame PISA 2012 são oriundos de unidades de ensino públicas.

Além do mais, esse problema se agrava quando se leva em conta o atendimento pré-escolar e localização da escola, conforme pode ser verificado na tabela A.4. Em linhas gerais, os efeitos indiretos são mais expressivos em países desenvolvidos, com destaque para França e Finlândia. Por outro lado, onde não se observa efeitos indiretos tão expressivos, como Japão e Coreia do Sul, fatores como tipo de escola, localização da escola e atendimento pré-escolar

¹⁶ Ver, por exemplo, Björklund, Lindahl e Plug (2004).

não são estatisticamente significantes. Ou seja, tais fatores são homogêneos, e, portanto, não exercem influência verificável indireta sobre o nível de esforço.

Tabela 2.7 – Decomposição do Efeito das Circunstâncias por Países no PISA

Continentes Países	Desigualdade PISA	$\aleph_i$	$\aleph_i^d$	% Efeito Total	% Efeito Direto	$\aleph_i^i$
América do Sul	0,20151	0,14126	0,15244	0,29897	0,24350	0,05547
Argentina	0,22500	0,13650	0,14835	0,39331	0,34066	0,05265
Brasil	0,18623	0,13574	0,14749	0,27111	0,20804	0,06307
Chile	0,17562	0,12993	0,14094	0,26013	0,19746	0,06267
Países da OCDE	0,19225	0,15734	0,16316	0,18158	0,15131	0,03026
Alemanha	0,18367	0,13295	0,14427	0,27615	0,21453	0,06162
Estados Unidos	0,17963	0,14907	0,15987	0,17012	0,10997	0,06016
França	0,20298	0,12776	0,16310	0,37060	0,19648	0,17411
Japão	0,18142	0,15335	0,15653	0,15471	0,13720	0,01752
Países Asiáticos	0,19408	0,14794	0,16909	0,23770	0,12876	0,10894
Chinês Taipei	0,16114	0,13883	0,14718	0,13846	0,08660	0,05186
Hong Kong	0,18598	0,14631	0,15764	0,21326	0,15238	0,06089
Shanghai	0,14753	0,12284	0,12943	0,16735	0,12267	0,04468
Singapura	0,18949	0,15955	0,16894	0,15803	0,10846	0,04956
Países Nórdicos	0,19411	0,14483	0,16438	0,25384	0,15316	0,10069
Dinamarca	0,18417	0,13493	0,15264	0,26739	0,17122	0,09617
Finlândia	0,18858	0,13187	0,15399	0,30071	0,18344	0,11727
Suécia	0,20620	0,15056	0,16629	0,26982	0,19353	0,07629

Fonte: Elaboração própria, a partir dos dados do PISA 2012. Significativos a 95%.

Sobretudo, seguindo os objetivos traçados nesta pesquisa, agora verificaremos os efeitos de um incremento no conjunto socioeconômico individual através do método IDC. Primeiramente, abordaremos os resultados da estimação do log da média geral das notas individual em função de uma dummy para a educação do pai e da mãe, como também para a estrutura familiar, tipo de escola, atendimento pré-escolar e localização da escola.

O Efeito de Tratamento Médio Quantílico – ETMQ – foi apurado para todas as variáveis de circunstâncias – exceto sexo e irmãos – denotadas na tabela 2.8. Os resultados foram estatisticamente significantes a 1%. Todavia, o gap médio observado para educação parental é em torno de 8%. Onde cerca de 2% é explicado por diferenças no nível educacional parental da distribuição. E 6% são devido às diferenças nos coeficientes médios entre filhos de pais com no mínimo nível superior e filhos de pais com no máximo nível médio. Quanto ao tipo de escola – pública ou privada – a diferença média observada no resultado do teste situa-se em 6%, onde 1,5% deste diferencial advêm de se estudar em escola privada. Destaca-se que a parte explicada pelos resíduos é negligenciável para todos os tratamentos propostos.

Tabela 2.8 – Efeito Tratamento Médio Quantílico – Circunstâncias – ETMQ

Desempenho Individual – log da média geral – Quantílico									
ETMQ***	Educação da Mãe			Educação do Pai			Estrutura Familiar		
	0.10	0.50	0.90	0.10	0.50	0.90	0.10	0.50	0.90
	0.0251	0.0202	0.0128	0.0247	0.0196	0.0120	0.0240	0.0144	0.0068
	0.0651	0.0633	0.0614	0.0592	0.0575	0.0566	0.0367	0.0360	0.0362
	0.0807	0.0857	0.0642	0.0709	0.0797	0.0626	0.0603	0.0505	0.0307
	Tipo de Escola			Pré-escola			Localização da Escola		
	0.0197	0.0139	0.0091	0.0282	0.0290	0.0233	0.0314	0.0271	0.0170
	0.0578	0.0495	0.0451	0.1040	0.1069	0.1103	0.0467	0.0397	0.0365
	0.0842	0.0654	0.0324	0.1383	0.1381	0.0996	0.0832	0.0661	0.0568

Fonte: Elaboração própria. Os tratamentos utilizados foram *** $p < 0.01$. Os valores em grupo de três para cada quantil representam respectivamente: efeito da característica; coeficientes e tratamento médio.

Por sua vez, o tratamento médio referente à pré-escola ocorre próximo a 14%. Entretanto, apenas 21% desse efeito é função do aluno ter frequentado a pré-escola. Contudo, na base analisada, não há informações sobre a idade de ingresso na pré-escola, o que limita essa análise. Pois, segundo Spinath *et al.* (2003), há uma importância fundamental da primeira infância – 0 a 6 anos – no desenvolvimento da habilidade cognitiva geral.¹⁷ No que se referem as demais variáveis – estrutura familiar e localização das escolas – respectivamente, observou-se um ETMQ variando em torno de 5% e 6%. Onde o efeito de ter uma família monoparental e estudar em escola localizada na zona urbana afeta o desempenho no teste em torno de 1,5% a 2% nessa ordem.

Porém, a fim de compreender como se comporta o padrão desses efeitos entre os países, estimou-se o mesmo tratamento isoladamente para cada país. Os resultados reportados na tabela A.9 do apêndice, indicam haver uma grande variação nos efeitos tratamento propostos entre os países. Por exemplo, referente à variável educação parental, nota-se que, embora, haja um diferencial médio semelhante entre os continentes que varia em torno 9% a 12%, o efeito tratamento de possuir pais mais educados na América do Sul é 7,5 vezes superior ao observado nos países da OCDE. E respectivamente 3 a 4,5 vezes maiores ao encontrado nos países asiáticos e nórdicos.

De forma diferente, observa-se que nos países sul-americanos, o diferencial médio entre os grupos de tratamento e controle – escola privada e pública – é substancialmente maior que nos países asiáticos. E respectivamente, 2 a 3 vezes mais do que o observado nos membros da OCDE e países nórdicos. Especificamente, esse efeito no Brasil é quase 7 vezes maior do que na França e 20 vezes superior ao apresentado nos Estados Unidos. Contudo, somente cerca de um terço dessa influência é explicado pela origem escolar do aluno. Além

¹⁷ Segundo Carroll (1997), existe um fator geral de inteligência potencial que é responsável pela facilidade em lidar com informações e problemas de uma determinada classe ou conteúdo.

disso, tais efeitos são bem semelhantes aos encontrados em alguns países da OCDE, asiáticos e nórdicos. Esses resultados parecem sugerir novamente baixa qualidade educacional.

Por outro lado, o efeito tratamento da pré-escola é bem mais expressivo nos países que obtiveram os melhores desempenhos no exame PISA. Por exemplo, na França e no Japão a influência dessa característica situa-se em 10%, contra 4% do que se observa na Argentina e Brasil. Em linhas gerais, todos os resultados, apontam que o baixo desempenho obtido, em especial, dos países da América do Sul, em relação aos demais é determinado crucialmente pelo conjunto de oportunidades. Pois, olhando de forma grosseira para a tabela A.4, no apêndice, percebe-se claramente que a desigualdade observada explicada pelo conjunto de características individuais – soma dos efeitos individuais – é de aproximadamente 0.35 nesta região, contra 0.11 na OCDE, 0.15 nos países asiáticos e 0.13 nos países nórdicos. Isto é, correspondem a 91% do efeito total verificado nos outros três continentes juntos.

Nesse sentido, uma questão merece destaque, pois, dado que, o efeito tratamento é bem mais significativo nos países sul-americanos, porque a elasticidade – transmissão – intergeracional estimada para estes é menor? Os resultados indicam que em variáveis de desempenho e não de resultado educacional, como anos de estudo, a qualidade da educação é captada. Isto é, não basta apenas ter pais com realização educacional – nível superior – mas sim, ter pais educados efetivamente transferindo conhecimentos e habilidades para seus filhos. Além disso, o papel da escola nesses países parece ser fundamental de acordo com os números apresentados. Ou seja, não basta continuar-se realizando educacionalmente os indivíduos, e, sim, capacitando-os para desenvolver e aplicar conhecimentos.

2.5 ALGUMAS CONSIDERAÇÕES

Nesse trabalho buscou-se compreender os mecanismos subjacentes que estão por trás do desempenho educacional. Elencaram-se três objetivos principais. Primeiro, verificar o grau de transmissão e correlação intergeracional da educação. Segundo, decompor os efeitos dos determinantes da transmissão intergeracional – conjunto de oportunidades – em diretos e indiretos. Terceiro, construir uma análise contrafactual originada pela equalização de circunstâncias socioeconômicas dos estudantes que prestaram o exame do PISA 2012. Podem-se destacar ao menos três importantes deduções possibilitadas pelos resultados: i) percebe-se uma baixa transmissão da educação parental, nos países em desenvolvimento, em

especial, no continente sul-americano; ii) verifica-se haver efeitos indiretos das circunstâncias sobre as variáveis de esforço, e os mesmos variam entre indivíduos, países e continentes. Por fim, corroborando com a baixa transmissão educacional dos pais evidenciada, o efeito tratamento é substancialmente maior nos países com essas características.

Adicionalmente, olhando para os resultados reportados na tabela A.4 e A.9, percebe-se que os fatores associados à escola, exceto a pré-escola, revelam ser incisivos apenas nos países em desenvolvimento, especificamente, nos países sul-americanos. Por outro lado, com exceção do Chile, o efeito tratamento do atendimento pré-escolar nessas economias é geralmente inferior ao apurado em países que obtiveram os melhores desempenhos. Esses resultados sugerem que a pré-escola na realidade importa, no entanto, novamente enfatiza-se a qualidade (ver, por exemplo, Foguel e Veloso, 2012). Pois, ao olhar-se para Chile, país sul-americano com melhor desempenho no teste, percebe-se que é o fator escolar associado ao ambiente socioeconômico individual mais importante. Já para os outros países deste continente – em geral que apresentaram os piores desempenhos – averiguamos que este fator não se apresenta com tal importância. Ou seja, diferentemente da grande maioria onde se observa os melhores resultados no teste à pré-escola nesses países exerce pouca influência. Isso reforça ainda mais a desconfiança sobre a qualidade da educação nesses países.

3 IGUALDADE DE OPORTUNIDADES: ANALISANDO O PAPEL DAS CIRCUNSTÂNCIAS NO DESEMPENHO DO ENEM.

3.1 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). Esta etapa seguirá o referencial teórico de O'Neill, Sweetman e van DeGaer (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 Romer (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 e 2007), que demonstra como os diversos insumos influenciam os resultados educacionais e com isso proporciona uma orientação para a aplicação de políticas públicas alocativas. A magnitude da influencia do esforço/talento individual será mensurada a partir do instrumental desenvolvido por Klein, Vela e Farre (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 à inexistência de variáveis instrumentais apropriadas no banco de dados.

Dito isso, o trabalho está organizado em cinco seções além desta introdução. Na seção 3.2, apresenta-se uma breve revisão de literatura relacionada à desigualdade de oportunidades e desempenho educacional. A seção 3.3 é destinada aos modelos teóricos, métodos de estimação e descrição do banco de dados. Na seção 3.4 avaliam-se os principais resultados e a seção 3.5 conclui o estudo.

3.2 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 frutos 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 circunstância, 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 à 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 é denominada 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. isto é, 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, Ferreira e Menéndez, 2007). A adaptação desta 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 2000¹⁸, 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

¹⁸ Foram analisados 14 países, sendo 9 da OCDE e 5 da América Latina.

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 Waltemberg (2012), analisam a desigualdade de oportunidade no desempenho educacional em seis países da América Latina¹⁹. Utilizando a metodologia desenvolvida por Checchi e Peragine (2010), os autores constatarem que o grau de desigualdade de oportunidade educacional, varia entre 1% a 25%, o que denota uma considerável heterogeneidade entre os seis países. Em Aguirreche (2012), investiga-se como nível de desigualdade de oportunidades de um país 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

¹⁹ Os países analisados foram Argentina, Brasil, Chile, Colômbia, México e Uruguai.

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, observa-se que as circunstâncias observadas representam cerca de um quarto das disparidades observadas (Ver Bourguignon, Ferreira e Menéndez, 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.

3.3 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, Sweetman e van DeGaer (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 e 2007), em conjunção com o instrumental desenvolvido por Klein, Vela e Farre (2010). Esta seção discutirá o passo a passo das duas abordagens, subseções 3.3.1 e 3.3.2, respectivamente. Por fim, na seção 3.3.3 será apresentado o conjunto de dados.

3.3.1 Modelo Bivariado

De acordo com O'Neill, Sweetman e van DeGaer (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:

$$F_z^*(y[e, x]|x) = F_e^*(e|x) \quad (3.1)$$

A equação 3.1 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 3.1 adicionada a suposição IND implica no Axioma de Identificação de Roemer (RIA).
- **RIA** (*Roemer's Identification Axiom*): $F_z^*(y[e, x]|x) = F_e^*(e|x) \Rightarrow e' = e''$. 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, Sweetman e van DeGaer (2001), isto possibilita desenhar o resultado – desempenho no ENEM – com uma função de $\pi \in [0,1]$ para diferentes valores de x . Sendo o conjunto 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 π . Deste modo, o conjunto de oportunidade do indivíduo do tipo x será:

$$S_x = (z, \pi) \in (\mathbb{R}^+ \times [0,1]) \mid z = F_z^{*-1}(\pi|x) \quad (3.2)$$

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, Sweetman e van DeGaer (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).

Seguindo O'Neill, Sweetman e van DeGaer (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:

$$\hat{f}_{(x_0)} = \frac{1}{Nh} \sum_{i=1}^n K\left(\frac{x_i - x_0}{h}\right) \quad (3.3)$$

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, definindo 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:

$$f[z|x] = \frac{f[z|x]}{f_x[x]} \quad (3.4)$$

em que $f_x[x]$ representa a distribuição marginal das circunstâncias e $f[z|x]$ a distribuição conjunta de z e x . Para estimar (3.4), 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:

$$\hat{f}_{xA}[x] = \frac{1}{nh_x} \sum_{i=1}^n \frac{1}{w_i} k\left[\frac{x - x_i}{w_i h_x}\right] \quad (3.5)$$

A distribuição conjunta das circunstâncias e resultados obtidos da equação (3.4) é obtida por:

$$\hat{f}_A[z|x] = \frac{1}{nh_z h_x} \sum_{i=1}^n \frac{1}{w_i^2} k\left[\frac{z - z_i}{w_i h_z}\right] \left[\frac{x - x_i}{w_i h_x}\right] \quad (3.6)$$

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:

$$w_i = \left[\frac{\hat{f}_g}{\hat{f}_k[z_i, x_i]} \right]^{1/2} \quad (3.7)$$

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, Sweetman e van DeGaer (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 (3.5) e (3.6). Adicionalmente, assume-se que a kernel é multiplicativa, conforme Trede (1998), tornando-a em uma simples expressão de fda. Dessa forma, substitui os termos da equação (3.4) pelas estimativas das equações (3.5) e (3.7), obtendo-se a distribuição condicional:

$$\hat{F}_{zA}^*[z|x] = \frac{\sum_{i=1}^n \frac{1}{w_i} K\left[\frac{x - x_i}{w_i h_x}\right] G\left[\frac{z - z_i}{w_i h_z}\right]}{\sum_{i=1}^n \frac{1}{w_i} K\left[\frac{x - x_i}{w_i h_x}\right]} \quad (3.8)$$

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:

$$\hat{S}_{x_i} = \{(z, \pi) \in (\mathbb{R}^+ \times [0,1]) | z = \hat{F}_{zA}^{*-1}[\pi|x]\} \quad (3.9)$$

3.3.2 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 e 2007), a FPE é dada por:

$$A_{it} = g(F_i^t, P_i^t, I_i S_i^t) \quad (3.10)$$

em que A_{it} é o vetor educacional realizado pelo i -ésimo estudante 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, Ferreira e Franco (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 a diferenças na composição sócioeconômica de seus alunos.

Cabe aqui ressaltar um aspecto relevante para a estimação de (3.10) referente a 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, Vella (2009, 2010) tratadas a seguir.

Considere os seguintes modelos simultâneos de transferência no desempenho educacional:

$$E_i^f = \alpha_0 + \beta_0 X_i + \beta_m E_i^m + \beta_p E_i^p + u_i, \quad i = 1, \dots, N \quad (3.11)$$

$$E_i^j = \alpha_0 + \gamma_j X_i + v_i^j, \quad j = m, p. \quad (3.12)$$

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 β'_s e os γ_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 os estimadores dos β'_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 (3.11) em razão do mesmo estar correlacionado com o termo de erro da equação (3.12) 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:

$$u_i = \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i \quad (3.13)$$

Sendo $\lambda^j = \frac{\text{cov}(v^j u)}{\text{var}(v^j)}$ quando não existir dependência entre a distribuição dos erros e os $X's$. Para tal, necessita-se estimar (3.12) os resíduos da forma reduzida duas vezes, isto é, tanto para a mãe quanto para o pai. Assim, estima-se:

$$E_i^f = \alpha_0 + \beta_0 X_i + \beta_m E_i^m + \beta_p E_i^p + \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i, \quad i = 1, \dots, N \quad (3.14)$$

onde ε_i denota o termo de erro de média zero o qual não é correlacionado com as variáveis explicativas inclusas. Contudo, a estimação de (3.14) não é possível com a ausência de restrições na forma reduzida apresentada em (3.12) 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 depende dos $X's$, ou seja, que em geral são heterocedásticos, o coeficiente em (3.13) torna-se:

$$\lambda^j = \frac{\text{cov}(v^j u | X)}{\text{var}(v^j | X)} = A^j(X) \quad (3.15)$$

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 método KV mostra:

$$\lambda^j = \frac{\text{cov}(v^j u | X)}{\text{var}(v^j | X)} = \rho^j \frac{\sqrt{\text{var}(u_i | X_i)}}{\sqrt{\text{var}(v_i^j | X_i)}} \quad (3.16)$$

O qual dada as suposições mostradas abaixo resulta em:

$$u_i = \rho^m \frac{H_{ui}}{\hat{H}_{vi}^m} \hat{v}_i^m + \rho^p \frac{H_{ui}}{\hat{H}_{vi}^p} \hat{v}_i^p + \varepsilon_i \quad (3.17)$$

No entanto, Farré, Klein e Vela (2013), impõem uma estrutura paramétrica através das seguintes suposições:

$$H_{ui}^2 = \exp(\alpha_1(Z_{ui}\theta_1)) \quad (3.18)$$

$$H_{vi}^{j2} = \exp(\alpha_2^j(Z_{vi}\theta_{2j})), j = m, p. \quad (3.19)$$

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 advogam que é possível estimar o seguinte modelo de função controle consistentemente.

$$E_i^f = \alpha_0 + \beta_0 X_i + \beta_m E_i^m + \beta_p E_i^p + \rho^m \frac{H_{ui}}{\hat{H}_{vi}^m} \hat{v}_i^m + \rho^p \frac{H_{ui}}{\hat{H}_{vi}^p} \hat{v}_i^p + \varepsilon_i, \quad i = 1, \dots, N \quad (3.20)$$

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 $H_v^j(X_i)$.

3.3.3 Descrição dos Dados

As estimativas da pesquisa serão possibilitadas pelos microdados do Exame Nacional do Ensino Médio – ENEM – no de 2010, captados no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP – no portal eletrônico do Ministério da Educação. 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, 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 destas 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 incorpora-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 3.1 – Descrição das Variáveis

Variável	Descrição
<i>Resultados individuais</i>	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.
<i>Renda familiar</i>	Somatório da renda familiar que são distribuídas em cinco classes de renda: <i>i)</i> Renda familiar mínima - Até um salário mínimo; <i>ii)</i> Renda familiar baixa - 1 a 3 salários mínimos; <i>iii)</i> Renda familiar média - 3 a 6 salários mínimos; <i>iv)</i> Renda familiar Alta - 6 a 9 salários mínimos; <i>v)</i> Renda familiar Máxima - acima de 9 salários mínimos.
<i>Escolaridades dos pais</i>	Agregadas em três estratos: <i>i)</i> Pais com ensino fundamental incompleto e analfabetos - menos de 4 anos de estudos; <i>ii)</i> Pais com ensino fundamental e/ou médio - de cinco a onze anos de estudos; <i>iii)</i> Pais com ensino superior ou pós-graduação - mais de onze anos de estudos.
<i>Tipo de Escola</i>	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;
<i>Sexo</i>	Feminino ou masculino;
<i>Raça</i>	Divididos em brancos e não brancos (aqui abrangendo pretos, pardos, amarelos e indígenas);
<i>Moradia</i>	Zona rural ou zona urbana
<i>Qualidade escolar</i>	Resultado do SAEB por unidade escolar.

Fonte: Elaboração própria, 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 as 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 3.1 sumariza os 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% vem de escola pública; aproximadamente 71% dos candidatos possuem renda familiar de no máximo três salários mínimos, em que 33% destes tem 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 de estudos, ou seja, possuem no máximo o primeiro ciclo ensino fundamental.

Tabela 3.1 – Análise Descritiva

Estudantes por Variáveis de Circunstâncias				
Sexo		Raça		
Masculino	Feminino	Brancos	Ñ Brancos	
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 própria, 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.

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 quanto comparada a escola privada.

3.4 PRINCIPAIS RESULTADOS

Nesta seção, serão apresentados os resultados da pesquisa. Primeiramente, na subseção 3.4.1, controem-se as densidades de Kernel sob perspectiva nacional e regional através de interações de circunstâncias. Nesta etapa admite-se a RIA. Por fim, subseção 3.4.2, infere-se sobre as relações das desigualdades de oportunidades no Brasil, considerando-se o esforço individual como uma variável observável, porém, omitida na inferência.

3.4.1 Primeiro Bloco

Neste primeiro momento, estima-se as densidades conjuntas de Kernel, descrita pela equação 3.9. De acordo com O'Neill, Sweetman e van DeGaer (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. Assume-se que baixo *background* social são aqueles indivíduos que detém as seguintes características de não responsabilidade: filhos de mães analfabetas, com renda familiar menor do que 1 (um) salário mínimo e que estudou 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 estudou

predominantemente em escola privada. O gráfico 3.1 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.

Gráfico 3.1 – Conjunto de Oportunidades – Brasil e Regiões

Gráfico 1A - Brasil

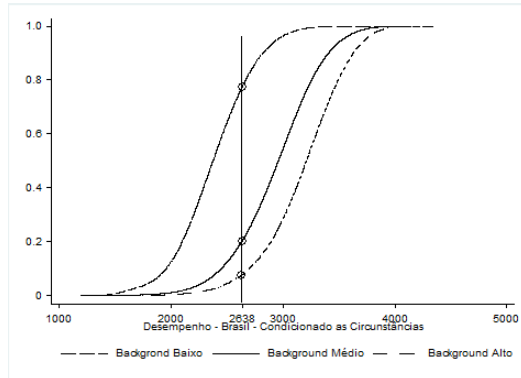


Gráfico 1B - Região Centro-Oeste

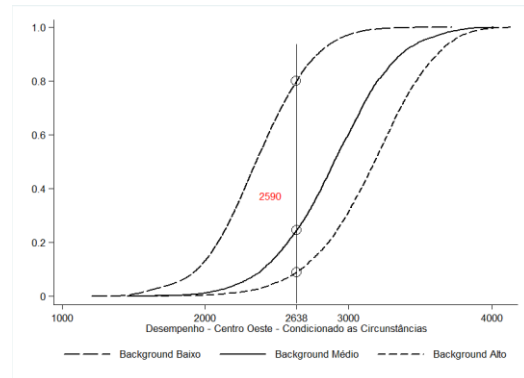


Gráfico 1C - Região Nordeste

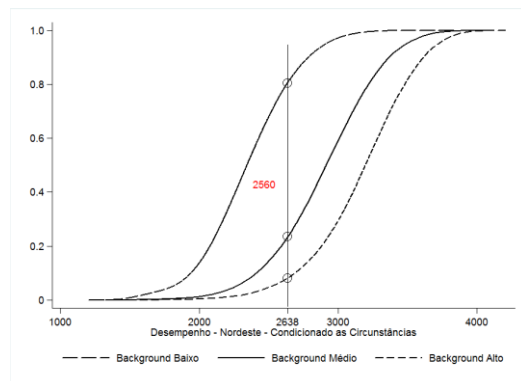


Gráfico 2D - Região Norte

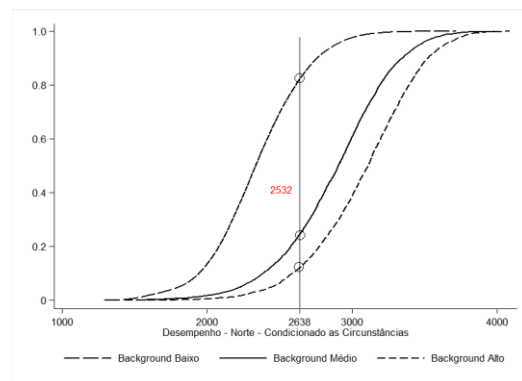


Gráfico 2E - Região Sul

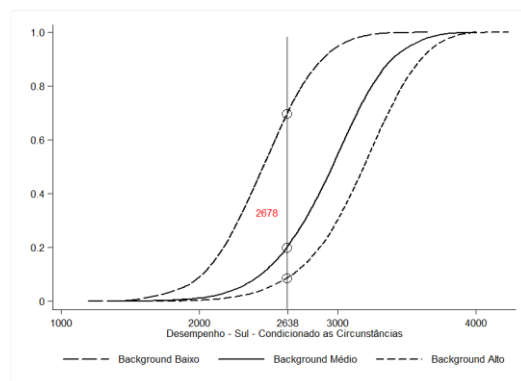
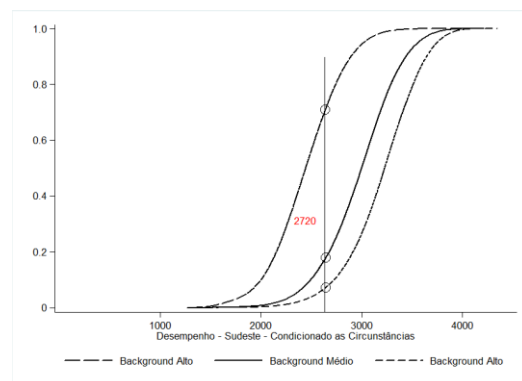


Gráfico 2F - Região Sudeste



Fonte: Elaboração própria, a partir de dados do ENEM.

A curva a direita representa os alunos que estão associadas às circunstâncias favoráveis e a curva a esquerda, por sua vez, os alunos associados as situações desfavoráveis. Ao observar o gráfico 3.1, 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 3.2, 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 menor²⁰ do que um indivíduo com alto *background* social.

Além do mais, a probabilidade de indivíduos com baixo *background* estar entre os 5% que obtém 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 desta 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

²⁰ O cálculo desta 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 $P_B = 0,2306$.

nacional. Respectivamente, a chance de um indivíduo pertencente ao percentil social inferior estar neste seletor grupo é de apenas 0,15% e 0,18%.

Tabela 3.2 – 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 própria, a partir de dados do ENEM.

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 nortista/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 obtém melhores resultados é de aproximadamente 1,74% contra 0,74%, 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 no 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 3.3 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.

Tabela 3.3 – 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 própria, a partir de dados do ENEM.

Nota: A renda corresponde: i) Renda familiar mínima: Até um salário mínimo; Renda familiar média: 3 a 6 salários mínimos; e v) Renda familiar Máxima: acima de 9 salários mínimos.

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 tem 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.

3.4.2 Segundo Bloco

Com a finalidade de iniciar a busca por respostas para questões como: quais fatores estão associados a esta 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 (3.10). Todavia, destaca-se que a fim de tornar os resultados paramétricos robustos será adotada a metodologia desenvolvida por Klein e Vela (2009, 2010) e Farré, Klein e Vela (2013) – daqui por diante, KV.

Os resultados da estimação OLS e KV são apresentados na Tabela 3.4. Eles assumem que o 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 3.4, notem que não há evidências de que as variáveis tem 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.

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. Neste sentido, os parâmetros associados a estas variáveis podem estar superestimados.

Para corrigir possíveis vieses decorrentes da situação descrita anteriormente, o estudo adota a metodologia de KV. De acordo com Klein e Vela (2010), o uso dessa abordagem evita o emprego de variáveis instrumentais em cenários em que os instrumentos não estão

facilmente disponíveis²¹, como é o caso aqui em questão. A diferença entre o coeficiente KV e OLS apurados, representa a medida do viés da variável educação obtida 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 (ρ mãe e ρ pai) que refletem o efeito indireto das variáveis não observadas.

Tabela 3.4 – Estimação KV e OLS

MODELO	OLS	KV	KV (<i>Bootstrap</i>)
Lognota	β_{ols}	β_{kv}	β_{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*
ρ mãe	-	0.02283*	0.02283*
ρ pai	-	0.01122	0.01123
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 (3.12) não somente captura os fatores não observáveis como, por exemplo, a habilidade. Entretanto, ρ^j positivo, sugere que o retorno da habilidade individual não observada é positivo, embora, isso pode incluir outras características como motivação, sorte. Estes 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 ρ^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, Klein e Vela (2013), no qual realizaram o

²¹ Ver, por exemplo, Angrist e Krueger (1991), Farré, Klein e Vela (2013).

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.

Observa-se também que, 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 brancos apresentam uma considerável vantagem no desempenho educacional sobre estudantes não brancos. Tais resultados apresentam-se mais robustos que em outros trabalhos, por exemplo, Farré, Klein e Vela (2013), pois eles, somente encontram efeitos significativos de sexo e raça após estimar com a função controle²², embora, a direção dos resultados sejam os mesmos.

A maior parte dos resultados encontrados apresenta-se consistentes com a teoria, uma vez que refletem 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, Pistolesi e Trannoy (2009) e Barros et. al. (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 o 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 esta tem influência significativa sobre o desempenho no exame. Em resumo, estudar em escola pública é uma grande desvantagem. Ao assumir este fato como verdadeiro, diante dos resultados empíricos apresentados destaca-se “o porquê disto”. Assim,

²² 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.

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.

3.5 CONSIDERAÇÕES FINAIS

Neste 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 no ano de 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 estes 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 obter melhores resultados. Por exemplo, um indivíduo com baixo *background* precisa esforçar em torno de 99,38% a 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 a robustez desses resultados, 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, são do sexo masculino, são considerados brancos e que estudaram em escolas 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 às discussões a respeito das políticas voltadas a 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.

4 A SORTE SORRI PARA TODOS? UMA NOVA ABORDAGEM EMPÍRICA DE DESIGUALDADE DE OPORTUNIDADES.

4.1 INTRODUÇÃO

A abordagem clássica tradicional da Teoria de Igualdade de Oportunidades – Roemer, 1998 – advoga que o sucesso ou fracasso individual é explicado por uma combinação de fatores determinísticos e aleatórios. Sinteticamente, o resultado individual, é determinado por três vetores de variáveis: i) variáveis de circunstâncias, as quais fogem do controle individual, de forma que, toda desigualdade social oriunda destas é considerada nociva, e, portanto, devem ser compensadas; ii) variáveis de esforço ou de responsabilidade individual, nas quais são observadas como origens de desigualdade social justa, e, desse modo, não devem ser anuladas perante esta abordagem; e por último, iii) por um termo de sorte bruta.

Em relação ao último conjunto, um valor positivo é habitualmente descrito como boa sorte. Segundo Lefranc, Pistolesi e Trannoy (2009), nos debates filosóficos sobre a igualdade de oportunidades, o conceito de sorte se refere a situações em que a responsabilidade do controle individual de escolha ou moral, não tem qualquer relação com a ocorrência dos resultados. É interessante destacar que, duas hipóteses chaves são assumidas nesse arcabouço teórico: a) a independência – exogeneidade – atribuída entre os vetores de variáveis de esforço e circunstâncias; e b) o caráter totalmente aleatório atribuído à sorte no resultado individual.

Essa caracterização abre espaço para, pelo menos, duas contestações: primeiro, a independência assumida entre as variáveis de circunstância e de esforço; segundo, a aleatoriedade total suposta entre as componentes do fator sorte que contempla o termo de erro.

Nesse ensejo, Bourguignon, Ferreira e Menéndez (2007), examinam o tópico (a), embora, haja reservas ao tratamento dado a endogeneidade das covariadas, em especial, devido à omissão de variáveis. Por outro lado, Lefranc, Pistolesi e Trannoy (2009) postulam diferentes vertentes filosóficas e econômicas referentes ao tópico (b). Entretanto, em nenhum momento cogita-se mensurar a participação da sorte no resultado individual. Além disso, os autores caracterizam a sorte, em todas as suas especificidades, como fator de não responsabilidade individual, atribuindo-a apenas a função de captar aspectos que fogem da responsabilidade individual. Nesse sentido, há uma questão fundamental a ser investigada.

Isto é, há componentes determinísticos no fator sorte que são tratados como aleatórios pela moderna teoria de igualdade de oportunidades? Se a resposta for sim, então quais seriam?

Nessa perspectiva de investigação, a priori faz-se necessário explicar uma visão geral da sorte na literatura. Nessa discussão, Lefranc, Pistolesi e Trannoy (2009), assume a existência de pelo menos quatro tipos de sorte nas mais influentes correntes econômicas e filosóficas, assim designadas.

1. **Sorte Background Social:** denominada por Rawls de loteria social. Neste tipo de sorte, os resultados individuais somente diferem em função das diferenças em suas conexões sociais – *rent seeking* – familiares.
2. **Sorte Constitutiva ou Genética:** atribuído aos casos onde as diferenças nos resultados individuais acontecem em razão da herdabilidade genética. Também conhecida como loteria natural.
3. **Sorte Opção:** denominada assim por Dworkin. Especificamente, considere dois indivíduos que ambos tenham que escolher entre duas loterias. Admita também, que o resultado da primeira loteria é dado como certo e o da segunda, aleatório. Adicionalmente considere que os indivíduos diferem em suas escolhas e resultados. Tal fato faz com que o risco assumido seja deliberado e o resultado individual seja fruto de tomar tal risco.
4. **Sorte Bruta:** Suponha que dois indivíduos possuam talento e background social idêntico. E que o diferencial de resultado ocorre apenas em função de um não ter como escapar de uma determinada loteria. Por exemplo, uma guerra, a qual somente um dos indivíduos é designado para batalha.

Posto isso, percebe-se claramente que a sorte background social é tratada adequadamente pela literatura tradicional de igualdade de oportunidades. Por outro lado, a sorte genética ou constitutiva, seja em função da difícil mensuração ou da ausência de fatores que captem seus efeitos é considerada como sorte bruta. Além dessa limitação, ao menos empírica, parece existir também uma clara controvérsia teórica, quanto à nomeada *sorte opção*. Uma vez que, a literatura descreve-a como parte da sorte em que o indivíduo tem controle ou responsabilidade individual de escolha. Contudo, em trabalhos seminais, por exemplo, Roemer (1998), tais fatores são desconsiderados. Em linhas gerais, nos atuais modelos de Igualdade de Oportunidades – IOPs – fatores referentes à genética, opção escolhida e sorte bruta compõe o termo de erro aleatório, e, portanto, origina dúvidas sobre os limites dessa especificação.

Nesse contexto, um importante avanço empírico pode ser alcançado se considerarmos a opção individual escolhida – *sorte opção* – como um dos fatores determinantes do resultado individual, uma vez que, é fruto de uma escolha racional baseada em expectativas observadas pelos agentes. Adicionalmente, a fim de melhor avaliar o comportamento do resultado individual, adota-se um procedimento simples que considera as características individuais não observáveis e algumas observáveis, porém, em grande parte não disponíveis.

Em outras palavras, a estratégia empírica proposta, leva em conta fatores não observáveis como motivação, herdabilidade genética, ambiente individual compartilhado e habilidade cognitiva geral. Ao invés de simplesmente supor uma aleatoriedade, em especial, justificada pela ausência de variáveis que representem tais fatores. Além disso, outros fatores, observáveis, mas, não disponíveis na maioria da base de dados, também são incorporados na análise. Diante do exposto, este trabalho tem como principal objetivo investigar se há componentes do fator sorte que influenciam o resultado individual, como também, objetiva controlar tanto o efeito de importantes características individuais observáveis, embora, algumas nem sempre disponíveis, como também, as principais características não observáveis que afetam o desempenho individual.

Para tanto, faz-se uso de uma amostra composta por 680 gêmeos monozigóticos coletados no Festival do Dia Anual dos Gêmeos em Twinsburg, Ohio, ocorridos no período de 1991 a 1994. Há duas características marcantes dessa base dados, primeiro, o controle imediato de diversos fatores consensualmente apontados na literatura como determinantes da desigualdade indesejada ou injusta. Segundo, e, vista como a principal vantagem dos dados utilizados, isto é, o controle substancial de diversas características não observáveis.

Por sua vez, a variável escolhida para representar a *sorte opção* é a filiação sindical, dado que, ser sindicalizado não necessariamente se caracteriza circunstância, nem muito menos, esforço individual. Uma vez que, é produto de escolha fundamentada em expectativas racionais dos indivíduos. Nesse sentido, a fim de proporcionar uma estrutura empírica satisfatória para nossa análise, adota-se a princípio a metodologia de inferências contrafatuais *Propensity Score Matching* desenvolvido por Rosenbaum e Rubin (1983), que se apresenta consistente com os objetivos propostos. Entretanto, faz-se outro exercício baseado na abordagem proposta por Ashenfelter e Rouse (1997), com a finalidade de superar algumas limitações identificadas nos atuais modelos de igualdade de oportunidades.

Esta pesquisa encontra-se subdividida em seis seções, além dessa introdução. Na segunda seção faz-se um resumo da literatura, focado essencialmente, em desigualdade de oportunidades, sorte, habilidade cognitiva geral, herdabilidade genética e ambiente

compartilhado. Na terceira seção apresenta-se a estratégia empírica. Na quarta descreve-se e analisam-se os dados. A quinta seção é destinada à apresentação e discussão dos principais resultados, e, por fim, a seção final é reservada as considerações finais.

4.2 REVISÃO DE LITERATURA

Nesta seção faz-se uma breve caracterização da teoria de igualdade de oportunidades, como também, a respeito das variedades da componente sorte, em especial, destacando a distinção entre a sorte opção e sorte bruta. Além disso, reportam-se sinteticamente alguns dos principais trabalhos que tratam das principais características individuais não observáveis destacadas nesse trabalho, por exemplo, habilidade cognitiva geral, herdabilidade genética e ambiente compartilhado.

4.2.1 Igualdade de Oportunidades

O aparato geralmente utilizado pelos modelos de Igualdade de Oportunidades – IOPs – defende que o resultado individual, w_i , é produto de fatores de responsabilidade e não responsabilidade individual. Em síntese, o resultado obtido pelo agente é composto por uma gama de fatores que representam: i) as circunstâncias, C_i , por exemplo, características familiares, socioeconômicas, etc.; ii) o esforço despendido, E_i , como, horas trabalhadas, anos de estudo, etc. e; iii) por um termo de sorte bruta, u_i . Onde $f(.)$ é uma função desconhecida.

$$w_i = f(C_i, E_i, u_i) \quad (4.1)$$

Nos últimos anos, há estudos na literatura que criticam a estrutura da equação (4.1), especialmente, envolvendo a suposta exogeneidade – independência – entre as variáveis de circunstâncias e esforço, como também, referente às componentes do termo de erro. Todavia, esses problemas parecem terem sido abordados de forma superficial e insipiente. Por exemplo, em Bourguignon, Ferreira e Menéndez (2007), o estudo propõe uma decomposição da desigualdade de renda em dois aspectos, um que contempla o conjunto de oportunidades, e outro que vislumbra o termo residual. Em síntese, buscam verificar a existência de efeitos

diretos e indiretos das circunstâncias sobre o resultado individual, onde os últimos agem sobre o esforço individual despendido. Os dados utilizados referem-se a ganhos salariais de homens e mulheres para o Brasil no ano de 1996. Os resultados indicam um efeito indireto das circunstâncias em torno de 30% sobre a desigualdade observada. Todavia, a estrutura utilizada para estimar tais efeitos não difere substancialmente da equação (4.1).

Por outro lado, Lefranc, Pistolesi e Trannoy (2009), relatam a existência e distinção dos componentes do termo de erro, em particular, aos fatores ligados à sorte. Na verdade, o estudo busca ofertar um modelo de igualdade de oportunidades flexível bastante para englobar as mais diversas concepções sobre a teoria de igualdade de oportunidades. Adicionalmente, faz-se uma aplicação empírica referente à renda na França durante o período 1979 a 2000. Os resultados revelam que o nível de desigualdade de oportunidades tende a diminuir e que o grau de risco das distribuições de renda, condicionado à origem social, são semelhantes em todos os grupos de origens sociais observados. Contudo, apesar desses estudos apontarem algumas limitações da equação (4.1), tais limitações não foram superadas. Além disso, em nenhum desses estudos atentou-se para as características individuais não observáveis.

Nesse sentido, a principal contribuição desse estudo é superar algumas importantes limitações estruturais e empíricas presentes na equação (4.1), enumeradas a seguir. Primeiro, considerar que as componentes do fator sorte que representa o termo de erro são totalmente aleatórios. Segundo, assumir que os indivíduos são amplamente idênticos, não se levando em conta fatores como: herdabilidade genética, ambiente compartilhado e a habilidade cognitiva geral dos mesmos. Nesses termos, parece que os resultados apresentados pela literatura predominante de IOPs possuem importantes limitações.

Posto isto, este estudo adota um recurso simples capaz de resolver os problemas acima mencionados. Na verdade, o principal diferencial desse estudo, é trabalhar com uma base de dados composta por gêmeos monozigóticos, isto é, gêmeos idênticos. De forma que, a própria base de dados controla por diversos fatores não observáveis, em especial, os cinco últimos citados acima. Além disso, engloba outras características individuais não disponíveis na maioria dos conjuntos de dados. Em outras palavras, o emprego dessa base tem uma característica marcante, assumindo, por exemplo, que a habilidade cognitiva geral, ambiente compartilhado e herdabilidade genética são fundamentais na formação dos talentos individuais, e, portanto, tais características influenciam substancialmente os resultados individuais obtidos. Como visto, uma simples abordagem pode tornar os resultados bem mais robustos. Disto isso, vale ressaltar, que a princípio faz-se uso de uma estratégia empírica que

estime a equação (4.1) não considerando essas limitações. Isto é, faz-se o que em geral se observa na literatura de IOPs. Em seguida, emprega-se a segunda estratégia com o propósito de identificar o *gap*, caso haja, da omissão desses aspectos nos resultados estimados. Em resumo, comparam-se os resultados negligenciando ou considerando as referidas limitações.

Nessa perspectiva, a subseção 4.2.2 detalha a importância do fator sorte no resultado individual. Por outro lado, a seção 4.2.3, destaca fatores como a herdabilidade genética, ambiente compartilhado e habilidade cognitiva geral.

4.2.2 Sorte

Nesta subseção detalharemos as quatro principais especificações do fator sorte abordado pela literatura econômica e filosófica. Todavia, primeiramente, faz-se uma distinção entre sorte bruta *versus* sorte opção. Uma vez que, essa diferenciação, é parte fundamental deste trabalho, pois, em geral, a literatura de igualdade de oportunidade trata a sorte opção como uma complementação da sorte bruta. E, ao que parece, conforme anteriormente abordado e discutido a seguir, esse não é o tratamento mais adequado para este problema. Para ter-se uma ideia, segundo Vallentyne (2002), há em geral, quatro classificações distintas sobre a sorte bruta na literatura, sendo elas: i) razoável não previsibilidade de escolha, ii) razoável inevitabilidade, iii) razoável incapacidade de influenciar, e, por fim, iv) não razoável influência deliberada. A primeira visão, conhecida como, não previsibilidade de escolha, defende que a determinação de um evento para um indivíduo somente é considerada função da sorte bruta, se e somente se, tal ocorrência não seja resultado esperado de suas escolhas.

Por exemplo, admita duas situações, na qual em ambas, há dois agentes similares e identicamente situados. No primeiro caso, suponha que não existem possibilidades de escolha. Isto é, ambos participam de uma loteria natural – exposição à queda de raios – na qual o resultado é totalmente aleatório e contabilizado a sorte bruta. Por sua vez, na segunda situação, suponha que ambos possuem duas escolhas – parar ou continuar andando na tempestade – embora, suas escolhas não exerçam qualquer efeito sobre as probabilidades de ser atingido ou não por um raio. Isto é, em ambas as situações, os agentes são expostos à mesma loteria natural. Além disso, admita também, que cada agente seja totalmente consciente do resultado de cada escolha. Suponha que cada agente decide parar e somente um deles é atingido por um raio. Tendo em vista, que a probabilidade de ser atingido por um raio,

dado as suas escolhas, eram razoavelmente previsíveis pelos agentes, o ponto de vista acima qualificaria esta sorte como bruta.

Dado que, nos casos acima citados, nitidamente não há razão para tratar ambos os eventos de forma diferenciada. Uma vez que, envolvem as desigualdades inevitáveis no resultado da sorte. Pois, não havia nada que os agentes pudessem fazer para alterar a sua exposição aos raios. Por isso, a presença de escolha, mesmo plenamente informada não necessariamente é suficiente para fazer de um resultado consequência de uma opção.

A segunda abordagem é, sem dúvidas, uma forma mais clara de compreender a sorte bruta. Isto é, considerá-la em termos de provável inevitabilidade de um resultado. Neste caso, a influência de um evento no resultado individual é devido à sorte bruta de um indivíduo, se e somente se, o mesmo não possuir razoável possibilidade de evitar sua ocorrência. Em linhas gerais, a sorte bruta, em termos de não evitabilidade, leva em conta se o indivíduo possuía em algum momento a possibilidade de evitar um determinado resultado, e, portanto, se o mesmo tiver como prever sua ocorrência não faz do resultado uma questão de sorte bruta. Por exemplo, sofrer um ataque de tubarão em área de risco poderia ter sido evitado se o banhista atendesse as placas de sinalização de perigo. Deve-se destacar a distinção entre a abordagem de não previsibilidade de escolha e a de inevitabilidade. Uma vez que, na segunda, a escolha importa e impacta consideravelmente os resultados.

Dito de outro modo, segundo Dworkin (1981), a sorte opção é a forma como um indivíduo leva vantagem ou não, através da aceitação de um risco deliberado e calculado, ou da antecipação de um evento na atenuação de um resultado ruim. Por outro lado, a sorte bruta na sua forma pura não são apostas deliberadas. De forma mais concisa, uma interpretação estrita da abordagem de inevitabilidade, é que, se o indivíduo tem de alguma forma como evitar sua exposição a determinado risco, não se deve considerar a diferença no resultado individual fruto da sorte bruta. Porém, há uma limitação nessa abordagem, pois não incorpora a razoabilidade de evitar o resultado em questão. Entretanto, em uma visão menos estrita, e, talvez mais prudente, compõe a sorte bruta apenas os fatos que influência nos resultados e que não são razoavelmente evitáveis pelo indivíduo. Todavia, destaca-se que, se existe dois indivíduos identicamente situados, e que ambos fazem escolhas diferentes, então, portanto, o impacto diferencial de suas escolhas é ao que parece, devido à opção escolhida.

Vale ressaltar que, a caracterização da sorte bruta como razoável inevitabilidade tem algumas limitações. Uma das principais seria o fato que é uma simplificação útil e adequada da incapacidade de influenciar. Essa caracterização ignora a possibilidade que os indivíduos possuem de influenciar seus resultados, modificando suas probabilidades. Nesse sentido,

segundo a literatura, uma visão mais geral seria a terceira caracterização. Ou seja, sorte bruta como razoável incapacidade de influenciar um resultado. Nesta abordagem, a sorte bruta influencia o resultado individual, se e somente se, o indivíduo não tem como influenciar razoavelmente a possibilidade ou probabilidade de sua ocorrência.

Essa caracterização advoga que a inevitabilidade é o modo no qual os resultados podem ser influenciados por um indivíduo. Porém, há um problema adicional. Pois, o “se e somente se” nesta especificação é problemática, uma vez que, exige que para determinado resultado seja atribuído à sorte não bruta ou opção, o indivíduo terá que possuir alguma influência mínima sobre o resultado. Contudo, se houver duas escolhas e as mesmas possuírem idênticos *payoffs* e probabilidades, então, deve-se considerar a influência da escolha no resultado como sorte bruta.

Nessa configuração, segundo Vallentyne (2002), uma forma razoável seria substituir o termo “se e somente se” pela expressão “na medida em que” e, em seguida, decompor os resultados decorrentes da sorte bruta e sorte opção de acordo com o nível de participação que o indivíduo exerce sobre o resultado. Embora, essa tarefa não seja trivial. Adicionalmente, há outra questão importante a entender. Visto que, até o momento considerou-se um ambiente em que os indivíduos possuem perfeita informação sobre as opções disponíveis. Entretanto, sabe-se que isso não retrata completamente a realidade. Pois, embora, em algumas situações, haja eventos não previsíveis, os mesmos podem ser influenciáveis a um determinado resultado. Todavia, como influenciabilidade é entendida nesse contexto como razoável previsibilidade, assim, portanto, por simplificação, as crenças falsas ou incompletas são ao que parece corretamente tratada como uma questão de sorte bruta. Contudo, o autor ressalva os casos onde os resultados das opções escolhidas por indivíduos idênticos, e que possuem previsibilidade imperfeita, poderia ser atribuída ao diferencial da sorte bruta, mesmo quando eles pudessem influenciar os resultados.

Em outro momento, Vallentyne (2002), também destaca que o simples fato de que a possibilidade de um evento seja resultado previsível de escolhas dos indivíduos não é uma condição suficiente para que o resultado seja atribuído à sorte opção. Entretanto, sugere ser uma condição necessária. Por fim, surge a seguinte caracterização de sorte bruta, isto é, a sorte bruta como não razoável influência deliberada. Em síntese, essa caracterização da sorte apregoa que a ocorrência ou não de um determinado evento é devido à sorte bruta de um indivíduo, na medida em que, o mesmo não pode ter razoável deliberação que influencie a possibilidade ou probabilidade de sua ocorrência ou não.

Em linhas gerais, a inevitabilidade bruta – núcleo – é a principal distinção entre a sorte bruta e as outras componentes do fator sorte, aqui especificamente, confrontada com a sorte opção. Contudo, também se devem levar em conta, os acontecimentos para os quais os indivíduos não têm a capacidade de influenciar a probabilidade. Como também, aqueles para os quais os mesmos, não têm consciência de sua disposição de impactar suas possibilidades, sejam por causa de problemas de crenças falsas ou incompletas. Adicionalmente, a literatura, em geral, inclui nessa gama, os fatos que não são razoavelmente propícios à influência deliberada, mesmo quando a rigor são tão influenciáveis, embora, não seja claro como isso ocorra.

Todavia, segundo Lefranc, Pistolesi e Tranoy (2009), os debates sobre igualdade de oportunidades tem nomeado ao menos quatro tipos de sorte, as quais foram sumariamente citadas. Isto é, sorte background social, genética, opção e sorte bruta.

A primeira, sorte background social, revela que os diferenciais existentes entre dois indivíduos identicamente talentosos e motivados, é função das conexões e fundo social das suas famílias. Isto é, os diferenciais resultantes são pré-determinados por fatores antecedentes, como origem familiar e social. O segundo tipo, sorte genética, advoga que os diferenciais resultantes, por exemplo, entre dois irmãos gêmeos decorrentes de um deles ter uma herdabilidade genética superior, de modo a gerar um talento especial, deve ser configurada como sorte genética ou constitutiva. Vale ressaltar que nesse caso não estar-se considerando gêmeos monozigóticos.

A terceira, conhecida como sorte opção ocorre, por exemplo, quando a diferença nos resultados existentes é decorrente apenas das escolhas frente às opções disponíveis, excetuando-se, as loterias não evitáveis. Nessa análise considera-se que os indivíduos sejam identicamente talentosos e motivados, e, que, além disso, possuam origens familiares e sociais similares. E por fim, a sorte bruta, já amplamente enfatizada e sinteticamente conceituada, como o tipo de sorte responsável por um resultado de uma loteria inevitável. Por exemplo, o diferencial no resultado econômico individual atribuído a dois irmãos identicamente talentosos e situados, onde os efeitos nocivos de uma guerra afeta somente um deles, pois apenas um deles foi designado para batalha.

4.2.3 Habilidade Cognitiva Geral, Herdabilidade Genética e Ambiente Compartilhado

Agora reportar-se alguns trabalhos referentes à habilidade cognitiva geral, herdabilidade genética e ambiente compartilhado. Entretanto, primeiramente apresenta-se uma definição fundamental. Seguindo a evolução histórica, desde o estudo secular de Spearman (1904), a literatura define a habilidade cognitiva geral (g) como a capacidade que os indivíduos adquirem competências e permitem-nos interagir em seu ambiente. Tais habilidades constituem a estrutura essencial das competências cognitivas individuais. Por exemplo, dominar linguagens, compreender fenômenos, enfrentar situações e problemas, construir argumentações, elaborar propostas, entre outras aptidões individuais.

Nesse contexto, Herskind *et al.* (1993), aponta que a importância genética varia de 0,4 a 0,6 contra 0,3 a 0,4 dos efeitos ambientais compartilhados em g em indivíduos de 4 a 20 anos. O estudo também mostra que as influências ambientais compartilhadas caem virtualmente à zero na idade adulta.

Em outro estudo, Chipuer, Rovine e Plomin (1990), encontra-se uma herdabilidade (h)² em cerca de 50%. Em outras palavras, a variação genética é responsável por cerca de metade da variação fenotípica em g .²³ Onde, a participação relativa das influências genéticas e ambientais sobre a capacidade cognitiva difere substancialmente ao longo da vida, com o aumento da herdabilidade e queda da influência ambiental compartilhada desde a infância e adolescência até a idade adulta. Nessa mesma linha, Wilson (1983), sugere que o ambiente compartilhado responde por 60% da variância em g na primeira infância.

Por sua vez, Trouton, Spinath e Plomin (2002), investigam as contribuições genéticas e ambientais em gêmeos de 2, 3, 4 e 7 anos de idade após mudanças e continuidade na linguagem e desenvolvimento cognitivo. Os resultados sugerem que os mesmos genes contribuem largamente para ambos os problemas de linguagem e desenvolvimento cognitivo e afetam o desenvolvimento normal e anormal. Em outro estudo, Ronald, Spinath e Plomin (2002), destacam a divergência entre os defensores e contrários da contabilização do talento individual. Para os adeptos, há fatores biológicos individuais pré-disponíveis como um potencial para o talento que pode ser identificado em algumas crianças e outras não. Para os opostos a esta concepção, a aquisição de habilidades por meio de exposição precoce, prática e

²³ O fenótipo são as características observáveis ou caracteres de um organismo ou população como, por exemplo: morfologia, desenvolvimento, propriedades bioquímicas ou fisiológicas e comportamento.

outros processos estão mais relacionados a fatores ambientais. Contudo, os resultados indicam que mais de 70% da diferença média entre o grupo de pontuação elevada e toda a amostra – crianças de 2 a 4 anos de idade – são determinadas por influências ambientais compartilhadas. E somente um quinto da diferença média é devido a fatores genéticos. Estes resultados corroboram o argumento de que os primeiros "talentos" são movidos por fatores ambientais compartilhados. Em suma, o potencial biológico especial não foi evidente na amostra investigada avaliada durante os anos pré-escolares.

Nessa mesma linha de pensamento, Spinath *et al.* (2003), destacam que g é uma das construções mais confiáveis e válidas sobre o comportamento humano. Segundo os autores, a estabilidade de g no longo prazo após a infância é maior do que qualquer outro traço comportamental e prevê resultados sociais importantes, tais como os níveis educacionais e ocupacionais muito melhor do que qualquer outra característica. Outro fato interessante é a convergência recente de resultados em direção da conclusão de que fatores genéticos contribuem substancialmente para as diferenças interindividuais de g .

Nesse mesmo estudo, os resultados remetem três importantes questões. Primeiro, o fenótipo g é claramente evidente nos primeiros anos de vida. Segundo, a influência genética é em torno de 20% a 30% menor na primeira infância do que a encontrada na infância média. Algo próximo de 40% inferior ao observado na adolescência, e por volta de 50% menor que o valor estimado após a adolescência. Em terceiro lugar, a influência ambiental compartilhada é maior na infância do que após a adolescência, quando sua importância declina a níveis insignificantes. No entanto, as estimações captam alta influência do ambiente compartilhado, algo em torno de 60%. Embora, os autores justificam que esse valor pode incluir influências específicas de gêmeos e, talvez, seja produto das medidas utilizadas.

Ratificando os resultados anteriores, Plomin e Spinath (2004), relatam que se sabe mais sobre a genética da inteligência do que sobre qualquer outra característica comportamental ou biológica. Adicionalmente, o estudo destaca que as duas das descobertas genéticas mais interessantes são: i) que a hereditariedade da inteligência aumenta ao longo do tempo de vida; ii) que os mesmos genes afetam diversas habilidades cognitivas.

Nesse contexto, Haworth *et al.* (2009), advogam que embora hajam muitas pesquisas genéticas que abordam variação normal em inteligência, pouco se sabe sobre a etiologia das habilidades cognitivas. Neste estudo, os autores investigam a genética e etiologias ambientais de g utilizando dados de 11.000 pares de gêmeos com idade variando entre 6 a 71 anos. Para tanto, usam um modelo de responsabilidade limiar de encaixe para estimar os parâmetros

genéticos e ambientais para aqueles que se situam entre os 15% top da distribuição de g . A influência genética para alta g foi substancial. Especificamente, estimado em 0,50 e variando num intervalo 0,41 a 0,60. Por outro lado, as influências ambientais compartilhadas responderam em média por 0,28 e variaram entre 0,19 a 0,37. O estudo concluiu que a variação genética contribui substancialmente para alta g na Austrália, Holanda, Reino Unido e Estados Unidos.

Nessa perspectiva, fica evidente a importância em controlar os fatores genéticos e ambientais compartilhados. Assim sendo, acreditamos haver um considerável ganho nesse estudo comparado aos demais estudos de IOPs. Pois, ao utilizar uma amostra de gêmeos monozigóticos controlamos de modo significativo as limitações a priori estabelecidas. Particularmente, em relação a diferenciar empiricamente a sorte opção da sorte bruta, como também, controlar os efeitos da sorte genética – herdabilidade – ou constitutiva, habilidade cognitiva geral, ambiente compartilhado, motivação, entre outros fatores não observáveis determinantes do resultado individual através da base de dados utilizada.

4.3 ESTRATÉGIA EMPÍRICA

Com o propósito de alcançar os objetivos traçados, este estudo utiliza a princípio o método de inferências contrafatuais *Propensity Score Matching* desenvolvido por Rosenbaum e Rubin (1983), doravante denominado *PSM*. De acordo com Cameron e Trivedi (2005), o *PSM* é um método de correspondência inexata, que ao invés de fazer correspondência com as variáveis explicativas faz correspondência com o escore de propensão. Neste processo, revela a probabilidade condicional de receber tratamento dado X , expressado por $p(X)$, como uma medida de correspondência. Conforme observado, se os dados justificam a correspondência em X , então, a correspondência com base no *PSM* é também justificada.

De acordo com a teoria contrafactual, a situação ideal seria observar o mesmo indivíduo em situações distintas, ou seja, participando e não participando de um determinado evento. Contudo, como não é possível observar o mesmo indivíduo, por exemplo, sendo fumante e não fumante, uma saída para isso é apurar a causalidade e tamanho da influência de determinados fatos através da comparação de dois grupos estatisticamente semelhantes.

Nesse sentido, este método, nada mais é, do que um estimador de correspondência baseado na variância que faz uso do escore de propensão estimado $\hat{P}(X, Z)$ a fim de buscar e impor pesos aos pares simétricos (i, j) com status de tratamento opostos, dados os valores de X e Z .

O objetivo, na verdade, é procurar uma correspondência exata, ou seja, o *PSM* busca encontrar pares para os quais a soma das probabilidades dentro de cada par seja igual a um. Caso isso não ocorra, então, as unidades de comparação são aquelas cujo escore de propensão é suficientemente próximo da unidade tratada. Em outras palavras, o *PSM* procura estabelecer um grupo de controle idêntico ao grupo de tratamento em função de certas características observáveis.

A fundamental hipótese assumida no *PSM* é que o vetor de covariadas X contempla todas as características responsáveis pelo resultado potencial na ausência do tratamento que o indivíduo possui ao decidir ingressar ou não em um determinado tratamento. Essa hipótese é denominada de condição de não confundimento ou seleção nos observáveis.

$$y_i(0) \perp D_i | X_i \quad (4.2)$$

Desse modo, para obter-se o efeito tratamento médio via *PSM*, faz-se necessário possuir indivíduos no grupo de controle que corresponda identicamente a cada indivíduo no grupo de tratamento, a fim de apurar qual seria seu resultado caso o mesmo não estivesse submetido ao tratamento proposto. Dito de outra forma faz-se imprescindível obter um grupo de controle que possua as mesmas características observáveis contidas no vetor X que as correspondentes do grupo de tratamento. Nessa configuração, deve-se atender a segunda hipótese, denominada hipótese de sobreposição. Em linhas gerais, a mesma estabelece que o grupo de controle contenha indivíduos com características similares aos indivíduos pertencentes ao grupo de tratamento. Tal hipótese é expressa por.

$$P[D_i | X_i < 1] \quad (4.3)$$

Uma limitação evidente do *PSM* é que ao assumir as hipóteses em (4.2) e (4.3), admite-se que o vetor de características observáveis contempla todos os fatores que determinam os indivíduos participar ou não tratamento. Isto é, a decisão de participar do tratamento é independente de fatores não observáveis. Dito de outro modo, se os fatores não observáveis determinarem a decisão de participar do tratamento, como também, influenciar o resultado potencial individual, o *PSM* está sujeito ao viés de seleção.

Quanto à questão de implementação, temos que considerar na execução do *matching* baseado em $p(X_i)$, segundo os autores, três questões relevantes: i) *matching* com ou sem

substituição; ii) o número de unidades usadas na comparação conjunta; e, iii) a escolha do método de *matching*.

No primeiro caso, *matching* sem substituição significa que qualquer observação no grupo de controle é comparada com não mais do que uma do grupo de tratamento, o qual é a correspondência mais próxima. Enquanto que, com substituição, poderá haver várias partidas. Contudo, se o *matching* adotado for sem substituição e o grupo de controle for muito pequeno, então, os matches podem ser muito próximos em termos de $p(X_i)$, o que irá aumentar consideravelmente o viés do estimador.

Por outro lado, a escolha do número de observações do grupo de controle envolve uma permuta entre viés e variância. Pois, usando um número mais próximo do grupo de tratamento, consegue-se reduzir o viés, embora, aumente-se a variância. Enquanto que, incluindo controles mais correspondentes, a variância se reduz, contudo, o viés aumenta particularmente, se as observações adicionais do grupo de controle são matches de qualidade inferior às observações do grupo de tratamento. O PSM tem como recurso parcial, o emprego de uma vizinhança predefinida em termos de um raio em torno da observação tratada, descartando assim matches que se encontram fora desta vizinhança. Em linhas gerais, utilizam os melhores resultados. Esse procedimento é conhecido como *matching calibrado*.

Contudo, a sensibilidade dos resultados perante o método utilizado não é passível de uma resposta simples. Pois, os resultados podem variar em diferentes amostras, dependendo do grau de sobreposição entre as observações do grupo de tratamento e do grupo de controle. Por um lado, caso os grupos sejam idênticos, em termos de sobreposição substancial dos seus escores de propensão, e se o grupo de controle é grande, então, os matches serão mais acessíveis. De modo que, combinar através do método de substituição é a melhor alternativa. Por outro lado, caso o grupo de controle seja pequeno e díspar em relação às observações do grupo de tratamento, então, os matches podem ser insatisfatórios, e, portanto, deve-se adotar o método sem substituição.

Nessa configuração, denote o grupo de controle do caso tratado i com características X_i como o conjunto $A_j(X) = \{j | X_j \in c(X_i)\}$, onde $c(X_i)$ representa as características da vizinhança de X_i . Além disso, suponha N_c representar o número de casos do grupo de controle e $w(i, j)$ o peso dado ao j th não tratado com o i th caso tratado, onde $\sum_j w(i, j) = 1$, sendo a fórmula geral do *matching* do estimador Efeito Tratamento sobre o Tratado – ATET – expressada por:

$$\Delta^M = \frac{1}{N_T} \sum_{i \in \{D=1\}} y_{1,i} - \sum_j w(i,j) y_{0,j}. \quad (4.4)$$

Onde $0 < w(i,j) \leq 1$ e $\{D = 1\}$ denota o conjunto dos indivíduos tratados, e j é um elemento do conjunto de *matching* de unidades de comparação. Diferentes *matching* estimadores são gerados por variar a escolha de $w(i,j)$. Por exemplo, o *matching* simples compara células com exatamente a mesma x discreta.

$$\Delta^M = \sum_{k \in \{x\}} w_k [\bar{y}_{1,k} - \bar{y}_{0,k}]. \quad (4.5)$$

Sendo $\bar{y}_1$ o resultado médio do tratado e $\bar{y}_0$ representa o resultado médio do não tratado, e w_k é o peso atribuído da k th célula. Ou seja, a fração das observações na célula k . Segundo Cameron e Trivedi (2005), um caso específico é encontrado em Dehejia e Wahba, (2002).

$$\Delta^M = \frac{1}{N_T} \sum_i \left(y_1 - \frac{1}{N_{C,i}} \sum_{j \in \{D=0\}} y_j \right) \quad (4.6)$$

Em que N_T é o número de casos do grupo de tratamento ($D = 1$) e $N_{C,i}$ corresponde ao número de casos do grupo de controle correspondente a i th observação. Por sua vez, o método de *matching* baseado na vizinhança mais próxima, o conjunto $A_i(X) = \{j | \min_j \|X_i - X_j\|\}$. Aonde $\| \cdot \|$ denota a distância euclidiana entre os vetores. Se $w(i,j) = 1$ na equação (4.1), então, $j \in A_i(X)$, e zero caso contrário, de forma que, esta especificação utiliza apenas um caso para construir o grupo de controle para os casos tratados. Outro tipo de estimador gerado é o *matching* kernel, o qual é expresso pela equação 3.7 a seguir.

$$w(i,j) = \frac{K(X_j - X_i)}{\sum_{j=1}^{N_{C,i}} K(X_j - X_i)}. \quad (4.7)$$

Onde K representa a função kernel anteriormente citada. Vale ressaltar que, uma das principais vantagens dos métodos apresentados até o momento, e que, os mesmos tem a vantagem de não assumir determinadas suposições da forma funcional para as equações de resultados na estimativa ATET e pode ser estimado por valores específicos de X . Por outro lado, tem a desvantagem de que se X apresenta alta dimensão, então o número de matches pode ser pequeno. Nesses casos, de acordo com a metodologia, o *matching* baseado em uma métrica de valor escalar torna-se mais atrativo.

Por outro lado, o método de estratificação ou *matching* intervalo baseia-se na ideia de dividir o intervalo de variação do *propensity score* em intervalos, de forma que, em cada intervalo da unidade de tratamento e controle há, em média, o mesmo *propensity score*.

Podem-se usar os mesmos blocos identificados pelo algoritmo utilizado para calcular os *propensity score*. Em seguida, calcula-se a diferença entre os resultados médios do grupo tratados e de controle. O ATET, na realidade é a média ponderada destas diferenças, com pesos sendo determinados pela distribuição das unidades tratadas através dos blocos. Uma das desvantagens deste método é que ele descarta observações nos blocos nas quais unidades tratadas ou de controle estão ausentes.

$$ATE_T^S = (N_b^T)^{-1} \sum_{i \in I(b)} Y_{1,i} - (N_b^C)^{-1} \sum_{j \in I(b)} Y_{0,j}. \quad (4.8)$$

Em que $I(b)$ é o conjunto de unidades no bloco b , N_b^T é o número de unidades tratadas no b th bloco. Assim, portanto, o efeito tratamento baseado na estratificação é definido como:

$$ATE^S = \sum_{b=1}^B ATE_T^S \times \left[\frac{\sum_{i \in I(b)} D_i}{\sum D_i} \right]. \quad (4.9)$$

Onde o termo entre parentes representa o peso de cada bloco dado à fração correspondente de cada unidade tratada e B é o número total de blocos.

Por sua vez, no *matching* raio o conjunto $A_i(p(X)) = \{P_j | \min_j \|P_i - P_j\| < r\}$, é baseado no *propensity score*. Isto diz que todos os casos pertencentes ao grupo de controle estimados através do *propensity score*, que pertençam ao raio r , é *matching* do i th caso tratado. Adicionalmente, pode-se expressar ATE e ATET em termos em $p(X)$, assumindo a condição de sobreposição $0 < p(X) < 1$. Os dois principais resultados são:

$$ATE = E \left[\frac{(D - p(X))y}{p(x)(1 - p(X))} \right]. \quad (4.10)$$

$$ATE_T = E \left[\frac{(D - p(X))y}{\Pr[D = 1](1 - p(X))} \right]. \quad (4.11)$$

Como relatado anteriormente, a fim de superar possível endogeneidade devido à omissão de variáveis que captem características, como, habilidade cognitiva geral, herdabilidade genética, fatores ambientais compartilhados, como também, outros aspectos não observáveis ou não disponíveis, adota-se a seguir uma proposta similar à desenvolvida por Ashenfelter e Rouse (1999). Na realidade, faz-me um *matching* – pareamento – simples. Em outras palavras, coparam-se os resultados obtidos entre os próprios irmãos gêmeos. Uma vez que, a estratégia anteriormente empregada, possibilita o uso de *matches* – pareamento de não gêmeos – distintos devido à variável de tratamento não necessariamente ser ponto de diferença entre os gêmeos. Por exemplo, é perfeitamente possível que ambos os gêmeos sejam

sindicalizados ou não. Ratificando tais condições, os próprios dados já justificam tal preocupação, pois apenas 23,53% dos indivíduos da amostra são sindicalizados.

Como visto, a principal finalidade dessa segunda abordagem é justamente levar em conta fatores que limitam significativamente as estimações da equação (4.1). Pois, ao comparar – *matching* simples – irmãos gêmeos monozigóticos entre si controlam-se de imediato as características não observáveis citadas. Adicionalmente, ao se fazer isso, torna os resultados quanto à participação da sorte opção nos resultados individuais bem mais robustos.

Nesse sentido, apresenta-se a seguir uma breve ilustração da estratégia adotada. Assim sendo, admita que o *matching* recorra ao melhor controle de pareamento entre o grupo de tratados e não tratados, e, portanto, utilizando como unidade de comparação o próprio irmão gêmeo. Para tanto, utilizamos as diferenças existentes entre as observações dos mesmos com o propósito de controlar possíveis efeitos fixos existentes.

Nessa perspectiva, admita que os resultados individuais alcançados – salário hora – expressos por y_{1i} e y_{2i} por cada par de gêmeos sejam respectivamente.

$$y_{1i} = X_i^1 \beta^1 + C_i + u_i. \quad (4.12)$$

$$y_{2i} = X_i^2 \beta^2 + C_i + u_i. \quad (4.13)$$

Onde X_i^n denota as variáveis observáveis de cada gêmeo. Como por exemplo, educação individual e parental, idade, sexo, estado civil, sindicalização, etc.; C_i representa os fatores fixos não observáveis, aqui assumidos como, por exemplo: habilidade cognitiva geral, herdabilidade genética, fatores ambientais compartilhados, como também, outras particularidades individuais não observáveis ou disponíveis. E u_i o termo de erro. Nesse sentido, se tomarmos a diferença entre as variáveis contidas em (4.12) e (4.13) obtém-se:

$$y_{1i} - y_{2i} = (X_i^1 \beta^1 - X_i^2 \beta^2) + (C_i - C_i) + u_i. \quad (4.14)$$

$$dy_i = dX_i \theta + \varepsilon_i. \quad (4.15)$$

Como assumimos que os fatores contidos em C_i são semelhantes, ou seja, que não há diferenças significativas em habilidade cognitiva geral, herdabilidade genética, ambiente compartilhado, entre outros fatores entre os irmãos gêmeos. Logo, $(C_i - C_i) = 0$. Isto é, anulam-se os efeitos fixos não diretamente observáveis. Diante disso, os resultados propostos em (4.15) apresentam-se como o mais apropriado aos objetivos propostos. Posto isso, faz-se a seguir a descrição e análise parcial dos dados.

4.4 Descrição e Análise dos Dados

Os resultados reportados neste trabalho foram extraídos de uma amostra de 680 gêmeos monozigóticos – idênticos – coletados no Festival do Dia Anual dos Gêmeos em Twinsburg, Ohio, ocorridos no período de 1991 a 1994. E disponibilizados no *homepage* da Universidade de Princeton. Os dados foram originalmente utilizados no trabalho de Ashenfelter e Krueger (1994).

Todavia, conforme relatado anteriormente, o fator crucial para adoção dessa base de dados é a possibilidade de controlar os fatores não observáveis até então ignorados, em especial, em aplicações empíricas. Além disso, destaca-se que as variáveis utilizadas nesta pesquisa são descritas no quadro 4.1, com especial atenção para dummy de filiação sindical.

Entretanto, cabe ressaltar que foram realizados dois testes de especificação do modelo com a finalidade de elencar o melhor grupo de covariadas.²⁴ Os dois testes verificam se há problemas de variáveis omitidas.²⁵ Além da observância dos resultados dos testes, procurou-se optar também por variáveis clássicas utilizadas na literatura.²⁶ Guiado por isto, regrediu-se o log do salário individual contra o conjunto de covariadas especificadas no quadro 4.1.

Quadro 4.1 – Conjunto de Informações

Variáveis	Descrição
<i>Salário hora</i>	Log do salário hora individual.
<i>Nível Educacional</i>	Anos de estudo individual.
<i>Educação Parental</i>	Anos de estudo da mãe e do pai.
<i>Idade</i>	Idade individual.
<i>Sexo</i>	Dummy para sexo, que assume valor 0 para mulheres e 1 para os homens.
<i>Raça</i>	Dummy para raça, que assume valor 0 para negros e 1 brancos.
<i>Autônomo</i>	Dummy para trabalho, que assume valor 0 para empregado e 1 autônomo.
<i>Sindicalizado</i>	Dummy para sindicalizado, que assume valor 1 para sindicalizados e 0 caso contrário.
<i>Casado</i>	Dummy para união, que assume valor 1 para casados e 0 caso contrário.
<i>Irmãos</i>	Números de irmãos.
<i>Diferenças</i>	Diferença entre as variáveis anteriormente descritas.

Fonte: Elaboração própria.

Posto isso, faz-se agora um breve relato a respeito do conjunto de dados. A tabela 4.2 sumariza alguns predicados da base. Por exemplo, observa-se que aproximadamente 60% dos

²⁴ Os resultados dos testes de especificação – *ovtest* e *linktest* – não rejeitaram a hipótese nula.

²⁵ Maiores detalhes em Chatterjee e Hadi (2013).

²⁶ Por exemplo, Ashenfelter e Krueger (1994).

gêmeos são do sexo feminino; onde somente 8% são declarados negros; 11,02% trabalham por conta própria; 89% são trabalhadores formais, e destes, 21,91% são sindicalizados.

Tabela 4.1 – Testes de Especificação do Modelo

Descrição dos Testes	Especificação dos Testes	
	Função do Teste	P-valor
Linktest	H0: O modelo não tem variáveis omitidas	0.9550
Ovtest		0.9785

Fonte: Elaboração própria, a partir dos dados.

Adicionalmente, constata-se que a 24% a 30,5%, respectivamente, mãe e pai, possuem no mínimo nível superior. Adicionalmente, verifica-se que em torno de 64% dos gêmeos possui nível superior. Esse último resultado denota teoricamente uma elevada mobilidade educacional entre as gerações.

Tabela 4.2 – Análise Descritiva

Características Individuais					
Sexo		Raça		Autônomo	Sindicalizados
Masculino 40,29%	Feminino 59,71%	Branco 92,21	Negro 7,79%	13,68%	23,53%
Nível Educacional					
Própria		Pai		Mãe	
Médio 36,48%	Superior 63,52%	Médio 69,56%	Superior 30,44%	Médio 75,59%	Superior 24,41%

Fonte: Elaboração própria, a partir de dados.

Posto isto, a seguir apresenta-se os principais resultados encontrados.

4.5 PRINCIPAIS RESULTADOS

Esta seção designa-se a apresentar e discutir os principais resultados desta pesquisa. Primeiramente, faz-se uma análise das diferenças nos resultados observados – salário hora – dos gêmeos e sua opção de escolha quanto a ser sindicalizado ou não.

Inicialmente, a tabela 4.3 reporta os resultados do log do salário hora – variável dependente – contra a dummy de filiação sindical – tratamento – que assume valor um caso o

trabalhador seja sindicalizado e zero caso contrário. Os resultados sugerem uma diferença salarial média de 28% entre os grupos analisados. Em suma, ser sindicalizado impacta em média 0,28 vezes a mais no resultado econômico individual do que em indivíduos que não são sindicalizados. Contudo, tais resultados não devem ser considerados como verdadeiros, uma vez que, não foi empregada nenhuma forma – variável – de controle. Os resultados foram obtidos usando o método de *bootstrap* com 1.000 replicações.

Tabela 4.3 – Diferença nas Médias

Log do Salário	
Sindicalizado	0.2832*** (0.0456)
Intercepto	2.3769*** (0.0286)
Observações	680

Fonte: Elaboração própria. Desvios Padrão (*).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Por outro lado, a tabela 4.4 relata a distribuição amostral perante a participação na variável de interesse – Sindicalização – adotada nesta pesquisa. A mesma informa que 76,47% dos indivíduos avaliados não são sindicalizados, ou seja, pertencem ao grupo de controle, contra 23,53% que formam o grupo de tratamento. Destaca-se que a propriedade de balanceamento, isto é, a hipótese 4.2 é satisfeita em todas as abordagens empregadas.

Tabela 4.4 – Distribuição do Propensity Score

Grupo	Frequência	%
Controle – Não sindicalizado	520	76,47
Tratamento – Sindicalizado	160	23,53
Observações	680	100

Fonte: Elaboração própria.

Verifica-se também, a partir da tabela 4.5, que em todos os casos, o escore de propensão foi restrito à região de suporte comum. Outra característica que merece destaque é a representatividade do grupo de tratamento perante o grupo de controle nos blocos estimados, com exceção do último bloco.

Posto isto, estima-se a probabilidade de ser sindicalizado, tanto para os participantes, quanto para os não participantes da entidade sindical. As variáveis de controle elegidas para estimar o *PSM* são as mesmas empregadas para delinear os grupos analisados. A tabela 4.6 reporta os fatores que explicam a sindicalização.

Tabela 4.5 – Impacto da Sindicalização

Mínimo $\hat{P}(x)$	Não Tratado	Tratado	Total
0.0	302	61	363
0.1	86	20	106
0.2	59	21	80
0.3	36	15	51
0.4	20	16	36
0.5	10	14	24
0.6	06	08	14
0.7	01	03	04
0.8	00	02	02
Total	520	160	680

Fonte: Elaboração própria.

As variáveis as quais os indivíduos não possuem qualquer responsabilidade individual, como idade, raça, gênero, educação parental e número de irmãos foram incluídas em função de representar o background social – sorte social – amplamente difundida na literatura como fonte de desigualdade injusta. Por outro lado, a adoção da idade ao quadrado busca captar uma tendência negativa nos salários à medida que os indivíduos envelhecem. Quanto a outras variáveis, por exemplo, se os indivíduos são trabalhadores autônomos, nível de estabilidade e se os mesmos são casados representam fatores referentes ao esforço individual.

Tabela 4.6 – Propensity Score Matching – Sindicalização

Descrição	B _{PMS}	σ	t	p-value
Idade	0.273**	0.084	3.22	0.001
Idade ²	-0.003**	0.000	-3.37	0.001
Branco	-1.249**	0.398	-3.14	0.002
Feminino	-0.478	0.268	-1.78	0.074
Autônomos	-2.252**	0.649	-3.47	0.001
Casados	0.333	0.328	1.02	0.310
Estabilidade	0.051**	0.017	2.96	0.003
Nº Irmãos	-0.006	0.057	-0.11	0.914
Educ. Pai	-0.123**	0.052	-2.34	0.019
Educ. Mãe	-0.032	0.065	-0.50	0.617
Intercepto	-3.872**	1.860	-2.08	0.037

Fonte: Elaboração própria.

* p<0.10; ** p<0.05; *** p<0.01

Adicionalmente, destaca-se que a variável dependente – salário hora – empregada na análise, capta características relevantes do esforço individual empregado. Os resultados apresentados informam, por exemplo, que a idade influencia em 0.27 na probabilidade de sindicalizar tanto para o grupo de controle quanto para o grupo de tratamento. Entretanto, um resultado que chama a atenção é o estado civil – casado – do indivíduo. Pois, realmente parece ser bem intuitivo que indivíduos com essas características busquem uma maior

segurança jurídica. Ou seja, o simples fato de ser casado aumenta a probabilidade de possuir filiação sindical. Entretanto, embora o coeficiente estimado seja positivo, o mesmo não é estatisticamente diferente de zero.

Por sua vez, os resultados reportados na tabela 4.7 revelam que o efeito tratamento sobre o tratado – filiação sindical – referente ao salário hora é em torno 20% maior comparado ao grupo de controle. Em suma, a grandeza econômica do tratamento é bastante expressiva. Sendo mais específico, a decisão de se sindicalizar aumenta em 20% a probabilidade de obter um salário hora 28% maior dos que os indivíduos não sindicalizados. Contudo, o efeito tratamento médio informa um impacto positivo de 0.135 na hora salarial dos tratados.

Outra questão a destacar, é que a influência apurada apresenta-se estatisticamente significativa em todos os modelos adotados. Especificamente, o modelo de distância com $r = 0.01$ e $r = 0.05$ são os que apresentam os coeficientes mais expressivos. Esses resultados parecem evidenciar o principal objetivo desta pesquisa, isto é, permitimos inferir sobre a existência de fatores – *sorte opção* – determinísticos contidos no termo de erro, até então, considerados como aleatórios, ao menos que, empiricamente.

Tabela 4.7 – Efeito Tratamento Sobre o Tratado – ATT – Sorte Opção

Matching	Tratados	Não Tratados	ATT	ATE	σ	t
<i>PSmatching logit</i>	103	335	0.182***	0.135***	0.069	4.230
	103	335	(0.182)***	-	-	-
<i>Vizinhança</i>	160	256	0.191***	-	0.086	2.234
	160	256	(0.191)***	-	0.080	2.157
<i>Radius $r = 0.01$</i>	97	286	0.262***	-	0.061	4.299
	97	270	(0.248)***	-	0.068	3.646
<i>Radius $r = 0.05$</i>	101	335	0.247***	-	0.059	4.218
	101	301	(0.242)***	-	0.058	4.159
<i>Estratificação</i>	101	303	0.198***	-	0.059	3.382
	101	303	(0.198)***	-	0.061	3.255
<i>Kernel</i>	160	520	0.201***	-	-	-
	103	301	(0.199)***	-	0.061	3.244

Fonte: Elaboração própria. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Valores em parênteses obtidos por bootstrap para 1000 replicações.

Contudo, embora, as estimações apresentem uma convergência em tornos dos valores relatados, com exceção, ao método da distância através de um raio. Os coeficientes encontrados neste último modelo estão em torno de 25% mais significativos. Apresenta-se a seguir os resultados da segunda estratégia empírica, a qual busca tratar a omissão das características individuais não observáveis. Isto é, essa estratégia objetiva suplantando importantes limitações empíricas e estruturais presentes na equação (4.1).

Nessa perspectiva, estimamos o modelo expresso pela equação (4.12) para o mesmo grupo de covariadas. Como esperado o efeito médio da variável de tratamento – sindicalização – sobre o salário hora se reduziu significativamente, uma vez que, os fatores omitidos superestimam os coeficientes obtidos na tabela 4.7.

Tabela 4.8 – Influência da Sorte Opção no Salário Hora

Diferenças	β_{OLS}	β_{1dif}	β_{1difIV}	β_{1dEF}
Tratamento	0.087**	0.091*	0.091*	0.091*
Educação	0.079***	0.081***	0.081*	0.081***
Casado	0.034	0.024	0.024	0.024
Estabilidade	0.023***	0.023***	0.023***	0.023***
Autônomos	-0.007	-	-	-
Intercepto	0.006	-	-	-
Observações	680	680	680	680

Fonte: Elaboração própria. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Os resultados reportados na tabela 4.8 ratificam as expectativas. Pois, de acordo com os números apresentados, nos quatro modelos estimados,²⁷ a influência média dessa componente da sorte, situa-se próximo a 0.09. Em suma, a opção de sindicalizar, controlado por fatores observáveis e não observáveis, responde em média por 9% das chances do indivíduo situar-se na distribuição salarial superior. Em linhas gerais, a sorte opção representada aqui pela filiação sindical, faz com que o indivíduo tenha 0.09 maiores oportunidades de estar entre os que têm um salário hora em média 28% superior.

4.6 ALGUMAS CONSIDERAÇÕES

Nesse trabalho, o problema principal era investigar se a denominada sorte opção teria participação determinística no resultado – salário hora – individual obtido, como também, superar algumas importantes limitações empíricas e estruturais na equação fundamental dos atuais modelos de igualdade de oportunidades. Para tanto, regrediu-se a equação (4.1) através do emprego da equação (4.12). Em palavras, controlou-se a influência de determinantes não observáveis ou não disponíveis na maioria das bases de dados nos resultados – salário hora – individuais. Em linhas gerais, buscou-se avançar estruturalmente e empiricamente a teoria de

²⁷ β_{OLS} – mínimos quadrados ordinários; β_{1dif} – primeiras diferenças; β_{1dEF} – primeiras diferenças variáveis instrumentais e β_{1d} – primeiras diferenças efeitos fixos.

igualdade de oportunidades. Pois, fez-se uma aplicação que parece ser mais realista e procura atender tanto a teoria filosófica quanto econômica. Posto isso, primeiramente, empregou-se o método *PSM* em cinco abordagens distintas. Adicionalmente, através do emprego de um enfoque alternativo consideraram-se características individuais não observáveis, como, habilidade cognitiva geral, herdabilidade genética, ambiente compartilhado, entre outros fatores. Para isso, utilizou-se uma amostra composta por 680 gêmeos monozigóticos.

Os resultados indicaram que a variável aqui assumida como *sorte opção* – filiação sindical – possui influência substancial no resultado salarial da amostra analisada. Na primeira abordagem, sem considerar os aspectos não observáveis, como alguns fatores observáveis, embora não disponíveis, os valores informam que o grupo de tratamento tem em torno de 20% maiores chances comparado aos indivíduos que compõem o grupo de controle de se localizar na distribuição que recebe em média 28% a mais. Por sua vez, na segunda abordagem, considerando agora, as limitações enumeradas em diversas oportunidades anteriores, essa influência declina para algo em torno de 9%.

Contudo, mesmo assim, a amplitude econômica desse efeito é sem dúvida muito significativa. Sintetizando, os principais objetivos traçados foram alcançados. Isto é, encontraram-se evidências significativas na amostra analisada que parte da sorte, até então desconsideradas nos modelos de IOP's, tem participação significativa no resultado individual. Em outras palavras, o termo de erro dos atuais modelos de IOPs não é formado por componentes totalmente aleatórios. Por outro lado, o uso da base de dados associado ao procedimento de *matching* simples – pareamento de irmãos gêmeos – composto por gêmeos monozigóticos propiciou um controle imediato de diversos fatores não observáveis ou não disponíveis responsáveis pelo desempenho individual.

Espera-se que futuras pesquisas sobre o tema em questão, lance luz sobre a participação – quantitativa – das características não observáveis. Por exemplo, o quanto características como habilidade cognitiva geral, herdabilidade genética, ambiente compartilhado respondem pelo resultado individual. Como também, adotar abordagens que busquem decompor o termo de erro, particularmente, sobre os efeitos das componentes do fator sorte.

5 CONCLUSÃO

Este estudo propôs uma investigação sobre a moderna teoria de desigualdade de oportunidades, particularmente, centrada na disparidade de oportunidades educacionais e de renda. Foram consideradas dois tipos de variáveis de resultados, isto é, no primeiro e segundo ensaio, utilizaram-se informações de desempenho de testes educacionais. No terceiro ensaio, adotou-se o nível de renda. Especificamente, os dados de desigualdade de oportunidades educacionais foram obtidos na base do PISA versão 2012 e do ENEM e SAEB 2010. Por outro lado, no último ensaio, empregou uma base composta por um conjunto de informações socioeconômicas individuais referentes a 680 gêmeos monozigóticos. Enfatiza-se que essa última base foi originalmente utilizada no trabalho Ashenfelter e Krueger (1994).

Os principais resultados apontaram essencialmente para quatro inferências categóricas: i) o elevado *gap* observado da transmissão intergeracional da educação, designadamente, devido à baixa transmissão educacional apurada nos países sul-americanos. A razão mais provável para isso seria devido à baixa qualidade educacional ofertada nesses países. Essa suposição é respaldada pela baixa nota média no teste PISA observada nessas economias; ii) efeito expressivo das circunstâncias individuais que o indivíduo enfrenta sobre o esforço despendido. Nesse caso, parece ser o provável resultado esperado, uma vez que, os indivíduos geralmente reagem positivamente a incentivos; iii) efeito positivo de parte da sorte – opção – no resultado salarial obtido. Quanto a esse resultado, pode-se assumir ao menos uma coisa, isto é, os principais resultados encontrados na moderna literatura de igualdade de oportunidades parecem conter graves limitações; e por fim, e considerado o mais importante resultado alcançado, iv) fatores até então desconsiderados como habilidade cognitiva geral, herdabilidade genética, ambiente compartilhado, entre outros, apresentaram-se decisivos na obtenção dos resultados individuais. Nitidamente, estes últimos resultados ratificam a consideração anterior e acentuam a existência de importantes limitações, em especial, empíricas na predominante teoria de igualdade de oportunidades. Em outros moldes, pode-se afirmar que as deduções feitas a partir de análises que não consideram a existência dessas características individuais não observáveis são ao que parecem frágeis e fortemente limitadas.

Contudo, é importante ressaltar alguns percalços enfrentados nesse estudo. Isto é, ao utilizar dados de testes que não penalizam estudantes e escolas envolvidas, como é o caso do PISA, faz-se necessário assumir, por exemplo, a hipótese na qual os indivíduos estejam sendo subestimados em virtude da ausência de incentivos que façam com que os mesmos resolvam

adequadamente ao teste. Todavia, quando se faz o uso dos dados do ENEM, a qual descarta tal possibilidade, em decorrência da sua utilização no ingresso da universidade pública no Brasil, ao que parece os resultados convergem para os mesmos problemas. Embora, para confirmar esse diagnóstico dever-se-ia empregar os mesmos mecanismos adotados no primeiro ensaio.

Por outro lado, no terceiro ensaio, pode-se destacar como principal limitação o uso do *PSM* para uma amostra tão pequena. Essa limitação é identificada especialmente pela baixa área de sobreposição do grupo de tratados e não tratados, visualizada a partir do gráfico B.1 do apêndice B. Entretanto, esclarece-se que o uso desse instrumental serviu exatamente para mostrar a restrição de qualquer método que desconsidera fatores essenciais nas suas análises. Especificamente, no caso em questão, corroborada pelas características individuais não observáveis.

Nessa configuração, acredita-se haver uma grande lacuna na atual teoria de desigualdade de oportunidades. Nesse sentido, ao longo do estudo, surgiram diversas oportunidades de trabalhos futuros, como por exemplo, mensurar a magnitude da influência de fatores não observáveis sobre o resultado econômico individual, seja em nível de desempenho educacional ou nível de renda obtida. Por outro lado, pode-se empregar uma maior amostra de gêmeos monozigóticos, e, portanto, mais rica de informações, de modo que, possibilite encontrar resultados mais informativos.

REFERÊNCIAS

AGUIRRECHE, Agurtzane Lecuona. **Inequality of opportunity in education**. 2012.

ALBERNAZ, Angela; FERREIRA, Francisco HG; FRANCO, Creso. **Qualidade e Equidade na Educação Fundamental Brasileiro**. Pesquisa e Planejamento Econômico, v. 33, n. 3, 2002.

ASHENFELTER, Orley C.; KRUEGER, Alan B. **Estimates of the Economic Returns to Schooling from a New Sample of Twins**. American Economic Review, v. 84, n. 5, p. 1157-73, 1994.

ASHENFELTER, Orley; ROUSE, Cecilia. **Income, schooling, and ability: Evidence from a new sample of identical twins**. The Quarterly Journal of Economics, v. 113, n. 1, p. 253-284, 1998.

ASHENFELTER, Orley; ROUSE, Cecilia. **Income, schooling, and ability: Evidence from a new sample of identical twins**. National Bureau of Economic Research, 1997.

AZAM, Mehtabul; BHATT, Vipul. **Like father, like son? Intergenerational education mobility in India**. Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit, 2012.

DE BARROS, Ricardo Paes. **Measuring inequality of opportunities in Latin America and the Caribbean**. World Bank Publications, 2009.

BAUER, Philipp; RIPHAHN, Regina T. **Heterogeneity in the intergenerational transmission of educational attainment: evidence from Switzerland on natives and second-generation immigrants**. Journal of Population Economics, v. 20, n. 1, p. 121-148, 2007.

BETTS, Julian; ROEMER, John E. **Equalizing opportunity for racial and socioeconomic groups in the United States through educational finance reform**. Department of Economics, UCSD, 2005.

BJÖRKLUND, Anders; LINDAHL, Mikael; PLUG, Erik. **The origins of intergenerational associations: Lessons from Swedish adoption data**. The Quarterly Journal of Economics, v. 121, n. 3, p. 999-1028, 2006.

BOURGUIGNON, François; FERREIRA, Francisco HG; MENENDEZ, Marta. **Inequality of opportunity in Brazil**. Review of income and Wealth, v. 53, n. 4, p. 585-618, 2007.

CAMERON, A. Colin; TRIVEDI, Pravin K. **Microeconometrics: methods and applications**. Cambridge university press, 2005.

CARD, David. **Estimating the return to schooling: Progress on some persistent econometric problems**. Econometrica, v. 69, n. 5, p. 1127-1160, 2001.

CARROLL, John B. **Psychometrics, intelligence, and public perception**. Intelligence, v. 24, n. 1, p. 25-52, 1997.

CHATTERJEE, Samprit; HADI, Ali S. **Regression analysis by example**. John Wiley & Sons, 2013.

CHECCHI, Daniele; PERAGINE, Vito; SERLENGA, Laura. **Fair and unfair income inequalities in Europe**. Discussion paper series//Forschungsinstitut zur Zukunft der Arbeit, 2010.

CHERNOZHUKOV, Victor; FERNÁNDEZ-VAL, Iván; MELLY, Blaise. **Inference on counterfactual distributions**. Econometrica, v. 81, n. 6, p. 2205-2268, 2013.

CHIPUER, Heather M.; ROVINE, Michael J.; PLOMIN, Robert. **LISREL modeling: Genetic and environmental influences on IQ revisited**. Intelligence, v. 14, n. 1, p. 11-29, 1990.

DAUDE, Christian. **Ascendance by Descendants? On Intergenerational Education Mobility in Latin America**. OECD Publishing, 2011.

DE BARROS, Ricardo Paes. **Measuring inequality of opportunities in Latin America and the Caribbean**. World Bank Publications, 2009.

DIAZ, Maria Dolores Montoya. (Des) **Igualdades de Oportunidades no Ensino Médio Brasileiro: Escolas Públicas e Privadas**. Economia, v. 13, n. 3a, 2012.

DWORKIN, Ronald. **What is equality? Part 2: Equality of resources**. Philosophy & Public Affairs, p. 283-345, 1981.

FARRÉ, Lúdia; KLEIN, Roger; VELLA, Francis. **A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY**. *Empirical Economics*, v. 44, n. 1, p. 111-133, 2013.

FERREIRA, F.; GIGNOUX, Jérémie. **Towards an Understanding of Socially Inherited Inequalities in Educational Achievement: Evidence from Latin America and the OECD**. Washington, DC, World Bank, en< [www. worldbank. org/lacopportunity](http://www.worldbank.org/lacopportunity), 2008.

FERREIRA, Francisco HG; GIGNOUX, Jérémie. **The measurement of inequality of opportunity: Theory and an application to Latin America**. *Review of Income and Wealth*, v. 57, n. 4, p. 622-657, 2011.

FERREIRA, Sérgio; VELOSO, Fernando A. **Mobilidade intergeracional de educação no Brasil**. *Pesquisa e Planejamento Econômico*, v. 33, n. 3, p. 481-513, 2003.

FIRMO, Marcio Gold; SOARES, R. **Uma análise da transmissão intergeracional de capital humano no Brasil**. 2008.

FLEURBAEY, Marc; SCHOKKAERT, Erik. **Unfair inequalities in health and health care**. *Journal of health economics*, v. 28, n. 1, p. 73-90, 2009.

FOGUEL, Miguel N.; VELOSO, Fernando A. **Inequality of opportunity in daycare and preschool services in Brazil**. *The Journal of Economic Inequality*, v. 12, n. 2, p. 191-220, 2014.

GAMBOA, Luis Fernando; WALTENBERG, Fábio D. **Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006–2009**. *Economics of Education Review*, v. 31, n. 5, p. 694-708, 2012.

GRAWE, Nathan D.; MULLIGAN, Casey B. **Economic interpretations of intergenerational correlations**. National bureau of economic research, 2002.

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

HANUSHEK, Eric. **The Production of Education**, Teacher Quality and Efficiency. 1970.

HANUSHEK, Eric A.; WÖßMANN, Ludger. **The role of school improvement in economic development**. National Bureau of Economic Research, 2007.

HERSKIND, Anne Maria et al. **The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900**. *Human genetics*, v. 97, n. 3, p. 319-323, 1996.

HERTZ, Tom et al. **The inheritance of educational inequality: International comparisons and fifty-year trends**. *The BE Journal of Economic Analysis & Policy*, v. 7, n. 2, 2007.

KLEIN, Roger; VELLA, Francis. (2009) **Estimating the return to endogenous schooling decisions via conditional second moments**. *Journal of Human Resources*, v. 44, n. 4, p. 1047-1065.

KLEIN, Roger; VELLA, Francis. (2010) **Estimating a class of triangular simultaneous equations models without exclusion restrictions**. *Journal of Econometrics*, v. 154, n. 2, p. 154-164.

KOENKER, Roger. **Quantile regression**. Cambridge university press, 2005.

LEFRANC, Arnaud; PISTOLESI, Nicolas; TRANNOY, Alain. **Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France**. *Journal of Public Economics*, v. 93, n. 11, p. 1189-1207, 2009.

O'NEILL, B., SWEETMAN, D. VAN DE GAER, D. (2001). **Equality of opportunity and kernel density estimation: an application to intergenerational mobility**. In: Fomby, T. & Hill, C. (eds.). *Applying kernel and nonparametric estimation to economic topics*. *Advances in Econometrics*, vol. 14, Stanford, Conn: JAI Press.

PLOMIN, Robert; SPINATH, Frank M. **Intelligence: genetics, genes, and genomics**. *Journal of personality and social psychology*, v. 86, n. 1, p. 112, 2004.

QUINTANO, Claudio; CASTELLANO, Rosalia; LONGOBARDI, Sergio. **A fuzzy clustering approach to improve the accuracy of Italian student data. An experimental procedure to correct the impact of outliers on assessment test scores**. *Statistica & Applicazioni*, v. 7, n. 2, p. 149-171, 2009.

RAWLS, John. **A Theory of Justice**. Cambridge. Mass.: Harvard University, 1971.

ROEMER, John E. **Equality of opportunity**. Cambridge: Harvard University Press, 1998.

RONALD, Angelica; SPINATH, Frank M.; PLOMIN, Robert. **The aetiology of high cognitive ability in early childhood**. *High Ability Studies*, v. 13, n. 2, p. 103-114, 2002.

ROSENBAUM, Paul R.; RUBIN, Donald B. **The central role of the propensity score in observational studies for causal effects.** Biometrika, v. 70, n. 1, p. 41-55, 1983.

SOLON, Gary. **Intergenerational mobility in the labor market.** Handbook of labor economics, v. 3, p. 1761-1800, 1999.

SPEARMAN, Charles. **"General Intelligence," Objectively Determined and Measured.** The American Journal of Psychology, v. 15, n. 2, p. 201-292, 1904.

SPINATH, Frank M. et al. **Phenotypic early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2–4 years.** Intelligence, v. 31, n. 2, p. 195-210, 2003.

TROUTON, Alexandra; SPINATH, Frank M.; PLOMIN, Robert. **Twins early development study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood.** Twin Research, v. 5, n. 05, p. 444-448, 2002.

UNESCO. **Relatório da Organização das Nações Unidas para a Educação, a Ciência e a Cultura.** 2005.

VALLENTYNE, Peter. **Brute Luck, Option Luck, and Equality of Initial Opportunities.** Ethics, v. 112, n. 3, p. 529-557, 2002.

VAN DE GAER, Dirk; MARTINEZ, Michel; SCHOKKAERT, Erik. **Measuring intergenerational mobility and equality of opportunity.** Katholieke Universiteit Leuven, Centrum voor Economische Studiën, Working Group Public Economics, 1998.

WOOLDRIDGE, Jeffrey M. **Econometric analysis of cross section and panel data.** MIT press, 2010.

APÊNDICE A: Resultados Adicionais

Tabela A.1 - Determinantes do Resultado Individual – Matemática

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1517*** (0.0011)	0.1691*** (0.0014)	0.1691*** (0.0009)
Dif. Educação	-0.0023*** (0.0002)	-0.0023*** (0.0003)	-0.0023*** (0.0003)
Tipo Escola	0.0486*** (0.0008)	0.0512*** (0.0011)	0.0512*** (0.0009)
Sexo	0.0244*** (0.0007)	0.0275*** (0.0009)	0.0275*** (0.0006)
Pré-escolar	0.1172*** (0.0011)	0.1230*** (0.0014)	0.1230*** (0.0015)
Local Escola	0.0431*** (0.0012)	0.0381*** (0.0016)	0.0381*** (0.0017)
Estrutura Familiar	0.0595*** (0.0010)	0.0607*** (0.0013)	0.0607*** (0.0011)
Irmãos	0.0231*** (0.0010)	0.0248*** (0.0012)	0.0248*** (0.0011)
Intercepto	5.5491*** (0.0030)	5.5138*** (0.0038)	5.5138*** (0.0027)
N	336286	336286	336286
adj. R ²	0.138		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * p<0.10, ** p<0.05, *** p<0.01.

Tabela A.2 - Determinantes do Resultado Individual – Língua

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1380*** (0.0011)	0.1496*** (0.0012)	0.1496*** (0.0012)
Dif. Educação	-0.0021*** (0.0002)	-0.0018*** (0.0002)	-0.0018*** (0.0003)
Tipo Escola	0.0480*** (0.0008)	0.0450*** (0.0009)	0.0450*** (0.0008)
Sexo	-0.0778*** (0.0007)	-0.0688*** (0.0007)	-0.0688*** (0.0009)
Pré-escolar	0.1190*** (0.0011)	0.1174*** (0.0012)	0.1174*** (0.0009)
Local Escola	0.0526*** (0.0012)	0.0500*** (0.0013)	0.0500*** (0.0016)
Estrutura Familiar	0.0509*** (0.0010)	0.0492*** (0.0011)	0.0492*** (0.0012)
Irmãos	0.0198*** (0.0010)	0.0230*** (0.0011)	0.0230*** (0.0011)
Intercepto	5.6355*** (0.0030)	5.6300*** (0.0033)	5.6300*** (0.0030)
N	336286	336286	336286
adj. R ²	0.153		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * p<0.10, ** p<0.05, *** p<0.01.

Tabela A.3 - Determinantes do Resultado Individual – Ciências

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1462*** (0.0010)	0.1604*** (0.0013)	0.1604*** (0.0011)
Dif. Educação	-0.0029*** (0.0002)	-0.0027*** (0.0003)	-0.0027*** (0.0003)
Tipo Escola	0.0445*** (0.0008)	0.0459*** (0.0011)	0.0459*** (0.0011)
Sexo	0.0015** (0.0007)	0.0055*** (0.0008)	0.0055*** (0.0009)
Pré-escolar	0.1108*** (0.0011)	0.1133*** (0.0014)	0.1133*** (0.0014)
Local Escola	0.0371*** (0.0012)	0.0342*** (0.0015)	0.0342*** (0.0015)
Estrutura Familiar	0.0531*** (0.0010)	0.0540*** (0.0012)	0.0540*** (0.0011)
Irmãos	0.0252*** (0.0009)	0.0288*** (0.0012)	0.0288*** (0.0011)
Intercepto	5.6040*** (0.0029)	5.5823*** (0.0037)	5.5823*** (0.0026)
<i>N</i>	336286	336286	336286
adj. <i>R</i> ²	0.127		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabela A.4 – Efeitos das Circunstâncias no PISA – Média Geral

Lugar	Educação Parental	Estrutura Familiar	Tipo de Escola	Atendimento Pré-escolar	Localização da Escola	Escolaridade Parental ($\bar{X}$)
América do Sul	0,0850833	-0,0143909	0,1413002	0,0548627	0,0998015	11,36677
Argentina	0,0856819	0,0194786	0,1424431	0,1254046	0,0243553	12,46967
Brasil	0,0596736	-0,0004727	0,1683148	0,0462792	0,0949779	10,10564
Chile	0,1981890	0,0281244	0,1009697	0,0313861	0,0883646	13,46962
Colômbia	0,0783930	0,0345007	0,0791773	0,0447753	0,0773341	11,42376
Peru	0,1128428	-0,008692*	0,0811457	0,0675148	0,1166911	11,45194
Uruguai	0,1106485	0,003729*	0,1398876	0,0778968	0,0338154	11,47351
Países da OCDE	0,187097	0,0180336	0,0380126	0,0524253	0,0327978	13,35801
Alemanha	0,2262079	0,0391014	0,0417868	0,0792220	0,0220426	14,99728
Canadá	0,1453420	0,0048673	0,0717503	0,0464819	0,0170054	15,00451
Estados Unidos	0,2008159	0,0558980	-0,014005*	0,008152*	0,0185935	13,67305
França	0,3098196	0,0241637	0,0238523	0,2058700	0,0241637	13,06789
Reino Unido	0,1920620	0,0319095	0,0613637	0,0778844	0,011594*	14,24681
Espanha	0,1331143	0,0229493	0,0332206	0,0688382	0,0247763	12,81644
Japão	0,2960487	0,0367766	-0,010416*	0,0638318	-	14,12523
México	0,0729893	0,0064641	0,0323874	0,0652580	0,0687198	11,14021
Países Asiáticos	0,0874985	0,0406226	0,000911*	0,1793669	0,0660958	12,06192
Shangai	0,1594430	0,0117946	0,0294805	0,1081157	-	12,72651
Singapura	0,2171583	0,0275654	-0,0652437	0,0951780	-	12,27486
Hong-Kong	0,1030831	0,0233843	-0,0561143	0,0979285	-	11,38295
Chinês Taipei	0,2957494	0,0384931	-0,0978440	0,0463951	-	13,01219
Coréia do Sul	0,2100192	0,0214467	0,0253231	0,016487*	0,0485138	13,99356
Países Nórdicos	0,1918707	0,0318528	0,0216254	0,0724996	0,0370241	14,76997
Suécia	0,1877121	0,0519021	0,0191649	0,0903231	0,0386639	14,09894
Dinamarca	0,2262079	0,0391014	0,0417868	0,0792220	0,0220426	14,99728
Finlândia	0,1933194	0,0199897	0,0201674	0,0931195	-0,005865*	14,86206

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

* Não significativos a 10%;

Tabela A.5 – Ranking pelo Índice de Correlação Intergeracional

Países	σ_{filhos} Variância	σ_{pais} Variância	β_1 (Persistência)	$\rho = \frac{\sigma_{\text{pais}}}{\sigma_{\text{filhos}}} \times \beta_1$	Ranking $\beta_1 \rho$
República Tcheca	0,17	0,15	0,42	0,36	1 3
República Eslovaca	0,22	0,18	0,40	0,34	2 5
Israel	0,22	0,21	0,35	0,34	3 6
Polônia	0,16	0,17	0,32	0,33	4 7
Rússia	0,17	0,12	0,31	0,22	5 32
França	0,20	0,20	0,31	0,32	6 10
Japão	0,18	0,14	0,30	0,23	7 31
Chinês Taipei	0,19	0,20	0,30	0,31	8 11
Hungria	0,18	0,24	0,29	0,38	9 2
Bulgária	0,23	0,26	0,28	0,33	10 8
Austrália	0,20	0,19	0,27	0,24	11 29
Áustria	0,18	0,20	0,26	0,29	12 13
Nova Zelândia	0,20	0,18	0,24	0,22	13 33
Eslovênia	0,19	0,19	0,24	0,25	14 35
Irlanda	0,17	0,20	0,24	0,28	15 27
Noruega	0,19	0,16	0,23	0,19	16 45
Alemanha	0,18	0,23	0,23	0,29	17 14
Dinamarca	0,18	0,23	0,23	0,29	18 15
Bélgica	0,20	0,24	0,22	0,26	19 24
Singapura	0,19	0,28	0,22	0,32	20 09
Montenegro	0,19	0,20	0,21	0,22	21 34
Coreia do Sul	0,17	0,18	0,21	0,23	22 31
Latvia	0,15	0,19	0,21	0,26	23 23
Grécia	0,19	0,25	0,20	0,27	24 21
Estados Unidos	0,18	0,25	0,20	0,28	25 18
Chile	0,18	0,34	0,20	0,38	26 1
Finlândia	0,19	0,24	0,19	0,25	27 26
Reino Unido	0,19	0,16	0,19	0,16	28 49
Suíça	0,17	0,26	0,19	0,29	29 16
Suécia	0,21	0,22	0,19	0,20	30 43
Lituânia	0,17	0,19	0,18	0,00	31 45
Holanda	0,18	0,23	0,18	0,22	32 33
Sérvia	0,19	0,22	0,18	0,21	33 41
Croácia	0,17	0,20	0,18	0,21	34 38
Luxemburgo	0,20	0,40	0,17	0,34	35 4
Islândia	0,20	0,17	0,16	0,14	36 54
Shangai	0,15	0,28	0,16	0,30	37 12
Estônia	0,14	0,15	0,16	0,16	38 51
Canadá	0,17	0,19	0,15	0,16	39 52
Tailândia	0,18	0,34	0,14	0,27	40 20
Romênia	0,17	0,22	0,14	0,18	41 47
Espanha	0,17	0,35	0,13	0,26	42 22
Itália	0,18	0,28	0,13	0,20	43 42

Portugal	0,19	0,44	0,12	0,28	44 17
Peru	0,21	0,45	0,11	0,25	45 28
Uruguai	0,21	0,39	0,11	0,21	46 39
Jordânia	0,21	0,27	0,11	0,14	47 40
Hong Kong	0,16	0,32	0,10	0,21	48 55
Liechtenstein	0,16	0,24	0,10	0,15	49 53
Turquia	0,17	0,45	0,09	0,25	50 25
Argentina	0,21	0,42	0,09	0,17	51 48
Tunísia	0,19	0,50	0,08	0,21	52 36
Colômbia	0,18	0,47	0,08	0,21	53 37
México	0,16	0,43	0,07	0,20	54 44
Qatar	0,26	0,30	0,06	0,07	55 58
Brasil	0,19	0,53	0,06	0,17	56 49
Macao-China	0,16	0,34	0,05	0,10	57 56
Indonésia	0,17	0,41	0,04	0,09	58 57

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Tabela A.6 - Determinantes do Resultado Individual – Matemática

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1430*** (0.0013)	0.1554*** (0.0016)	0.1554*** (0.0016)
Dif. Educação	-0.0011*** (0.0003)	-0.0009*** (0.0003)	-0.0009*** (0.0003)
Tipo de Escola	0.0519*** (0.0010)	0.0546*** (0.0013)	0.0546*** (0.0013)
Sexo	0.0362*** (0.0008)	0.0383*** (0.0010)	0.0383*** (0.0009)
Pré-escolar	0.1036*** (0.0013)	0.1090*** (0.0017)	0.1090*** (0.0018)
Local Escolar	0.0419*** (0.0014)	0.0379*** (0.0018)	0.0379*** (0.0019)
Migrante	0.0149*** (0.0012)	0.0161*** (0.0016)	0.0161*** (0.0020)
Repetente	0.1167*** (0.0011)	0.1220*** (0.0014)	0.1220*** (0.0019)
Perseverança	0.0172*** (0.0004)	0.0165*** (0.0005)	0.0165*** (0.0005)
Esforço Real	0.0032*** (0.0002)	0.0040*** (0.0003)	0.0040*** (0.0002)
Esforço Potencial	-0.0009*** (0.0001)	-0.0009*** (0.0001)	-0.0009*** (0.0001)
Intercepto	5.5103*** (0.0041)	5.4741*** (0.0053)	5.4741*** (0.0048)
N	225629	225629	225629
adj. R^2	0.188		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * p<0.10, ** p<0.05, *** p<0.01.

Tabela A.7 - Determinantes do Resultado Individual – Língua

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1286*** (0.0012)	0.1369*** (0.0015)	0.1369*** (0.0016)
Dif. Educação	-0.0009*** (0.0002)	-0.0010*** (0.0003)	-0.0010*** (0.0004)
Tipo de Escola	0.0509*** (0.0010)	0.0486*** (0.0011)	0.0486*** (0.0012)
Sexo	-0.0661*** (0.0008)	-0.0591*** (0.0009)	-0.0591*** (0.0011)
Pré-escolar	0.1023*** (0.0013)	0.1019*** (0.0015)	0.1019*** (0.0020)
Local Escolar	0.0521*** (0.0014)	0.0478*** (0.0016)	0.0478*** (0.0022)
Migrante	0.0214*** (0.0012)	0.0170*** (0.0014)	0.0170*** (0.0017)
Repetente	0.1160*** (0.0010)	0.1193*** (0.0012)	0.1193*** (0.0018)
Perseverança	0.0150*** (0.0004)	0.0127*** (0.0005)	0.0127*** (0.0005)
Esforço Real	0.0047*** (0.0002)	0.0055*** (0.0003)	0.0055*** (0.0003)
Esforço Potencial	-0.0009*** (0.0001)	-0.0009*** (0.0001)	-0.0009*** (0.0001)
Intercepto	5.5790*** (0.0040)	5.5710*** (0.0047)	5.5710*** (0.0054)
<i>N</i>	225629	225629	225629
adj. <i>R</i> ²	0.206		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabela A.8 - Determinantes do Resultado Individual – Ciências

Variáveis	OLS	QREG	SQREG
Educação Parental	0.1377*** (0.0012)	0.1489*** (0.0016)	0.1489*** (0.0016)
Dif. Educação	-0.0017*** (0.0002)	-0.0017*** (0.0003)	-0.0017*** (0.0003)
Tipo de Escola	0.0489*** (0.0010)	0.0507*** (0.0012)	0.0507*** (0.0013)
Sexo	0.0129*** (0.0008)	0.0157*** (0.0010)	0.0157*** (0.0011)
Pré-escolar	0.0945*** (0.0013)	0.0979*** (0.0016)	0.0979*** (0.0015)
Local Escolar	0.0360*** (0.0014)	0.0320*** (0.0018)	0.0320*** (0.0018)
Migrante	0.0255*** (0.0012)	0.0213*** (0.0015)	0.0213*** (0.0016)
Repetente	0.1061*** (0.0010)	0.1130*** (0.0013)	0.1130*** (0.0013)
Perseverança	0.0160*** (0.0004)	0.0144*** (0.0005)	0.0144*** (0.0004)
Esforço Real	0.0035*** (0.0002)	0.0041*** (0.0003)	0.0041*** (0.0004)
Esforço Potencial	-0.0009*** (0.0001)	-0.0009*** (0.0001)	-0.0009*** (0.0001)
Intercepto	5.5606*** (0.0040)	5.5394*** (0.0051)	5.5394*** (0.0055)
<i>N</i>	225629	225629	225629
adj. <i>R</i> ²	0.173		

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Desvios padrão em parênteses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabela A.9 - Efeito de Tratamento Médio – Circunstâncias – por Países

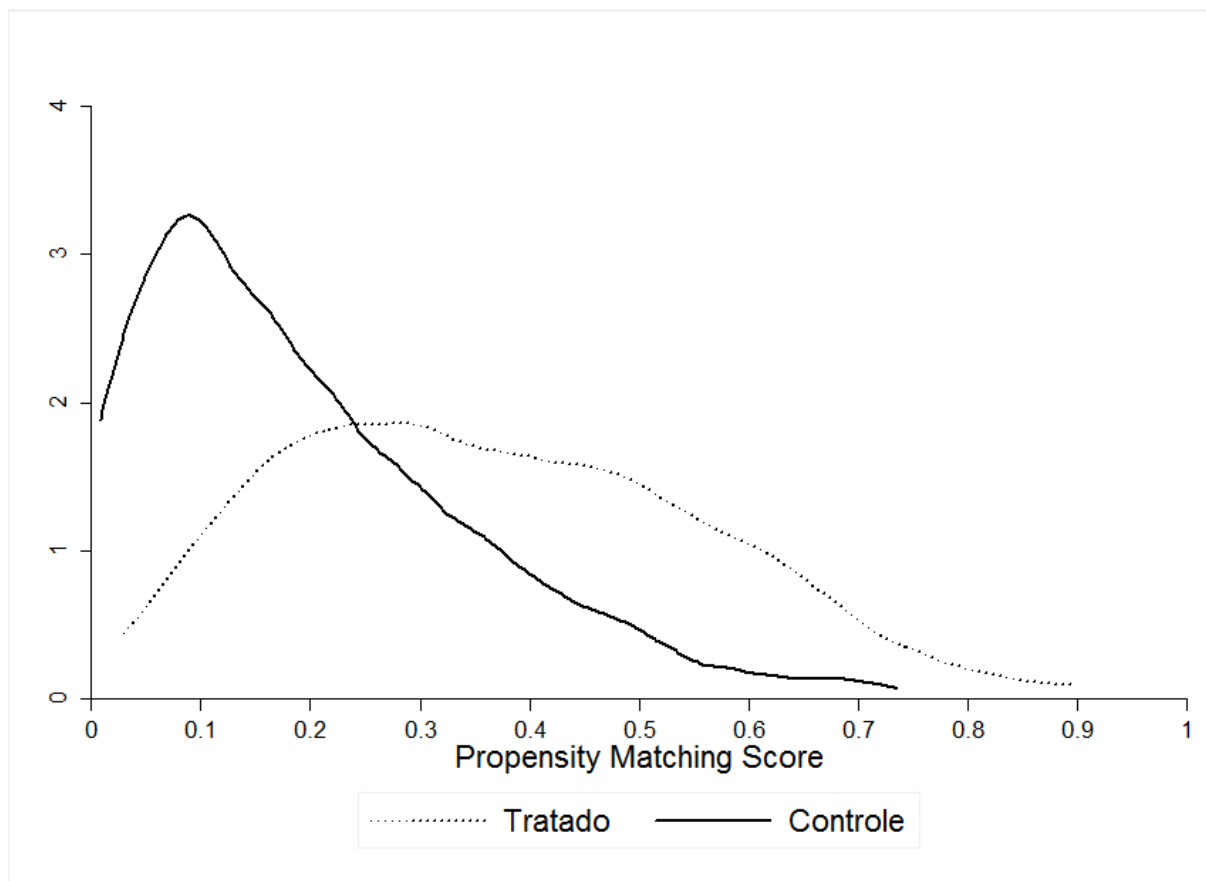
Continentes/Países	Educ. Mãe	Educ. Pai	Estrut. Fam.	Tipo Escola	Pré-escolar	Local Escola
América do Sul	0.098093 0.118730	0.058824 0.115888	0.024489 0.032333	0.051932 0.175486	0.056411 0.103262	0.064881 0.153443
Argentina	0.048386 0.104354	0.044594 0.082832	0.026608 0.031527	0.038053 0.146395	0.044710 0.142076	0.075749 0.084823
Brasil	0.058463 0.098093	0.071051 0.105993	0.020118 0.048574	0.054305 0.200697	0.036645 0.076164	0.057033 0.132065
Chile	0.039796 0.153724	0.045090 0.151598	0.021034 0.034362	0.042969 0.165054	0.072531 0.112729	0.051453 0.147919
Países da OCDE	0.013222 0.088401	0.012544 0.083916	0.007323 0.033916	0.016521 0.054301	0.041449 0.088213	0.016772 0.047812
Alemanha	0.026368 0.097151	0.024343 0.096465	-0.000638 0.001665	0.033427 0.069008	0.065402 0.124583	-0.043289 0.090472
França	0.046355 0.112805	0.037920 0.090926	0.028868 0.041832	0.001680 0.032044	0.100873 0.181397	0.093763 0.164834
Japão	0.001870 0.094489	0.000880 0.085052	0.003176 0.049396	0.020460 0.006075	0.095392 0.091853	- -
Estados Unidos	0.024951 0.104685	0.024790 0.106881	0.017103 0.065594	0.022829 0.010004	-0.024881 -0.029285	0.013143 0.014046
Países Asiáticos	0.021931 0.095099	0.021905 0.099773	0.013826 0.041630	0.002749 0.001243	0.099789 0.175045	-0.006460 0.099658
Singapura	0.004782 0.112488	0.004448 0.116527	0.015869 0.039765	0.037399 -0.020264	0.058787 0.100485	- -
Shangai	0.021960 0.085086	0.022025 0.087289	0.007872 0.014803	0.023597 0.046244	0.089232 0.199583	- -
Chinês Taipei	0.022646 0.100599	0.021297 0.099438	0.023363 0.059754	-0.014602 -0.104437	0.026113 0.090202	0.010898 0.104399
Hong Kong	0.008349 0.074560	0.009217 0.077834	0.011071 0.024750	-0.013261 -0.056790	0.013131 0.090307	- -
Países Nórdicos	0.032741 0.085580	0.026820 0.072986	0.010518 0.037608	0.001404 0.024242	0.053805 0.080825	0.001146 0.040688
Suécia	0.025901 0.060257	0.018805 0.054815	0.026544 0.050230	0.013286 0.035795	0.061552 0.100419	-0.005258 0.049242
Dinamarca	0.038928 0.100514	0.029816 0.084131	0.023696 0.045692	0.014031 0.047543	0.063090 0.101739	-0.017105 0.007524
Finlândia	0.036946 0.079705	0.025590 0.059861	0.013247 0.037939	-0.007247 0.011107	0.071405 0.084542	-0.002194 0.012848

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

Tabela A.10 – Testes de Especificação do Modelo

Descrição do Testes	Especificação dos Testes	
	Função do Teste	Resultado
Backward selection	Probabilidade de Remover Variável	Nenhuma
Forward selection	Probabilidade de Adicionar Variável	Todas
Forward hierarchical selection	Probabilidade de Ordem Lógica das Variáveis	Correta
Ovtest	Testa se há variáveis omitidas	Há variáveis omitidas

Fonte: Elaboração própria, a partir dos dados do PISA 2012.

APÊNDICE B: Resultados Adicionais**Gráfico B.1 – Sobreposição Propensity Score Matching**

Fonte: Elaboração própria.