



ECONOMIA APLICADA

BRAZILIAN JOURNAL OF APPLIED ECONOMICS

Vol. 19 – nº 2

Abril – Junho 2015

ARTIGOS

Impacto da violência nas escolas paulistas sobre o desempenho acadêmico dos alunos

Evandro Camargos Teixeira, Ana Lúcia Kassouf

Brazilian Female Labor Market: Racial-Skin Color Discrimination and Inefficiency

Gustavo Andrey de Almeida Lopes Fernandes

Estabilidade dos Professores e Qualidade do Ensino de Escolas Públicas

Leandro Oliveira Costa, Ronaldo de Albuquerque e Arraes, Daniel Barboza Guimarães

Determinantes socioeconômicos, estruturas produtivas regionais e condição ocupacional no Brasil, 2000-2010

Ana Carolina da Cruz Lima, Rodrigo Simões, Ana Maria Hermeto

Desigualdade de oportunidades na formação de habilidades: uma análise com dados longitudinais

Igor Vieira Procópio, Ricardo da Silva Freguglia, Flávia Chein

Eficiência nos mercados futuros agropecuários brasileiros

Marcos Aurelio Rodrigues, João Gomes Martines Filho

Desigualdade de oportunidades no acesso ao Ensino Superior no Brasil: uma comparação entre 2003 e 2013

Márcia Marques de Carvalho, Fábio D. Waltenberg

ECONOMIA FEA-RP/USP



ECONOMIA APLICADA

Vol. 19 – nº 2 Abril – Junho 2015

PAPERS

Impact of violence in São Paulo schools on the academic performance of students

Evandro Camargos Teixeira, Ana Lúcia Kassouf

Brazilian Female Labor Market: Racial-Skin Color Discrimination and Inefficiency

Gustavo Andrey de Almeida Lopes Fernandes

Stability of Teachers and Public Schools Education Quality

Leandro Oliveira Costa, Ronaldo de Albuquerque e Arraes, Daniel Barboza Guimarães

Socioeconomic determinants, regional structures and occupational status in Brazil, 2000-2010

Ana Carolina da Cruz Lima, Rodrigo Simões, Ana Maria Hermeto

Inequality of opportunity in skill formation: an analysis with longitudinal data

Igor Vieira Procópio, Ricardo da Silva Freguglia, Flávia Chein

Efficiency in Brazilian Agricultural Futures Markets

Marcos Aurelio Rodrigues, João Gomes Martines Filho

Inequality of opportunity in access to higher education in Brazil: A comparison between 2003 and 2013

Márcia Marques de Carvalho, Fábio D. Waltenberg

ISSN 1413-8050



9 771413 805001

APOIO:



Ministério da Educação

Ministério da Ciência e Tecnologia



ECONOMIA APLICADA

Vol. 19 – nº 2

Abril – Junho 2015

Editor

Reynaldo Fernandes

Editores Associados

Luiz Guilherme Scorzafave
Angelo Costa Gurgel
Rodrigo De Losso
Cláudio Ribeiro de Lucinda & Walter Belluzzo Junior

Sergio Naruhiko Sakurai
Eliezer Martins Diniz
Irineu E. de Carvalho Filho

Conselho Editorial

Affonso Celso Pastore
USP
Cássio E. Camargo Rolim
UFPR
Clélio Campolina Diniz
CEDEPLAR
Eleutério F. S. Prado
USP
Francisco Anuatti Neto
USP/RP
Geraldo Sant'ana de Camargo Barros
ESALQ/USP
José Marcelino da Costa
NAEA/PA
Marcelo Portugal
UFRGS
Márcio Gomes Pinto Garcia
PUC/ RJ
Paulo César Coutinho
UnB
Pierre Perron
Boston University
Ricardo R. Araújo Lima
UnB
Roberto Smith
UFCE
Rogério Studart
UFRJ
Sérgio Werlang
FGV/RJ
Victor Bulmer-Thomas
University of London
Wilson Suzigan
Unicamp

Antônio Barros de Castro
UFRJ
Cláudio Monteiro Considera
UFF
Denisard C. de Oliveira Alves
USP
Fernando de Holanda Barbosa
FGV-UFF
Geoffrey J. D. Hewings
University of Illinois
Gustavo Maia Gomes
IPEA
José A. Scheinkman
Princeton University
Maria José Willumsen
Florida International University
Mário Luiz Possas
UFRJ
Paulo Nogueira Batista Júnior
FGV/SP
Pedro Cesar Dutra Fonseca
UFRGS
Robert E. Evenson
Yale University
Rodolfo Hoffmann
ESALQ/USP
Russell E. Smith
Washburn University
Tomás Málaga
FGV/SP
Werner Baer
University of Illinois



A Revista ECONOMIA APLICADA é uma publicação trimestral do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo. Os conceitos emitidos são de responsabilidade exclusiva do(s) autor(es). A simples remessa do trabalho implica que o(s) autor(es) concordam que, em caso de aceitação para publicação, a ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS) passa a ter os direitos autorais para a veiculação dos artigos, tanto em formato impresso como eletrônico a eles referentes, os quais se tornarão propriedade exclusiva da ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS). É permitida a reprodução total ou parcial dos artigos desta revista, desde que seja explicitamente citada a fonte completa. Revista financiada com recursos do Programa de Apoio a Publicações Científicas do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, entidade governamental brasileira promotora do desenvolvimento científico e tecnológico. Esta revista está sendo indexada por Journal of Economic Literature, versões *eletronic on line* e CD-ROM e Scielo Brasil.

Secretaria e Divulgação: Juliana de Freitas Vitória Capa: Christof Gunkel
Revisão linguísticas (Português e Inglês): Iulo Feliciano Afonso (Solution Translations
Eireli — Epp)
Diagramação: Bruna Carolina Garcia

Endereço

Revista Economia Aplicada, Departamento de Economia, FEA-PR/USP
Avenida do Bandeirantes, 3900, Ribeirão Preto, SP.
CEP 14040-905
Telefone: (0xx16) 3315-3910
email:revecap@usp.br
site: <http://www.usp.br/revecap>

Assinaturas

tipo	Brasil	Exterior*
Individual	R\$50,00	R\$ 90,00
Institucional	R\$70,00	R\$110,00

*inclui custos de postagem

FICHA CATALOGRÁFICA

Elaborada pela Seção de Publicações e Divulgação do SBD/FEA/USP.

Economia Aplicada / Departamento de Economia, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo – v. 1, n. 1 – Ribeirão Preto: FEA-RP, 1997-

Trimestral

Publicado anteriormente sob responsabilidade de FEA/USP, FIPE e FEA=RP
ISSN 1413-8050

1. Economia I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto. Departamento de Economia.

CDD 22.ed – 330.

ECONOMIA APLICADA

Vol. 19 – nº 2

Abril – Junho 2015

ARTIGOS

- 221 **Impacto da violência nas escolas paulistas sobre o desempenho acadêmico dos alunos**
Evandro Camargos Teixeira , Ana Lúcia Kassouf
- 241 **Brazilian Female Labor Market: Racial-Skin Color Discrimination and Inefficiency**
Gustavo Andrey de Almeida Lopes Fernandes
- 261 **Estabilidade dos Professores e Qualidade do Ensino de Escolas Públicas**
Leandro Oliveira Costa , Ronaldo de Albuquerque e Arraes , Daniel Barboza Guimarães
- 299 **Determinantes socioeconômicos, estruturas produtivas regionais e condição ocupacional no Brasil, 2000-2010**
Ana Carolina da Cruz Lima , Rodrigo Simões , Ana Maria Hermeto
- 326 **Desigualdade de oportunidades na formação de habilidades: uma análise com dados longitudinais**
Igor Vieira Procópio , Ricardo da Silva Freguglia , Flávia Chein
- 349 **Eficiência nos mercados futuros agropecuários brasileiros**
Marcos Aurelio Rodrigues , João Gomes Martines Filho
- 369 **Desigualdade de oportunidades no acesso ao Ensino Superior no Brasil: uma comparação entre 2003 e 2013**
Márcia Marques de Carvalho , Fábio D. Waltenberg

IMPACTO DA VIOLÊNCIA NAS ESCOLAS PAULISTAS SOBRE O DESEMPENHO ACADÊMICO DOS ALUNOS

EVANDRO CAMARGOS TEIXEIRA *
ANA LÚCIA KASSOUF †

Resumo

O objetivo desse artigo é mensurar o impacto da violência sobre o desempenho escolar dos alunos do estado de São Paulo no ano de 2007. Para tal, é utilizado um modelo logit multinível, que possibilita a análise hierárquica dos dados. Os resultados demonstram que a violência nas escolas diminui em 0,54% a probabilidade de os alunos da terceira série do ensino médio apresentarem um desempenho satisfatório em matemática. A partir desse resultado, espera-se que políticas públicas sejam implementadas no intuito de diminuir a violência e elevar a probabilidade de desenvolvimento econômico do estado e consequentemente do país.

Palavras-chave: Violência; Desempenho Escolar; Desenvolvimento Econômico.

Abstract

The aim of this paper is to measure the impact of violence on school performance of students in the state of São Paulo in 2007. To this end, we use a multilevel logit model which enables the analysis of hierarchical data. The results show that violence in schools decreased by 0.54% the likelihood of third graders from the school present a satisfactory performance in Mathematics. From this result, it is expected that public policies are implemented in order to reduce violence and increase the probability of state economic development and consequently the country.

Keywords: Violence; School Performance; Economic Development.

JEL classification: I21, K42

DOI: <http://dx.doi.org/10.1590/1413-8050/ea124436>

* Professor Adjunto II do Departamento de Economia Rural da UFV. Viçosa, Minas Gerais, Brasil. E-mail: evandro.teixeira@ufv.br

† Professora Titular do Departamento de Economia, Administração e Sociologia da Esalq/USP. Piracicaba, São Paulo, Brasil. E-mail: anakassouf@usp.br

1 Introdução

A educação é uma variável chave na determinação do progresso econômico de qualquer país. Nesse sentido, indivíduos mais instruídos elevam o nível de capital humano e conseqüentemente as taxas de crescimento econômico. No entanto, um relevante fenômeno social vem diminuindo o nível de capital humano e afetando, em conseqüência, o desenvolvimento socioeconômico dos países: a violência nas escolas.

De acordo com Abramovay & Rua (2002), em todo o mundo ocidental moderno, a ocorrência de violência nas escolas não é um fenômeno recente. Desde os primeiros estudos realizados sobre o assunto, nos Estados Unidos, na década de 1950, diversas dimensões desse fenômeno passaram por grandes mudanças e os problemas decorrentes assumiram maior gravidade. Algumas dessas notáveis transformações foram: o surgimento de armas nas escolas; a disseminação do uso de drogas; e a expansão do fenômeno das gangues, influenciando na rotina das escolas, eventualmente associadas ao narcotráfico.

Guimarães (1998) afirma que, no Brasil, a violência nas escolas tem aumentado bruscamente nos últimos anos, sendo caracterizada principalmente pela presença do narcotráfico, de gangues e de pobreza acentuada. Nesse sentido, em um estudo realizado em diversas capitais do país, Abramovay & Rua (2002) concluíram que a violência física e a violência contra a propriedade são predominantes nas escolas dentre todos os tipos de violência, tendo diversos desdobramentos negativos sobre a qualidade do ensino e a aprendizagem. Tal efeito tem proporções elevadas para o país ao diminuir o nível de capital humano, tão importante no processo de desenvolvimento econômico.

Diante da importância do tema, este trabalho tem como objetivo analisar o impacto da violência nas escolas sobre o desempenho acadêmico dos alunos paulistas da terceira série do ensino médio em matemática no ano de 2007. O trabalho está dividido em mais cinco seções, além desta introdução. Na próxima seção serão apresentadas as evidências empíricas que concernem o tema e em seguida as estatísticas descritivas do banco de dados utilizado. Posteriormente, serão descritos o método e as variáveis a serem utilizadas. Logo em seguida, são apresentados os resultados da estimação do modelo proposto. Por fim, serão apresentadas as considerações finais.

2 Evidências Empíricas

A violência escolar tem despertado muita atenção por parte das autoridades públicas. Mas, apesar da importância do tema, poucas pesquisas têm sido realizadas. A maioria dos trabalhos foi realizada nos últimos anos e apresenta diversas facetas acerca da violência nas escolas: alguns analisam o impacto da violência dentro das escolas, outros a violência no entorno das escolas e os demais analisam ambos os impactos.

O primeiro trabalho relevante que versa o tema foi o de Grogger (1997). O autor estimou as conseqüências da violência nas escolas sobre o desempenho acadêmico dos alunos do ensino médio de escolas públicas nos Estados Unidos. O autor classificou cada tipo de violência ocorrido nas escolas como sério, moderado, menor ou não existente. Além da violência escolar, Grogger (1997) levou em consideração características dos estudantes que aumentam a propensão à violência, tais como cor, *background* familiar, além de características das escolas e da região onde a mesma está inserida. O autor concluiu

que níveis moderados de violência diminuam a probabilidade de conclusão do ensino médio em 5,1%, assim como diminuem a frequência escolar em 6,9%.

Seguindo a mesma linha do trabalho de Grogger (1997), Hurt et al. (2001), Henrich et al. (2004) e Ratner et al. (2006) também examinaram a associação entre exposição à violência nas escolas e desempenho escolar. Os primeiros autores citados levaram em consideração em sua amostra alunos afro-americanos de seis a sete anos. Os resultados mostraram que a exposição à violência está positivamente relacionada à baixa *performance* escolar, sintomas de ansiedade e baixa autoestima.

Por sua vez, Henrich et al. (2004) utilizaram como amostra 759 alunos norte-americanos numa análise *cross-section*. Todavia, diferentemente dos trabalhos até então realizados, esse estudo concluiu que a vitimização não está relacionada a baixos níveis de proficiência. Entretanto, a vitimização apresentou uma relação inversa com a sensação de segurança nas escolas, no caso específico de meninos que relataram não ter suporte da família.

Já Ratner et al. (2006) utilizaram como amostra filhos de mulheres que receberam assistência pré-natal e foram continuamente examinadas pelos hospitais da cidade de Detroit (EFUA). Os resultados corroboraram com a maior parte da literatura, que concerne o tema, ou seja, a violência está negativamente relacionada às medidas de proficiência.

No mesmo ano, McGarvey et al. (2006), diferentemente dos trabalhos até aqui apresentados, levaram em consideração o crime tanto dentro como no entorno das escolas e seu impacto sobre o desempenho acadêmico de alunos do ensino fundamental e médio em Atlanta (EUA) entre 1999 e 2002. Os autores utilizaram dois métodos: Mínimos Quadrados Ordinários — MQO e estimação via variáveis instrumentais (número total de adultos na escola, distância da escola à habitação pública mais próxima, número de estações ferroviárias no bairro e distância da escola às estações de trânsito ferroviário). Os resultados obtidos pelos dois métodos foram similares. Um incidente de violência em uma escola faz com que o desempenho acadêmico caia em 2,5% (pelo método de MQO) e em 4,0% (pela estimação via variáveis instrumentais)

Ainda nesse mesmo ano, em um estudo menos abrangente, Carroll (2006) calculou o efeito de atos violentos e criminosos em escolas públicas da Carolina do Norte (EUA), entre 2003 e 2004, sobre os níveis de desempenho acadêmico de alunos da oitava série, controlando por uma série de fatores pertinentes. A fim de desmembrar o efeito da violência sobre o desempenho dos alunos em leitura e matemática e corrigir possíveis correlações estatísticas causadas pela possibilidade de omissão de variáveis, o autor usou o método de Regressões Aparentemente Não Correlacionadas — SUR. Como resultado, verificou-se que um crime ou violência diminuam marginalmente o desempenho do aluno em 0,138% e 0,143% em matemática e leitura, respectivamente.

Ammermueller (2007) analisou os efeitos da violência escolar sobre a *performance* dos estudantes em um âmbito diferente dos trabalhos até então realizados: em onze países da Europa. Assim como verificado na literatura, o autor também concluiu que a violência afeta negativamente o desempenho acadêmico dos alunos. No entanto, o autor chegou a outras conclusões interessantes. Em primeiro lugar, Ammermueller (2007) verificou que alunos da oitava série, do sexo feminino e oriundos de municípios de pequeno porte apresentaram menor probabilidade de serem vítimas da violência escolar. Além disso, ao analisar a amostra da quarta série o autor ainda percebeu que o fato de o

aluno ser imigrante aumenta a possibilidade de vitimização, enquanto que o *background* familiar diminui tal probabilidade.

Nesse sentido, enfatizando a importância das características do aluno e de seu *background* familiar, Aizer (2008) destacou que as vítimas da violência possuem várias características em comum: em sua maioria são negras, pobres e de famílias com baixa escolaridade. Assim, a autora procurou resposta para o seguinte questionamento: a baixa *performance* escolar é melhor explicada pela exposição à violência ou pelas características e o *background* familiar dos alunos? Aizer (2008) utilizou efeitos fixos a fim de considerar características não observáveis das famílias dos alunos, assim como dos bairros onde os mesmos residem, concluindo que as características e o *background* familiar são os maiores responsáveis pelo decréscimo no desempenho acadêmico dos alunos.

No mesmo ano, Hjalmarsson (2008) examinou empiricamente um aspecto diferente dentro do tema: os efeitos de apreensões e encarceramento juvenis aos 16 anos sobre a conclusão do ensino médio aos 19 anos nos Estados Unidos. O autor concluiu que jovens que são presos têm 25% menor probabilidade de completar o ensino médio em relação a jovens com a mesma idade que não foram presos.

No Brasil, a literatura é ainda mais incipiente, sendo raros os trabalhos realizados que tratam desse tema. Severnini (2007) utilizou os indicadores de violência escolar estabelecidos no Sistema de Avaliação da Educação Básica — SAEB de 2003, concluindo que alunos oriundos de escolas mais violentas tiveram, em média, pior desempenho em matemática e português. O autor constatou também que existe uma relação indireta entre violência e proficiência acadêmica, que se evidencia por meio da elevação da rotatividade dos professores.

Gama (2009) analisou a relação entre violência e proficiência escolar de alunos da quarta série do ensino fundamental no município de São Paulo em 2005, encontrando resultados muito diferentes de Severnini (2007). O autor concluiu de forma surpreendente que alunos de escolas públicas de regiões violentas têm melhor desempenho em matemática e português para crimes violentos, crimes contra o patrimônio, e roubo consumado. Além disso, utilizando equações quantílicas, Gama (2009) ainda constatou que a violência afeta mais os alunos da cauda superior da distribuição de notas.

Becker & Kassouf (2012) analisaram um aspecto diferente com relação aos trabalhos realizados até então no país: o impacto da violência no entorno das escolas sobre a violência dentro das mesmas. Para isso, as autoras utilizaram dados da Prova Brasil para estimar um modelo *logit* de efeitos fixos, concluindo que a possibilidade de observar um ato agressivo de um aluno é maior em ambientes escolares com traços da violência, onde ocorreram crimes contra patrimônio, contra a pessoa, tráfico de drogas ou atuação de gangues. Becker & Kassouf (2012) também observaram que a possibilidade de um aluno cometer uma agressão física é 3,54 vezes maior em escolas onde foi observado o mesmo comportamento por parte de um professor.

Diante da literatura que concerne o tema apresentada acima, este trabalho inova ao analisar a relação entre violência e desempenho escolar no estado de São Paulo em 2007 por meio de um modelo multinível. Tal método considera a hierarquia existente nos dados (nível individual e nível agregado) e permite ao pesquisador verificar o impacto individual da inclusão de cada variável contextual (dentre elas, a violência nas escolas) no modelo sobre a *performance* acadêmica dos alunos.

3 Estatísticas Descritivas

Nesta seção, será apresentada uma breve descrição da base de dados utilizada neste trabalho, tendo como fonte o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Saresp) 2007.¹ No que tange ao desempenho dos alunos da terceira série do ensino médio na prova de matemática, verifica-se que somente 3,7% apresentaram um resultado considerado “avançado”. O desempenho dos demais alunos foi considerado “básico” (24,79%) ou ainda “abaixo do básico” (71,51%).

Com relação ao gênero dos alunos, percebe-se que há uma predominância das mulheres em relação aos homens: 54,73% são do sexo feminino e 45,27% do sexo masculino. No que tange à cor, percebe-se a maioria dos alunos é branca (54,37%) ou parda (31,79%) em relação a negros (10,39%), amarelos (2,30%) e indígenas (1,15%).

A Tabela 1 apresenta o nível socioeconômico dos estudantes por meio de alguns indicadores relevantes, o que compõe o denominado *background* familiar, ou seja, características socioeconômicas individuais dos alunos da terceira série do ensino médio, que podem afetar seus respectivos níveis de desempenho escolar na prova de matemática. Dada a indisponibilidade do nível de renda dos alunos diretamente, tais indicadores têm o objetivo de verificar a presença ou não de alguns bens e condições mínimas de moradia.

Percebe-se que a grande maioria dos alunos possui em seus domicílios, principalmente, os seguintes bens: tv em cores, rádio e geladeira. Em contrapartida, a maior parte dos alunos não possui aspirador de pó. Com relação às condições de moradia, a maioria dos alunos as possui minimamente. Cabe salientar que rua sem calçamento é o problema estrutural mais recorrente entre os alunos.

Uma variável relevante a ser analisada é a inserção do aluno no mercado de trabalho. Aproximadamente 54,46% dos alunos trabalham e 45,64% não trabalham. Mais importante que isso, do total de alunos que trabalham, 44,76% relatam que tais atividades atrapalham seus estudos.

A Tabela 2 apresenta, respectivamente, o grau de escolaridade dos pais dos alunos (pais e mães), que também compõe o *background* familiar. Verifica-se que o nível de escolaridade dos pais é relativamente baixo, pois a maior parte terminou apenas o ensino fundamental (57,54% e 56,27%, respectivamente) e um contingente relativamente pequeno possui curso superior (8,61% e 9,26%, respectivamente).

Nas Tabelas 3, 4 e 5, são apresentadas as condições estruturais, assim como os principais problemas enfrentados pelos diretores das escolas onde os alunos estudam. A Tabela 3 descreve como são as condições dos principais equipamentos da escola. As quadras e os espaços para recreação e refeições são os que menos atendem aos anseios dos alunos, segundo os diretores. Já em relação aos laboratórios, a situação é mais dramática, pois simplesmente 62,57% dos diretores afirmam que os mesmos não existem em suas escolas.

Com relação às condições e conservação do ambiente escolar, vide Tabela 4, conclui-se que a maioria dos diretores as considera boas ou regulares. A maior demanda por parte dos alunos, de acordo com os diretores, é que seja melhorada a limpeza e a conservação da área externa.

¹Ver São Paulo (2007).

Tabela 1: Nível socioeconômico dos alunos — proporção (%)

Indicadores Socioeconômicos	Não Possui	Possui
Tv em cores	2,92	97,08
Videocassete ou DVD	13,87	86,13
Computador	44,57	55,43
Rádio	7,84	92,16
Máquina de lavar roupas	20,18	79,82
Aspirador de pó	67,93	32,07
Telefone fixo	27,38	72,62
Telefone celular	11,03	88,98
Geladeira	5,28	94,72
Carro	45,01	54,99
Banheiro em casa	2,03	97,97
Rua com calçamento	14,87	85,13
Água encanada em casa	3,25	96,75
Eletricidade em casa	1,44	98,56

Fonte: Elaborada pelos autores a partir de informações de São Paulo (2007).

Tabela 2: Grau de escolaridade dos pais dos alunos — proporção (%)

Grau de parentesco	Níveis de escolaridade dos pais			
	Ensino fundamental	Ensino médio	Ensino superior	Nunca frequentou
Mães	57,54	31,08	8,61	2,77
Pais	56,27	31,18	9,26	3,29

Fonte: Elaborada pelos autores a partir de informações de São Paulo (2007).

A Tabela 5 apresenta as principais condições adversas relatadas pelos diretores das escolas. Os diretores, em sua maioria, concordam que existe precariedade em termos de infraestrutura física nas escolas (65,27%), falta de motivação por parte dos professores (77,20%) e apoio insuficiente da comunidade (59,26). Com relação à variável foco deste trabalho — insegurança na escola (vandalismo, roubos, etc.) — 41,71% dos diretores concordam que existem tais transtornos nas escolas em que trabalham. Assim, um percentual substancial das escolas que fazem parte da base de dados que compõe este trabalho é acometido por violência, o que denota a importância do presente estudo.

4 Metodologia

Segundo Riani & Rios-Neto (2004), existem três conjuntos de fatores que afetam de forma distinta o desempenho escolar de um indivíduo: familiar, escolar e comunitário. O primeiro altera a demanda intradomiciliar por educação, enquanto que os outros dois alteram o lado da oferta educacional. Muitos trabalhos utilizam variáveis relacionadas aos três conjuntos de fatores citados no mesmo nível, estimando os determinantes educacionais por meio do método de Mínimos Quadrados Ordinários (MQO). Esse tipo de análise apresenta di-

Tabela 3: Condições dos principais equipamentos das escolas — proporção (%)

Indicadores	Condições				
	Precárias	Regulares	Boas	Excelentes	Não existe (m)
Prédio escolar	8,00	36,47	49,74	5,79	-
Salas de aula	5,58	37,41	52,07	4,94	-
Laboratórios	9,03	13,96	12,87	1,57	62,57
Sala de informática	10,94	24,95	46,38	12,16	5,57
Biblioteca	10,87	24,38	36,00	10,19	18,56
Quadras	26,40	32,25	27,77	5,61	7,97
Sala para os professores	9,14	32,55	47,07	8,29	2,95
Espaço para recreação	13,26	32,41	33,94	7,38	13,01
Espaço para refeições	14,47	32,68	35,04	7,40	10,41

Fonte: Elaborada pelos autores a partir de informações de São Paulo (2007).

Tabela 4: Condições e conservação do ambiente escolar — proporção (%)

Indicadores	Condições			
	Precárias	Regulares	Boas	Excelentes
Condições dos vidros, lâmpadas, encanamentos	12,12	39,71	42,54	5,63
Limpeza e conservação dos banheiros	5,77	29,17	55,22	9,84
Conforto e conservação do mobiliário	6,72	40,72	48,22	4,34
Limpeza e conservação da área externa	4,82	27,99	56,90	10,29
Limpeza e conservação das paredes e muros	6,86	34,87	51,66	6,61
Aparência geral da escola	6,76	30,57	53,35	9,32

Fonte: Elaborada pelos autores a partir de informações de São Paulo (2007).

Tabela 5: Condições adversas observadas nas escolas — proporção (%)

Indicadores	Concordam	Discordam
Precariedade da infra-estrutura física	65,27	34,73
Insegurança na escola (vandalismo, roubos, etc)	41,71	58,29
Poucas oportunidades de desenvolvimento	32,56	67,44
Falta de motivação dos professores	77,20	22,80
Apoio insuficiente da equipe escolar	36,96	63,04
Apoio insuficiente da comunidade	59,26	40,74

Fonte: Elaborada pelos autores a partir de informações de São Paulo (2007).

ficuldades tanto conceituais como estatísticas.

De acordo com Shirasu & Arraes (2014), a estrutura do sistema educacional é organizada de modo hierárquico ou em níveis. Assim, um grupo de alunos constitui uma turma, o agrupamento de turmas compõe as escolas, o conjunto destas está inserido em um município, os municípios estão agrupados em estados e assim por diante. Nesse sentido, diversos fatores afetam a distribuição dos alunos entre as escolas, como questões socioeconômicas, disponibilidade de vagas ou mesmo a distância da escola ao local de residência. Tais fatores implicam na não aleatoriedade dessa distribuição ao fazerem com que os alunos de uma mesma escola sejam mais parecidos do que os de escolas diferentes. A semelhança advém desse contexto, onde as escolas estão inseridas, além do fato de compartilharem os mesmos professores e se beneficiarem da mesma infraestrutura.

Dessa forma, segundo os autores, não se pode tratar cada indivíduo de uma escola como independente dos demais alunos da mesma. Nesses casos, há violação de uma importante hipótese assumida pelo modelo de regressão linear clássico: de que os termos de erro são independentes e identicamente distribuídos, o que torna a estimação por mínimos quadrados ordinários inadequada. Tal inadequação, segundo Lee (2001), ocorre em função dos seguintes problemas: vies de agregação, estimação errônea do erro padrão, heterogeneidade dos coeficientes da regressão e diferenças entre agrupamentos.

Lee (2010) afirma que os trabalhos que não levam em consideração a natureza multinível da escola quase sempre apresentam resultados enviesados e, geralmente, a direção desse vies tende a subestimar o efeito-escola. Por levar em consideração a natureza hierárquica dos dados educacionais e a violação da hipótese de independência do termo de erro, alguns estudos na literatura têm utilizado os modelos multiníveis.

Assim, conforme salientado por Ferrão et al. (2002), na modelagem de dados educacionais, a aplicação dos modelos de regressão multinível possui duas vantagens em relação aos modelos de regressão clássica. Em primeiro lugar, na presença de correlação intraclasse, a estimação dos parâmetros do modelo via regressão clássica produz estimativas do erro padrão muito pequenas. Pode-se dizer assim que as estimativas produzidas pelos modelos de regressão multinível são geralmente mais conservadoras.

Além disso, ao decompor a variância do erro segundo os níveis hierárquicos, o modelo de regressão multinível permite ao analista a melhor compreensão e/ou explicação do processo que está sendo modelado. Nesse sentido, torna-se mais simples, por exemplo, estudar a capacidade explicativa de variáveis intraescolares face às extraescolares.

Conclui-se, portanto, que ao se trabalhar com dados em diferentes níveis é imprescindível o uso dos modelos hierárquicos. Nesse sentido, no intuito de analisar o impacto da violência nas escolas sobre o desempenho dos alunos será utilizado um modelo logístico hierárquico. Como já salientado, quando se incorpora a modelagem hierárquica no modelo logístico, pressupõe-se a existência de um conjunto de dados hierárquicos, onde a variável dependente é medida no nível mais baixo e as variáveis explicativas em todos os níveis existentes. Neste trabalho, o nível mais baixo contempla o indivíduo — nível individual — e o nível mais alto, as características da escola — nível agregado.²

²Ver Hox (1995).

Para a especificação do modelo de nível individual, y_{ij}^* é considerada uma variável contínua latente, não observável, que representa o desempenho acadêmico do aluno i na escola j (nível agregado). Essa variável pode ser definida a partir da seguinte relação linear:

$$y_{ij}^* = \beta_{0j} + X_{kij}\beta_{kj} + \epsilon_{ij} \quad (1)$$

onde:

X_{kij} = vetor das k variáveis independentes medidas no nível individual;
 β_{0j} = intercepto; β_{kj} = vetor dos k parâmetros a serem estimados pelo modelo;
 ϵ_{ij} = termo aleatório, com distribuição normal, com média zero e variância constante.

Considere uma variável binária observada y_{ij} , que mede o desempenho acadêmico do aluno, definido por:

$$y_{ij} = 1 \text{ se } y_{ij}^* > 0, \\ y_{ij} = 0 \text{ caso contrário}$$

tal que:

$$E(y_{ij}^* | X'_{ijk}) = X'_{ijk}\beta_{kj} \quad (2)$$

Ao se especificar uma função de probabilidade logística obtém-se o “modelo de chance proporcional” com coeficientes randômicos da seguinte forma:

$$P(y_{ij} = 1 | X_{ij}) = \frac{\exp(X'_{ijk}\beta_{kj})}{1 + \exp(X'_{ijk}\beta_{kj})} \quad (3)$$

Esse modelo é semelhante ao modelo de regressão logistic padrão. A diferença é que a análise multinível considera a natureza hierárquica dos dados e permite analisar a relação entre variáveis medidas no nível agregado e outras medidas no nível individual. Alguns ou todos os parâmetros estimados podem conter um termo aleatório que varia segundo a unidade de análise medida no nível 2, ou seja, no nível agregado. Neste trabalho, apenas o intercepto é suposto aleatório, no qual o modelo de nível 2 é especificado pelas equações a seguir:

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad (4)$$

$$\beta_{kj} = \gamma_{k0}, \text{ com } k \neq 0 \quad (5)$$

onde: μ_{0j} = termo residual no nível agregado com distribuição normal $\mu_j \approx (0, \sigma_\mu^2)$.

Definido e justificado o método a ser utilizado, é importante analisar a base de dados utilizada, que neste estudo tem como fonte o Saresp 2007 — (São Paulo 2007). Essa base de dados possui características que corroboram com a estrutura hierárquica das relações. O Saresp é um teste aplicado desde 2001 no estado de São Paulo que avalia as habilidades cognitivas dos alunos nas disciplinas de matemática e português. O desempenho do estudante é mensurado de acordo com a Teoria de Resposta ao Item, comumente chamada

de TRI. Essa variável é o indicador das habilidades e competências dos estudantes, ou seja, é o indicador de qualidade na sua formação. São avaliados alunos do ensino fundamental e do ensino médio de escolas públicas e privadas. Juntamente aos resultados dos testes são incorporadas as características socioeconômicas dos alunos, assim como informações relevantes a respeito das escolas.

A variável dependente é uma medida categórica do desempenho acadêmico dos alunos da terceira série do ensino médio em matemática, tendo como fonte primária os dados do Saesp (2007). Para classificar o desempenho dos alunos é necessário estabelecer um ponto de corte entre essas categorias. Neste trabalho, para o aluno que teve seu resultado na prova de matemática classificado como “avançado” ou “básico” atribui-se (1) na variável profmat. Em contrapartida, no caso do aluno que obteve resultado na prova classificado como “abaixo do básico” atribui-se (0) para tal variável.

A opção em se utilizar uma variável dependente categórica é bastante plausível. A nota do aluno, calculada a partir da TRI, é uma estimativa dos acertos, levando-se em consideração que o aluno tenha respondido a todas questões da prova, o que na prática não ocorre. Assim, no Saesp nunca é utilizada a média de acerto pura e simples, mas sim a proficiência média do aluno gerada a partir da TRI. Por isso, a nota individual do aluno não é divulgada, pois ela não corresponde exatamente aos seus acertos nas provas específicas.

Após o cálculo da média de proficiência de todos os alunos da escola, é realizada a distribuição dessas médias nos níveis de desempenho: abaixo do básico, básico e avançado para cada disciplina e ano/série avaliados. Assim, é determinado em quais escolas predominam alunos que estão nos níveis mais avançados ou básicos. Acredita-se que a distribuição dos alunos nos níveis expressa de forma mais adequada a realidade em termos de aprendizagem nas escolas, pois demonstra mais claramente a quantidade de alunos que precisam de maior atenção por parte dos professores e/ou direção das escolas.

A média simples da proficiência dos alunos é um resultado muito limitado, como qualquer medida de tendência central, e oferece menos subsídios no intuito de verificar quais alunos necessitam de mais atenção. Assim, no cálculo do Índice de Desenvolvimento da Educação do estado de São Paulo (IDESP) utiliza-se a distribuição dos alunos nos níveis como forma de caracterizar melhor a população da escola, obtendo-se, assim, informações se a escola possui mais alunos nos níveis inferiores ou superiores, o que determinará projetos pedagógicos dirigidos a serem encaminhados à mesma (São Paulo 2014).

Com relação às variáveis de controle, estas foram escolhidas de acordo com a literatura sobre os determinantes do desempenho educacional. Dessa forma, no nível individual foram utilizadas medidas que caracterizam a demanda escolar, ou seja, variáveis que representam características específicas dos alunos e de sua família.

No nível agregado são incluídas variáveis que caracterizam a oferta educacional, que possam influir no desempenho dos alunos. Com esse objetivo, serão utilizadas as seguintes variáveis: violência na escola, percentual de professores efetivos, participação da comunidade na escola, conservação das instalações físicas e dos equipamentos da escola, e existência e quantidade de equipamentos em condições de uso na escola.

4.1 Variáveis a nível individual

A maior parte da literatura brasileira, que analisa os determinantes do desempenho acadêmico dos alunos, demonstra que os fatores que melhor explicam o desempenho escolar são as características familiares e ambientais do aluno.

Neste trabalho, as duas primeiras variáveis consideram características físicas dos alunos: “sexo” e “cor”. Assim, inicialmente, foi incluída uma variável binária que identifica o sexo do aluno: (0) para feminino e (1) para masculino. Posteriormente, foram incluídas cinco variáveis binárias que denotam a cor do estudante: branco, negro, pardo, amarelo e indígena.

Uma variável de extrema relevância, mas que apresenta uma dificuldade inerente ao questionário aplicado junto aos alunos é o nível socioeconômico, pois não existe nenhuma variável que denote o valor da renda obtida pela família do estudante. Assim, será construída por Análise de Componentes Principais — ACP,³ a variável “nível socioeconômico” dos alunos, que inclui os seguintes elementos, tendo como base os critérios de classificação socioeconômica da Associação Brasileira de Empresas de Pesquisas — ABEP (2007):⁴

- número de automóveis;
- número de aparelhos de TV em cores;
- número de rádios;
- número de empregadas domésticas;⁵
- número de banheiros;
- posse de máquina de lavar roupa;
- posse de geladeira e freezer;
- posse de vídeo cassete ou dvd;
- nível de instrução do chefe de família.

Nesse sentido, com o intuito de construir o indicador de nível socioeconômico dos alunos foram criadas:

- três variáveis binárias que denotam a posse de geladeira, freezer, vídeo cassete ou dvd e máquina de lavar roupa.

³A Análise de Componentes Principais — ACP é uma técnica de análise multivariada que consiste em transformar um conjunto original de variáveis em outro conjunto — os Componentes Principais — CP. Os CP's são combinações lineares das variáveis originais e são estimados de forma a captar o máximo de variação total dos dados. O processo de estimação é tal que o primeiro CP capta o máximo de variância possível, o segundo capta o máximo possível do restante de variância, o terceiro o máximo possível do restante de variância, e assim sucessivamente. Segundo Mingoti (2005), o principal objetivo é o de explicar a estrutura de variância-covariância de um vetor aleatório, composto de p-variáveis aleatórias, por meio da construção de combinações lineares das variáveis originais. Tais combinações lineares são chamadas de componentes principais e não são correlacionadas entre si. Se existirem p-variáveis originais é possível obter-se p componentes principais. No entanto, em geral deseja-se obter 'redução do número de variáveis a serem avaliadas e interpretação das combinações lineares construídas.

⁴A única variável não incluída foi o número de empregadas domésticas, que não está disponível em São Paulo (2007).

⁵Com as informações disponibilizadas em São Paulo (2007) não foi possível a construção dessa variável.

- variáveis que identificam o número de automóveis, de rádios, de aparelhos de TV em cores e de banheiros no domicílio do aluno com a seguinte ponderação: (0) se o aluno não possui, (1) se o aluno possui uma unidade, e (2) se o aluno possui mais de uma unidade.
- variáveis que correspondem ao nível de instrução do pai e da mãe⁶ do aluno com a seguinte ponderação de acordo com o nível de escolaridade: (0) nunca frequentou a escola, (1) ensino fundamental — até a quarta série, (2) ensino fundamental — até a oitava série, (3) ensino médio e (4) ensino superior.

Percebe-se pelo processo de construção das variáveis que compõem a Análise de Componentes principais — ACP que quanto maiores seus valores, melhores tendem a ser as condições socioeconômicas dos alunos. Nesse sentido, espera-se uma relação diretamente proporcional entre a variável nível socioeconômico e o desempenho escolar.

As últimas variáveis a serem incluídas no nível individual são aquelas que identificam se o aluno exerce algum tipo de trabalho. Assim, foram incluídas seis variáveis binárias:

- Trabalho1: aluno nunca trabalhou ou não está trabalhando atualmente;
- Trabalho2: aluno trabalha sem jornada fixa, até 10 horas semanais;
- Trabalho3: aluno trabalha de 11 a 20 horas semanais;
- Trabalho4: aluno trabalha de 21 a 30 horas semanais;
- Trabalho5: aluno trabalha de 31 a 40 horas semanais;
- Trabalho6: aluno trabalha mais de 40 horas semanais.

Espera-se que quanto maior o número de horas de trabalho exercidas pelo aluno, menor tenda a ser o seu desempenho escolar. Segundo Kassouf (2007), boa parte dos prejuízos na educação ocorre em função do trabalho desde a infância. De acordo com a autora, a baixa escolaridade e o desempenho escolar mais baixo, consequências do trabalho infantil, restringem as oportunidades de emprego a postos de baixa qualificação e, portanto, de baixa remuneração.

4.2 Variáveis a nível agregado

As variáveis incluídas no nível agregado possuem a finalidade de captar as condições de oferta escolar, que impactam sobre o desempenho escolar dos alunos e terão como base o questionário respondido pelo diretor da escola.

A variável “violência” avalia se na escola ocorrem roubos, vandalismo, agressões, e etc. Caso o diretor reporte que há violência na escola esta variável assume o valor 1. Caso contrário, seu valor é 0. A existência de violência na escola diminuiu principalmente a frequência escolar e a concentração dos alunos. Nesse sentido, espera-se que quanto maior a violência nas escolas menor tenda a ser a performance escolar.

⁶Essas duas variáveis caracterizam a escolaridade apresentada pelo chefe de família, já que não é possível identificá-lo diretamente no questionário.

De acordo com Carroll (2006), a qualidade dos professores pode ter um efeito importante sobre o desempenho dos alunos na prova. Nesse caso, professores mais experientes, com maior titulação, e efetivos corroboram para um desempenho escolar mais elevado. Assim, foram incluídas cinco variáveis binárias que denotam qual o percentual de professores efetivos na escola:

- Profefet1: se até 25% dos professores são efetivos;
- Profefet2: de 26% a 50% dos professores são efetivos;
- Profefet3: de 51% a 75% dos professores são efetivos;
- Profefet4: de 76% a 90% dos professores são efetivos;
- Profefet5: de 91% a 100% dos professores são efetivos;

Essas cinco variáveis binárias incluídas medem também a rotatividade dos professores, pois um professor efetivo tende a mudar menos de trabalho. Como especificado por Abramovay & Rua (2002), a excessiva rotatividade dos professores nas escolas é um dos fatores que explicam o mau desempenho escolar. Nesse sentido, espera-se que quanto maior o percentual de professores efetivos mais elevado tenda a ser o desempenho escolar.

No caso das três demais variáveis de controle do nível agregado (comunidade, conservação e equipamento) também será utilizada a técnica de ACP a fim de se extrair um componente principal. A variável “comunidade” assinala a frequência em que se dá a participação da comunidade nos seguintes aspectos: colaboração dos pais na solução dos problemas da escola; presença e apoio dos pais nos eventos e festas que a escola promove; participação dos pais no Conselho de Escola; e prestação de serviços à escola (pintura, reparos, outros cuidados, doação de bens). Para cada um desses itens foi criada uma variável com a seguinte ponderação: (0) raramente ou nunca, (1) regularmente, (2) muitas vezes. Assim, espera-se que quanto maior o valor da variável “comunidade” mais elevado tenda a ser o desempenho educacional.

Por sua vez, a variável “conservação” avalia a conservação e o funcionamento das seguintes instalações e equipamentos da escola: prédio escolar; salas de aula; laboratórios; sala de informática; biblioteca; salas para exibição de filmes e DVD’s; quadras, salas para os professores; sala específica para o professor coordenador; salas destinadas ao grêmio estudantil; espaço para recreação; espaço para as refeições; condições dos vidros, lâmpadas e encanamentos; limpeza e conservação dos banheiros; condições de conforto e conservação do mobiliário escolar; limpeza e condição da área externa (pátio, jardins, e etc.); condições de limpeza e conservação das paredes e muros; e aparência geral da escola (cortinas, filtros, objetos de decoração). Para cada um desses itens também foi criada uma variável com a seguinte ponderação: (0) precárias, (1) regulares, (2) boas e (3) excelentes. Nesse caso, quanto maior a conservação das instalações e dos equipamentos da escola e, portanto, o valor da variável “conservação”, mais elevada tende a ser a *performance* escolar.

Por fim, inclui-se a variável “equipamento”, que identifica se existem e quantos são os equipamentos em condições de funcionamento na escola, tais como: televisores; videocassetes e/ou dvd’s; retroprojetores; computadores para uso dos professores; computadores para uso dos alunos; computadores para uso administrativo; impressoras; máquinas copiadoras (xerox); antena

parabólica; e aparelhos de som. A partir de cada uma dessas informações uma variável será criada com a seguinte ponderação: (0) não existem, (1) nenhum, (2) alguns, (3) metade, e (4) todos. Assim, quanto maior o valor da variável “equipamento”, a tendência é que o desempenho escolar seja mais elevado.

5 Resultados

Na Tabela 6, são apresentados os resultados⁷ da estimação dos efeitos fixos e aleatórios estimados pelo modelo *logit* multinível de intercepto randômico. Por meio dos resultados, percebe-se que todos os coeficientes foram estatisticamente significativos a um nível de significância de 1% e apresentaram os sinais esperados.

Foram testadas quatro especificações, concomitante à inclusão gradativa das variáveis contextuais. Nesse sentido, a especificação mais simples é o modelo não condicional (modelo 1), na qual nenhuma variável do nível agregado é considerada na análise, e o modelo 4 é o mais complexo, no qual todas as variáveis contextuais fazem parte da análise.

Uma questão que surge nesse momento é a seguinte: se cada modelo representa a inserção gradativa de variáveis contextuais, qual a importância da inclusão das mesmas na análise? Para responder a tal questionamento será utilizado o método proposto por Raudenbush & Bryk (2002). Esses autores especificaram uma forma de verificar em quanto a inserção das variáveis do nível agregado contribuem para reduzir o componente da variância estimada relativo ao intercepto. Quanto menor tal componente de variância, mais elevado tende a ser o poder de explicação (percentual de variância explicada) das variáveis contextuais em relação ao desempenho escolar individual dos alunos. Assim, Raudenbush & Bryk (2002) calculam o percentual da variância explicada pelas medidas contextuais, de acordo com a seguinte expressão:

$$\%VE = \left(\frac{\tau_{00}(\text{não condicional}) - \tau_{00}(\text{condicional})}{\tau_{00}(\text{não condicional})} \right) \times 100 \quad (6)$$

sendo:

$\%VE$ = percentual da variância explicada; T_{00} = componente da variância do intercepto estimado pelo modelo não condicional e pelo modelo condicional (com pelo menos uma variável de nível agregado).

Com o intuito de auxiliar na escolha do modelo mais adequado, os percentuais de variância explicada foram calculados e reportados na última linha da Tabela 1. Comparando-se o modelo 2 (condicional) com o modelo 1 (não-condicional), chega-se à conclusão de que a inclusão da violência escolar explica 1,66% da variância do intercepto, ou seja, 1,66% da variação no desempenho escolar dos alunos da terceira série do ensino médio em matemática são explicados pelos diferentes níveis de violência entre as escolas, dado o efeito fixo (variáveis do nível individual).

Em seguida, no modelo 3, acrescentam-se as variáveis que especificam o percentual de professores efetivos nas escolas. Consequentemente, o percen-

⁷O software estatístico utilizado na estimação do modelo foi o *Hierarchical Linear and Non-linear Modeling* — HLM 6.08. O modelo utilizado não permite o controle da possível presença de endogeneidade entre algumas variáveis explicativas e a variável dependente, configurando-se assim uma limitação do trabalho. No entanto, apesar de tal limitação, acredita-se que o modelo apresenta resultados razoáveis.

Tabela 6: Resultados da estimação do modelo *logit* para o desempenho em matemática da terceira série do ensino médio

Variáveis	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Efeitos Marginais ⁸
Nível individual					
Intercepto	-5,140162*	-5,23954*	-5,649791*	-5,693699*	
Sexo (ref. fe- minino)	0,665155*	0,66598*	0,664569*	0,665355*	1,43
Cor (ref. branco)					
Negro	-0,393777*	-0,394746*	-0,395059*	-0,396588*	-1,23
Amarelo	0,173192*	0,173596*	0,173084*	0,174005*	0,95
Pardo	-0,237264*	-0,237522*	-0,237262*	-0,237695*	-0,50
Indígena	-0,693242*	-0,694945*	-0,695509*	-0,69989*	-6,93
Trabalho (ref. Trabalho1)					
Trabalho2	-0,711313*	-0,71114*	-0,71239*	-0,713012*	-2,30
Trabalho3	-0,262729*	-0,263154*	-0,261916*	-0,262328*	-0,89
Trabalho4	-0,303945*	-0,304479*	-0,303328*	-0,304124*	-1,13
Trabalho5	-0,302434*	-0,303063*	-0,301528*	-0,301662*	-0,94
Trabalho6	-0,369175*	-0,369692*	-0,367484*	-0,368164*	-1,00
Nível Socioe- conômico	0,103155*	0,103086*	0,102863*	0,102801*	0,06
Nível agregado					
Violência	-	-0,261226*	-0,257718*	-0,141005*	-0,54
Professor Efetivo (ref. Profefet1)					
Profefet2	-	-	0,472538*	0,450266*	3,61
Profefet3	-	-	0,737299*	0,699328*	5,33
Profefet4	-	-	0,739263*	0,714058*	5,39
Profefet5	-	-	0,83447*	0,775984*	7,84
Comunidade	-	-	-	0,032077*	0,04
Conservação	-	-	-	0,061518*	0,05
Equipamento	-	-	-	0,024145***	0,03
Efeito Aleatório					
Componente da Variância	0,759	0,747	0,730	0,712	
% Variação Explicada		1,66	3,88	6,25	

Legenda: *, **, *** denotam níveis de significância de, respectivamente, 1%, 5% e 10%.

⁸ Dado que o modelo *logit* é log-linear, para que sejam analisados os efeitos marginais de cada variável sobre a probabilidade de o aluno apresentar um desempenho escolar avançado, foi calculada a razão de chance de o evento ocorrer, que é igual à exponencial do coeficiente estimado. Segundo Noronha & Andrade (2007), esse valor é igual a: $[\exp(\text{coeficiente estimado} \times \text{desvio padrão do coeficiente estimado}) - 1] \times 100$.

tual de variância explicada alcança um patamar mais elevado: 3,88. Por fim, no modelo 4, são incluídas as variáveis que denotam a qualidade da escola: participação da comunidade, conservação de instalações e equipamentos, e existência de equipamentos importantes na escola. A adição destas variáveis foi responsável, em conjunto com as demais variáveis contextuais do nível agregado, pela explicação de 6,25% da variância do intercepto. Portanto, percebe-se que o modelo 4 deve servir de base para analisar o impacto da violência sobre o desempenho escolar, pois o mesmo é o que melhor explica a variância do intercepto.

A partir dessa definição, inicialmente, será analisado o impacto da variável foco deste trabalho — violência nas escolas — sobre o desempenho dos alunos da terceira série do ensino médio em matemática. Conclui-se que uma elevação na violência reduz em 0,54% a probabilidade de que o desempenho dos alunos da terceira série do ensino médio no estado de São Paulo seja satisfatório em matemática. Esse resultado vai ao encontro das análises realizadas por McGarvey et al. (2006), e Carroll (2006). No primeiro trabalho, a violência reduz o desempenho dos alunos em 2,5% (pelo método de MQO) e em 4,0% (pela estimação via variáveis instrumentais); e já no segundo em 0,138% e 0,143% em matemática e leitura, respectivamente.

O que explica tal resultado e existem outras consequências? Como salientado anteriormente por Abramovay & Rua (2002), pode-se concluir que a violência escolar tem consequências perversas sobre a qualidade do ensino e o nível de aprendizagem dos alunos, materializando-se na diminuição no nível de concentração, menor frequência às aulas, além da elevação na desmotivação e rotatividade por parte dos professores. Além disso, a violência escolar afeta indiretamente o nível de capital humano, pois jovens com menor nível de desempenho e aprendizagem tendem a se tornar profissionais menos qualificados, que consequentemente terão menos condições de contribuir no processo de desenvolvimento econômico do país.

Cabe salientar que a literatura mais recente tem levado em consideração um aspecto adicional: a violência pode causar impactos de acordo com o horizonte temporal considerado. Mesmo não analisando especificamente a violência escolar, Chamarbagwala & Mórán (2011) e Justino et al. (2014) utilizaram dados em *coorte* para avaliar efeitos de curto e longo prazo das guerras civis sobre o desempenho dos alunos. Os primeiros autores concluíram que o impacto dos 36 anos de guerra civil na Guatemala reduziu de forma mais severa a escolaridade de homens e mulheres residentes nas zonas rurais e no longo prazo, considerando três *coortes*: 1960-1978, 1979-1984 e 1985-1996. Assim, a escolaridade de homens e mulheres nas zonas rurais da Guatemala decresceu nos três *coortes* citados em 6, 15 e 23%; e 3, 12 e 30%, respectivamente.

Por sua vez, Justino et al. (2014) avaliaram o impacto da violência no Timor Leste em 1999, durante a ocupação da Indonésia, sobre o desempenho dos alunos no ensino primário em 2001 (curto prazo) e 2007 (longo prazo). Os autores concluíram que o impacto é maior sobre os meninos no longo prazo, onde boa parte destes abandonou a escola para se dedicar às atividades militares.

Além da violência escolar, outro aspecto contextual que impacta na proficiência dos alunos em matemática é o percentual de professores efetivos na escola, representado pelas variáveis binárias *profefet2*, *profefet3*, *profefet4* e *profefet5*, sendo omitida a variável de referência — *profefet1* —, que identifica as escolas onde apenas até 25% dos professores são efetivos. Como esperado, todos os coeficientes estimados das variáveis binárias apresentaram

um sinal positivo, o que significa que escolas com uma maior proporção de professores efetivos elevam a probabilidade de os alunos apresentarem um desempenho escolar avançado.

Para se ter uma ideia da importância dessas variáveis sobre a proficiência escolar, o efeito marginal calculado para a variável *profefet5* foi de 7,84, ou seja, o fato de uma escola possuir de 91 a 100% de professores efetivos aumenta em 7,84% a probabilidade de o aluno ter um bom desempenho em matemática.

Cabe ainda sublinhar que se levarmos em consideração que o percentual de professores efetivos é uma boa *proxy* da rotatividade dos professores, confirma-se o resultado encontrado por Severnini (2007). Nesse sentido, uma elevação da rotatividade dos professores diminuiu o nível de performance escolar.

Ainda no nível agregado, as variáveis comunidade, conservação e equipamento, ambas construídas por meio de Análise de Componentes Principais — ACP, possuem a função de qualificar a escola onde o aluno estuda. A variável comunidade representa a participação e contribuição da comunidade nas atividades implementadas pelas escolas. Por sua vez, conservação descreve qual é o estado de conservação das instalações e equipamentos que a escola possui. Já a variável equipamento denota a existência e quantidade de equipamentos disponíveis na escola para utilização.

Os sinais dos coeficientes estimados para essas três variáveis contextuais apresentaram o sinal esperado, ou seja, positivo. Porém, os valores estimados dos efeitos marginais foram relativamente pequenos. Assim, uma elevação das variáveis comunidade, conservação e equipamento é responsável por um aumento de 0,04 0,05%, e 0,03% na probabilidade de o aluno apresentar um desempenho satisfatório em matemática, respectivamente.

Analisados os impactos das variáveis contextuais, é importante verificar os efeitos fixos, ou seja, a relação entre as variáveis do nível individual e a probabilidade de o aluno apresentar um desempenho satisfatório em matemática. A variável sexo do aluno assume 1 se o aluno for do sexo masculino e 0 se for do sexo feminino. Percebe-se por meio da Tabela 1 que o fato de o aluno ser do sexo masculino eleva em 1,43% a probabilidade do mesmo apresentar um desempenho avançado em matemática.

Considerando agora a cor do aluno, foram criadas cinco variáveis binárias: branco (omitida), negro, amarelo, pardo e indígena. Pode-se concluir que o fato de o aluno ser branco eleva a probabilidade de um melhor nível de proficiência em matemática em relação a negros, pardos e indígenas da ordem de 1,23%, 0,5% e 6,93%, respectivamente. Já em relação ao aluno da cor amarela a conclusão é inversa. Se o aluno for amarelo, sua probabilidade de apresentar um melhor desempenho em matemática aumenta em 0,95%.

As variáveis que representam a inserção do aluno no mercado de trabalho são muito importantes, pois como especificado anteriormente por Kasouf (2007), trabalhar desde a infância tende a diminuir a proficiência escolar. Foram incluídas seis variáveis binárias: trabalho1 (omitida), trabalho2, trabalho3, trabalho4, trabalho5 e trabalho6. Como esperado, as variáveis apresentaram um sinal negativo, podendo-se concluir que o fato de o aluno trabalhar diminui a sua probabilidade de apresentar um desempenho satisfatório em matemática. Tal efeito é potencializado no caso da variável trabalho2. Nesse caso, o fato de o aluno trabalhar até 10 horas semanais diminui a probabilidade do mesmo alcançar um resultado adequado em matemática em 2,3%, quando comparado ao aluno que não trabalha.

Por fim, assim como nos trabalhos de Grogger (1997) e Aizer (2008), leva-se em consideração neste estudo o *background* familiar, representado pela variável nível socioeconômico. Tal variável foi construída também via Análise de Componentes Principais — ACP, levando-se em consideração atributos socioeconômicos dos alunos, como escolaridade dos pais, além da posse e quantidade de eletrodomésticos, como rádio, tv, máquina de lavar, freezer, geladeira, etc. O coeficiente estimado dessa variável apresentou um sinal esperado positivo. Nesse sentido, segundo o modelo, um maior nível socioeconômico do aluno eleva sua probabilidade de apresentar desempenho satisfatório em matemática em 0,06%.

6 Considerações finais

O principal objetivo deste trabalho foi o de mensurar o impacto da violência nas escolas sobre o desempenho dos alunos paulistas da terceira série do ensino médio em matemática no ano de 2007. Quando são analisados os fatores que determinam a proficiência dos alunos verifica-se que existe uma vasta gama de fatores explicativos: características individuais dos alunos (sexo e cor), *background* familiar (renda domiciliar e escolaridade dos pais), e características da escola (infraestrutura, qualidade dos professores, e violência).

Percebe-se, assim, que o desempenho escolar dos alunos pode ser influenciado por variáveis de níveis hierárquicos distintos, o que foi negligenciado por diversos trabalhos já realizados. Tal negligência teve como consequência estimações imprecisas, onde ignorou-se a possível existência de falácia ecológica e de falácia atomística.

Diante disso, este estudo contribui de forma relevante na literatura econômica do crime, pois utiliza um modelo *logit* multinível, controlando a relação existente entre violência nas escolas e o desempenho escolar. Nesse caso, considerou-se dois níveis: individual e agregado.

Os resultados do trabalho demonstraram que a violência escolar diminui a probabilidade de o aluno da terceira série do ensino médio apresentar um desempenho satisfatório na prova de matemática. Nesse sentido, a modelagem multinível permite que se conclua o seguinte: um aluno do mesmo sexo, mesma cor e nível socioeconômico similar pode apresentar um rendimento escolar menor caso estude em uma escola com mais violência.

Porém, não é possível afirmar com exatidão de que forma o impacto da violência se materializa sobre o desempenho escolar. No entanto, pode-se inferir que a violência nas escolas diminui o nível de concentração dos alunos, reduz a frequência às aulas e também aumenta a desmotivação e rotatividade por parte dos professores.

Dessa forma, a violência escolar afeta ainda o nível de capital humano, pois crianças e adolescentes que apresentam um menor nível de desempenho escolar e aprendizagem tendem a se tornar profissionais menos qualificados. Consequentemente, eles terão menos condições de contribuir no processo de desenvolvimento econômico do país.

Diante dessa grave consequência surge a necessidade urgente de que sejam implementadas políticas públicas, que reduzam a violência nas escolas, aperfeiçoem o nível de ensino e, consequentemente, elevem o nível de capital humano, tão importante no processo de desenvolvimento econômico de qualquer país.

Referências Bibliográficas

- ABEP (2007), *O novo critério padrão de classificação econômica Brasil/2008*, number 41 p., Campinas.
- Abramovay, M. & Rua, M. (2002), *Violências nas escolas*, Vol. 1 of 400 p., 2. ed. edn, UNESCO, Brasília.
- Aizer, A. (2008), Neighborhood violence and urban youth, Working Paper NBER 13773, National Bureau of Economic Research - NBER, New York. 47p.
- Ammermueller, A. (2007), Violence in european schools: victimization and consequences, Discussion Paper 07-004, Centre for European Economic Research - ZEW, Mannheim. p. 1-40.
- Becker, K. L. & Kassouf, A. L. (2012), Violência nas escolas: uma análise da relação entre o comportamento agressivo dos alunos e o ambiente escolar, in 'Encontro Nacional de Economia. Porto de Galinhas, XL, 2012. Anais... Porto de Galinhas', p. 15 p.
- Carroll, B. (2006), *The effects of school violence and crime on academic achievement*, 33 p., Davidson College.
- Chamarbagwala, R. & Mórán, H. E. (2011), 'The human capital consequences of civil war: Evidence from guatemala', *Journal of Development Economics* 94(1), 41-61.
- Ferrão, M., K., B. & Santos, D. (2002), 'O impacto de políticas de não repetência sobre o aprendizado dos alunos da quarta série', *Pesquisa e Planejamento Econômico* 32(3), 495-514. Rio de Janeiro.
- Gama, V. (2009), Uma análise de relação entre violência escolar e proficiência no município de são paulo, Mestrado em economia aplicada, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.
- Grogger, J. (1997), 'Local violence and educational attainment', *Journal of Human Resources* 32(4), 659-682. Wisconsin.
- Guimarães, A. (1998), *Escola, galeras e narcotráfico*, 248 p., Rio de Janeiro: UFRJ.
- Henrich, C. C., Schwab-Stone, M., Fanti, K., Jones, S. M. & Ruchkin, V. (2004), 'The association of community violence exposure with middle-school achievement: a prospective study', *Journal of Applied Developmental Psychology* 25(3), 327-348. Atlanta.
- Hjalmarsson, R. (2008), 'Criminal justice involvement and high school completion', *Journal of Urban Economics* 63(2), 613-630. Atlanta.
- Hox, J. (1995), *Applied multilevel analysis*, 126 p., T.T. Publikatie, Amsterdam.
- Hurt, H., Malmud, E., Brodsky, N. L. & Giannetta, J. (2001), 'Exposure to violence. psychological and academic correlates in child witnesses', *Archives of Pediatrics Adolescent & Medicine* 155(12), 1351-1356.

Justino, P., Leone, M. & Salardi, P. (2014), 'Short and long-term impact of violence on education: The case of timor leste', *World Bank Economic Review* 28(2), 320–353.

Kassouf, A. L. (2007), 'O que conhecemos sobre o trabalho infantil?', *Nova Economia* 17(2), 323–350. Belo Horizonte.

Lee, V. E. (2001), What are multilevel questions, and how might we explore them with quantitative methods?, in 'Estudos em Avaliação Educacional, São Paulo', Vol. 24, pp. 31–45.

Lee, V. E. (2010), 'A necessidade dos dados longitudinais na identificação do efeito-escola', *Revista Brasileira de Estudos Pedagógicos* 91(229), 471–480. Brasília.

McGarvey, M., Smith, W. J. & Walker, M. B. (2006), The interdependence of school outcomes and school and neighborhood crime, Working Paper 07-19, Georgia State University, Department of Economics, Nebraska. 22 p.

Mingoti, S. (2005), *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*, 292 p., Belo Horizonte: UFMG.

Noronha, K. S. & Andrade, M. (2007), 'O efeito da distribuição de renda sobre o estado de saúde individual no brasil', *Pesquisa e Planejamento Econômico* 37(3), 521–556. Rio de Janeiro.

Ratner, H. H., Chiodo, L., Covington, C., Sokol, R. J., Ager, J. & Delaney-Black, V. (2006), 'Violence exposure, iq, academic performance, and children's perception of safety: evidence of protective effects', *Merrill-Palmer Quarterly* 52(2), 264–287.

Raudenbush, S. W. & Bryk, A. (2002), *Hierarchical linear models: applications and data analysis methods*, 485 p., 2 nd ed. edn, London: Sage.

Riani, J. L. R. & Rios-Neto, E. (2004), Impacto dos fatores familiares, escolares e comunitários na quantidade e qualidade do ensino no estado de minas gerais, in 'SEMINÁRIO SOBRE ECONOMIA MINEIRA, 11., 2004, Diamantina. Anais... Diamantina: Cedeplar', 22 p.

Severnini, E. (2007), A relação entre violência nas escolas e proficiência dos alunos, Mestrado em economia, Pontifícia Universidade Católica do Rio de Janeiro.

URL: Disponível em: <www.maxwell.lambda.ele.puc-rio.br>. Acesso em: 20 ago. 2010.

Shirasu, M. R. & Arraes, R. A. E. (2014), Determinantes da evasão e repetência escolar, in 'X Encontro Economia do Ceará em Debate. Fortaleza'.

São Paulo (2007), 'Secretaria de educação do estado de são paulo. sistema de avaliação de rendimento escolar do estado de são paulo (saresp). são paulo'. 1 cd-rom.

São Paulo (2014), 'Secretaria de educação do estado de são paulo. sistema de avaliação de rendimento escolar do estado de são paulo (saresp). são paulo'.
URL: Disponível em:< www.educacao.sp.gov.br/saresp>.

BRAZILIAN FEMALE LABOR MARKET: RACIAL-SKIN COLOR DISCRIMINATION AND INEFFICIENCY

GUSTAVO ANDREY DE ALMEIDA LOPES FERNANDES *

Resumo

O mito da democracia racial é ainda muito presente na sociedade brasileira, embora a literatura econômica aponte para a existência de um alto grau de discriminação racial. Este artigo visa inferir sua existência nos salários eficientes do mercado de trabalho feminino, levando-se em conta o problema do viés de seleção. As causas da ineficiência são também investigadas. Os resultados mostram que o diferencial de salário eficiente, não explicadas por diferenças no capital humano, é significativo, sendo maior entre as trabalhadoras mais eficientes. Por conseguinte, políticas tradicionais possuem apenas um efeito limitado para reduzir as diferenciais salariais entre raças.

Palavras-chave: Discriminação Racial; Discriminação de Gênero; Fronteira de Produção Estocástica

Abstract

The myth of racial democracy is still widespread in the Brazilian society, although economic literature has continuously documented pervasive racial discrimination. This study analyzes racial discrimination in the Brazilian female labor market using a Mincer stochastic wage frontier, corrected for sample selection. The results, using Oaxaca-Blinder decomposition, showed that an efficient wage differential, not explained by differences in human capital does exist and is large among the most efficient workers. Wage inefficiencies are also explained and it is a claim of the study that traditional policies might be a tool in reducing underpayment dispersion, but cannot cope with pure discrimination.

Keywords: Racial discrimination; Female Labor Market; Affirmative Action

JEL classification: J71

DOI: <http://dx.doi.org/10.1590/1413-8050/ea85456>

* Professor Assistente do Departamento de Gestão Pública da EAESP-FGV. São Paulo, São Paulo, Brasil

1 Introdução

Racial conflicts are almost absent in Brazil and that might be the reason why Brazilians, in general tend to believe that discrimination is not a relevant problem. Freyre (1933), one of the most important Brazilian anthropologists, wrote a book containing the results of his lifelong research, in which he epitomized racial democracy in Brazil. That belief is still subscribed to by the Brazilian elite, which believes that racial democracy continues to prevail in the country¹.

This view has been seriously questioned in the economic literature. Henriques (2001), using the Pesquisa Nacional por Amostra de Domicílios (PNAD) — the national household sample survey carried out annually since 1976 by Instituto Brasileiro de Geografia e Estatística, IBGE, the Brazilian Bureau of the Census, presents strong evidence against Freyre's perception of an ideal racial democracy, as far as economic outcomes are concerned. His study has shown that in Brazil, Afro-Brazilian men earn only 40 to 50 percent of what white men earn. The average monthly income in all regions of Brazil was the highest for white men and the lowest for non-white women, with white women earning more than non-white men in all regions of the country.²

The Brazilian Census has a peculiar way of classifying people, when it comes to race. Actually, people are not classified by race; they are, instead, classified by color. The miscegenation of races resulted in a special color for some people who were classified as *pardo* by IBGE, meaning an individual of mixed race. *Pardos* are treated as persons of African descent. If we add the black and *pardo* people, Brazil has the largest African descent population in the world, outside Africa. For 2005, PNAD data indicated that 47.37 percent of the Brazilian population was white while black and *pardo* accounted for around 52.37 percent of the 187 million inhabitants of Brazil. This means the size of the black population, including persons classified as *pardo*³, is second only to that of Nigeria.

However, if one looks at the distribution of income in Brazil, it is easy to see that the black constitute only about 26.7 percent of the population in the highest income decile, while they represent 59.76 percent of the population in the poorest one. Figure 1 shows the income distribution as per 2005 PNAD data. It clearly shows that the black population in Brazil has a strong presence among the poor and a very weak presence among the high-income brackets.

Barros et al. (2000) and Ferreira (2000) argue that differences in educational attainment are a major factor contributing to unequal distribution of income in Brazil. The presence of a large proportion of blacks among the poorest might be because education levels are very different between racial groups. The discrepancy in years of schooling between blacks and whites is 2.13 years, according to 2005 PNAD data. It might look relatively small but it is rather large when compared to the seven years of average schooling for

¹Broadly speaking, racial democracy is the conception of an ideal society, of equals, in which the color of the skin presents no barriers for social mobility. The opposite conception is discrimination based on social classes. For a detailed description about the Brazilian racial debate, see Telles (2004).

²Average monthly income in 2005 was Real the Brazilian currency) 1077.27 for white men, 679.83 for white women, 549.29 for nonwhite men and 356.89 for non-white women (25 years or older).

³This approach will be followed in the rest of the article.

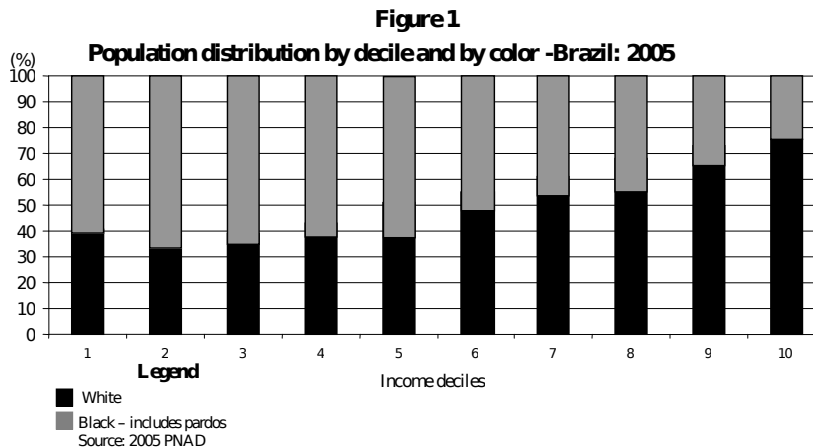


Figura 1: Population distribution by decile and by color

the entire population above 25 years of age. The picture becomes clearer when one looks at the evolution of education among groups. Defining a cohort as someone born between 1929 and 1974, average years of schooling for the whole population increased steadily, but the difference in years remained constant between blacks and whites throughout this period.

The differences between skin color groups become even more pronounced when a gender distinction is made. Kassouf (1998) shows that even after controlling for human capital, measured by educational attainment, a woman could make only 75 percent of a man's salary in Brazil in 1995. More recently, Carvalho et al. (2006) found that 97% of the difference between white men and black women wages is due to discrimination. Both papers, showing gender wage differentials favoring men, support the view presented in this study - i.e. if both race and gender differences are taken into account, then black women are likely to be at the bottom of the Brazilian socio-economic ladder. In fact, the average wages of black women in 2005, according to PNAD, were 64 percent of those of a black man and only 33 percent of the average salary of a white man.

Racial discrimination, from the perspective of economics, is measured by wage differentials unexplained by differences in contribution to economic outcomes, which are usually explained according to a scheme presented by Mincer (1970, 1974) that became known in the literature as the Mincer Equation. Oaxaca (1973) and Blinder (1973), working with Mincer's formulation, developed a methodology to decompose the wage gap among groups of persons into a first component attributed to individual characteristics (or human capital) that affect productivity and another component associated with discrimination. This framework has also been used to analyze variants of decomposition techniques by simulating counterfactual distributions, combining data on individual characteristics from one distribution with estimated parameters from another.

Zuchi & Hoffman (2004) and Zuchi (2006) show that wage differentials between white and blacks⁴, unexplained by human capital assets as measured

⁴They also include *pardos* as part of the Black population.

by years of education and experience between whites and blacks, could be as high as 30.9 percent for 2003. Arcand & d'Hombres (2004) worked with data of the 1998 survey covering men in the 25-65 age group. Applying basically the Mincer Oaxaca framework but using a quantile Mincer-Oaxaca regression analysis, they encountered wage differential favoring male workers that goes from 38.95 percent for the first decile up to 48.1 percent for the last decile of the wage distribution. Their results show rising wage differentials as one goes up in the wage distribution scale and the wage gap. According to Arcand and d'Hombres, discrimination is more prevalent as the color of the skin gets darker. Matos & Machado (2006), also using PNAD data, applied the Oaxaca decomposition to Mincer equations and found that from 1987 to 2001, race and gender had a significant contribution to explain wage inequalities in Brazil.

The brief literature review has documented pervasive racial discrimination in Brazil and there is little doubt that dark skin is a handicap to social advancement. Our aim is to go even further, analyzing the Brazilian female labor market, so that discrimination will be the result of the Oaxaca decomposition in an estimated efficient Mincer wage function corrected for selection bias. Simultaneously, the factors that might contribute to pulling black female workers away from their respective Mincer efficient frontiers will be presented.

After this introduction, estimation procedures are in the second part detailed. Data is described in the following section. PNAD 2005, a nationwide household survey is used. In the fourth part, before the conclusion, results are presented. Racial discrimination seems to be very important in Brazil even when one looks to efficient wages, which are larger the more efficient the worker is. This analysis adds to the evaluation of policy instruments that could be used to improve efficiency as well as reduce wage disparities within wage groups, leading to improvement in the wage distribution. Moreover, if a wage gap attributed to racial discrimination is encountered, it can be argued that the effect of traditional remedies, very often proposed to eliminate or diminish discrimination, reach a bound given by the estimated value of pure form of discrimination. In this sense, even the most efficient black worked would still be discriminated.

2 Conceptual Framework

The methodology used to infer discrimination in the labor market relies heavily on the stochastic frontier model which has been traditionally used in the framework of studies concerning productive efficiency ever since Aigner et al. (1977) and Meeusen & Van Den Broeck (1977) defined and used the concept in a simultaneous and yet independent way. One pioneering application of this technique in labor economics was that of Robinson & Wunnava (1989). Other prior studies include Hofler & Murphy (1992, 1994). The former estimated the extent to which effective wages of workers are below the potential maximum they could earn, given their marginal productivity, while the latter estimated the worker's reservation wage. Both papers adopted the framework of search theory. McClure et al. (1998) also take up the question of wage inefficiency but compare Stochastic Frontier estimations for the United States and Canada. Dawson et al. (2001) analyze the question of relative underpayments

to male and female workers. Prieto et al. (2003) use the Stochastic Frontier wage model to estimate returns to education for potential and effective wages. They estimated the impact of formal schooling on workers potential maximum wages. Lang (2005), analyzing discrimination against migrant labor in Germany, claims that human-capital endowments determine earning potentials rather than actual earnings, and uses stochastic earnings frontier to estimate differences between natives and migrants for the year 2000.

This article adds to this body of literature by using the Oaxaca decomposition in a Mincer frontier wage function corrected for sample selection bias. It is our correction for sample selection that makes it possible to obtain consistent estimates of the Stochastic Frontier function, thus allowing an Oaxaca decomposition based on a consistent estimation procedure. Furthermore, by separating efficient from inefficient wages, the analysis allows a better understanding of the unexplained wage differential between groups of laborers.

2.1 The Stochastic Frontier

The starting point of our estimation procedure will be the definition of a Stochastic Frontier Mincer (log) earnings function, obtained from a set of variables, all of them reflecting their income generating inputs according to the following specification:

$$\ln(w_i^*) = x_i' \beta + v_i \quad (1)$$

Where w_i^* is the potential wage earned by the worker, x_i is a column vector of worker characteristics, and occupational and regional differences in labor markets, β is a vector of parameters associated with each variable to logarithm of potential or efficient wage earned by the worker and v_i a vector of parameters associated with each variable to logarithm of potential or efficient wage earned by the worker and:

$$\ln(w_i) = \ln(w_i^*) - u_i \quad (2)$$

where u_i is a random, non-negative variable.

The potential wage of each worker is the frontier wage or the efficient wage. Effective or observed wage is given by the Stochastic Frontier minus the non-negative random variable that captures the difference between potential and effective wage:

$$\ln(w_i) = x_i' \beta + v_i - u_i \quad (3)$$

The earning function is perceived as a frontier. It represents the potential or maximum wage that a worker would receive given her capital endowment as well as occupational and regional characteristics of the labor market. The non-negative random term u_i represents the extent to which observed earnings fall short of potential earnings.

The difference between efficient and observed wage a worker earns is explained by a set of variables reflecting individual and labor market characteristics, $S_i = (S_{i1}, \dots, S_{im})$ and a new random term such that:

$$u_i = S_i' \delta + \eta_i \quad (4)$$

where $\delta' = (\delta_1, \dots, \delta_m)'$ is a vector of parameters to be estimated, $u_i \geq 0$ and $\eta_i(0, \sigma_u^2)$ are truncated at the point $-S_i' \delta$, in such a way that $\eta_i \geq -S_i' \delta$. Under these assumptions, u_i is a normally distributed variable with its mean depending on the specific explanatory variables reflecting individual and labor market characteristics and truncated at zero, such that $N^+(S_i' \delta, \sigma_u^2)$. Following Aigner et al. (1977), v_i is an idiosyncratic effect specific to an individual worker that could enter the model with either sign, such that we can suppose $v_i(0, \sigma_v^2)$, independent of and of the regressors.

The end result is the Mincer Stochastic Frontier model represented by equation 3 above. The frontier or potential wage is given by equation 1. The inefficiency term is given by u_i , a random variable of particular interest on this setting. Since the data is in log terms, u_i is a measure of the percentage by which the particular individual fails to achieve the particular potential ideal wage rate. The final Mincer Stochastic Frontier model (Equation 3) has a composed error term $\varepsilon_i = v_i - u_i$ ⁵

2.2 Sample Selection

Once a distribution for u_i is assumed, the next step is the definition of the likelihood function for ε_i and to obtain a set of parameters β and δ that maximizes the likelihood. However, there is an additional problem that has to be dealt with in order to obtain consistent estimates of parameters β : sample selection.

In this study, a two-step Heckman procedure will be used to take into account the selection process that might be present in our estimation of the Mincer Stochastic Frontier equation. In the first step of the Heckman procedure, the selection process, responsible for selection bias problems is studied with the so-called selection model. The bias is caused by existence of differences between employed and non-employed (hence earn no wages) women, which are related to their income. So it is necessary to compare these groups (employed and non-employed women) to find out what the differences are. For this purpose, generally, a probit model is estimated⁶.

Following the economic literature, in our study, the dependent variable in the probit analysis is a dummy variable indicating whether or not the female is employed. Independent variables in the model are the relevant characteristics of the respondents, available in the data set, such as education, age, formal or informal jobs, number of children 10 years old or younger, and the total number of children in the household. In the probit analysis, we estimate the effects of these variables on the employment decision. Besides that, the effect of unmeasured characteristics which are related to the employment decision

⁵To complete the specification of the model and to allow estimation is required a specification of the distribution of the error term. Kumbhakar & Lovell (2003) suggest three possible distributions for the inefficiency term: exponential, half normal and truncated normal. We are going to use the truncated normal, differently from Lang (2005) who used the exponential distribution and afterwards explained the estimated inefficiency as function of explanatory variable in a two-step estimation procedure. We follow Prieto et al. (2003) argument favoring a joint estimation of the frontier parameter vector and of the parameter vector associated to the variables explaining inefficiency for. However, as shown by Greene (2002), the estimated inefficiency differs very little when different distributions are assumed for . The distribution will matter for the specification of the likelihood function to be used in the estimation procedure.

⁶We assume that the error term of this model is normally distributed, one of the assumptions underlying the Heckman model

process can be estimated through the residuals of the probit analysis. After all, the variation which remains in the dependent variable after removing the effect of the known factors can only be caused by the influence of unknown factors. That is the reason why in the Heckman (n.d.) procedure, the residuals of the selection equation are used to construct a selection bias control factor, which is equivalent to the Inverse Mills Ratio⁷.

In the second step of the Heckman procedure, Stochastic Frontier analysis of the effects of human capital on the logarithm of wage is carried out, which we call the Mincer Stochastic Frontier equation. In this analysis, we use the Inverse Mills Ratio as an additional independent variable. Because we now have a control factor in the analysis for the effect of the logarithm of wage related unmeasured characteristics, which are also related to the employment decision, the human capital predictors in the Mincer Stochastic Frontier equation are free from this effect: the Mincer Stochastic Frontier analysis, therefore, produces unbiased coefficients.

2.3 Limitations

Before going to the description of the practical estimation of the Heckman model, a word of caution should be brought forward. Although theoretically the procedure sounds appropriate, applying it in practice is not so simple. An important condition for its use is that the selection equation contains at least one variable which is not related to the dependent variable in the substantial equation, otherwise severe problems of multicollinearity may arise and the addition of the correction factor to the substantial equation may lead to difficulties in estimation and to unreliable coefficients.

Formalization of the Mincer Stochastic Frontier model, corrected for sample selection, follows Greene (2003). Assuming that the sample selection follows a general framework, the equation that determines the sample selection process is⁸:

$$\ln(w_i^l) = z_i' + \xi_i \quad (5)$$

while the equation of primary interest is equation (2). The sample selection rule is that wage is observed only if $w_i^l > 0$, which in our analysis means that the observed wage is greater than the reservation wage, and the individual is participating in the labor force. Let us assume that v_i and ξ_i have a bivariate normal distribution with zero mean and correlation ρ , and that

$$E(u_i \xi_i) = E(v_i u_i) = 0 \quad (6)$$

Under those assumptions, the estimated Inverse Mills Ratio is introduced in the Mincer Stochastic Frontier as an additional regressor to correct for sample selection. The equation in the second step is estimated by maximum likelihood, with the Inverse Mills Ratio as additional regressor.

$$\ln(w_i | z_i^* > 0) = x_i' \beta + \lambda \frac{\phi(z_i' \gamma / \sigma_\xi)}{\Phi(z_i' \gamma / \sigma_\xi)} + v_i - u_i \quad (7)$$

⁷For a definition of the Inverse Mills Ratio, see Greene (2002).

⁸ Z_i is a column vector of worker characteristics, and occupational and regional differences in labor markets. It shares the same variables as X_i and has at least one additional variable which must be not related to the dependent variable in the substantial equation.

This new model will henceforth be called the Mincer-Heckman stochastic earning frontier model. It can be estimated for each of the two groups of female workers: black and white. Furthermore, for each group of laborers, the parameter vector δ , associated with the mean vector of the truncated normal distribution, is jointly estimated. This estimation is carried in such a way that we can have an estimate of the percentage return on human capital for individuals on the efficient frontier, corrected for sample selection, as well as estimation of the contributing factor to inefficiency in each of the two female labor markets. Once the Mincer-Heckman stochastic frontier is estimated, the next step is to observe if wage differential between black and white still persists.

2.4 Discrimination

Economic discrimination is a broad and inaccurate concept, despite the existence of large differences between population groups which are elaborately empirically documented, showing great stability in due course. A widely accepted definition comes from the fact that identical objects have the same price. Thus, considering the workforce as a commodity, two individuals with identical productivity and in this sense equal, should have the same wage.

Discrimination occurs when workers with the same productive abilities but belonging to different population groups, earn different wages, invalidating the hypothesis of neutrality of trade. Therefore, economic theories that try to explain the phenomenon of discrimination aim at finding the precise conditions in the labor market in which identical goods are priced distinctly.

The job offer for two individuals is considered essentially identical if they display the same degree of productivity in the material process of production of goods. The idea of productivity in this context should be seen in a broad manner, including a variety of factors, such as work attendance, cooperation skills, and expertise — in addition to the actual expected future productivity, among other things.

Evidently, equality of economic agents makes sense only in relation to variables that are exogenous to the process of discrimination. Be a vector of productive characteristics that determine the productivity and observable by the business as well as being exogenous, a dummy that adopts the value 1 if the individual belongs to the segment of the measured population.

$$\alpha = (Y|X, Z = 1) - (Y|X, Z = 0) \quad (8)$$

In a fully egalitarian society α is equal to zero, where Y is the expected value of Y , conditioned by X , i.e., all-or-equal, with the exception of the social group.

2.5 Oaxaca-Blinder Decomposition

Oaxaca-Blinder decomposition is used to estimate the wage earning differential that can not be attributed to differences in human capital endowment. As it was just shown, unexplained differences in wages between the two groups are usually attributed to racial discrimination. If separate wage regression equations are fitted for black and white female workers, the underlying regression models are:

$$\ln(w_{w,i}) = x_{w,i}'\beta_w + \varepsilon_{w,i}, i = 1, \dots, n_w \quad (9)$$

$$\ln(w_{b,i}) = x_{b,i}'\beta_b + \varepsilon_{b,i}, i = 1, \dots, n_b \quad (10)$$

where n_w and n_b represent black and white female samples, respectively.

Regressor vectors include socio-demographic variables, such as occupation, region, and human capital variables. A comparison of these two regressions allows us to perform what has been known in the literature as Oaxaca (1973) decomposition, which is based on a comparison of regression functions. For any two vectors of characteristics,

$$\ln(w_{w,i}) - \ln(w_{b,i}) = x_{w,i}'(\beta_{w,i} - \beta_{b,i}) + (x_{w,i} - x_{b,i})'\beta_b \quad (11)$$

The second term in this decomposition is identified with differences in human capital that would explain wage differences due to differences in human capital endowment between white and black female laborers, assuming that labor markets respond to these differences naturally. The first term shows the differential in log wages that is attributable to differences unexplainable by human capital; holding these factors constant at x_w makes the first term attributable to other factors, such as discrimination. Therefore, it is not easy to see that the first term can be used to estimate α :

$$\ln(w_{w,i}) - \ln(w_{b,i}) = x_{w,i}'\alpha + (x_{w,i} - x_{b,i})'\beta_b \quad (12)$$

In this study, the Oaxaca decomposition is performed using estimates generated by the Mincer-Heckman frontier functions for white and black female workers as estimators for β_w and β_b . The difference between this procedure and Oaxaca's is that here, Mincer-Heckman Stochastic Frontiers are used to perform an Oaxaca style decomposition for estimating potential wage for both groups of women, while the conventional procedure uses OLS estimation or Heckman two stage estimation. If the hypothesis that is rejected, then the data shows evidence of wage differentials not attributable to differences in human capital between the two groups of laborers which can be considered an estimate of pure racial discrimination, based on estimated potential wage for black and white female workers, corrected for selection bias.

Dispersion of efficiency within each female group still remains to be explained. Inefficiency dispersion is analyzed with the estimated underpayment for black and white female workers, derived from the estimation of the Mincer-Heckman Stochastic Frontiers. Estimation of the frontiers follows the Kumbhakar and Lovell (2003) procedure, assuming that the random inefficiency term u , in Equation 6 follows a truncated normal distribution⁹.

3 The Data

Our empirical analysis uses 2005 PNAD data. After data cleaning, when observations with any missing information were eliminated, the sample was reduced to 140,078 individuals, 25 years or older. Out of this sample, 61,183

⁹The normal-truncated normal model used in the estimation of the Mincer Stochastic Frontier and the resulting likelihood function is derived by in Kumbhakar & Lovell (2003). Once the log likelihood is maximized with respect to the unknown parameters, all the information necessary to estimate underpayment for each individual in the sample is at hand. The approximated formula to estimate the underpayment was developed by Jondrow et al. (1982).

were female workers who constituted the sample used in the study, with 30,489 classified as white and 30,551 black, based on the reported color of their skin. Table 1 and 2 present a summary description of variables for all female workers used for model estimation. Their average monthly wage was 519.95, average years of schooling was 9.0, meaning that the mean female worker was almost finishing high school (in Brazil it takes 10 years to finish high school). 12.2 percent of female workers had at least 15 years of schooling, which is required to complete some college study.

Tabela 1: Variables Description

Wage ⁽¹⁾	Wage in Reales
Educ	Years of Formal Education
Educ17 ⁽²⁾	Individual has at least under graduation degree
Exper	Years of experience in the labor market
SE	Individual lives in the Southeast Region
N	Individual lives in the North Region
NE	Individual lives in the Northeast Region
S	Individual lives in the South Region
CW	Individual lives in the Center West Region
Urban	Individual lives in an urban place
Maid formal	Individual works formally as a servant in someone's home
Maid informal	Individual works informally as a servant in someone's home
Formal Labor	Individual has a formal job
Age	In years
Unionized	Individual is a members of a union
Children	Number of Children
Indust. Sector	Individual works in the industrial sector
Child<10 ⁽³⁾	Individual has a less-than-10-year child
Dependent - Partner	Individual is the partner of the head of the family
Dependent - Son	Individual is the son of the head of the family
Dependent - Others	Individual is neither the partner nor the son of the head of the family

The average number of children was 1.8; 17.3 percent were working in domestic servant type of jobs that can be broadly described as “domestic assistance”; 4.3 percent had formal employment; and 16.7 percent were unionized. The northeast and southeast regions have high concentration of female workers and accounted for close to 60 percent of the female labor force. Northeast is the least developed region of Brazil and the southeast the most developed, as measured by per capita income. Thirty-nine point seven percent of black women had at least one child 10 years old or younger, while the figure for white female workers was 35 percent.

Table 2 also presents the same results, for white and black female workers, respectively. Average wage for white female workers was 759.14 reals while for black female workers it was 396.49 reals. A white female worker, on average, earns 1.9 times more money per month than an average black female. If one looks at education and experience, years of education of whites is 26 percent higher than for blacks. Higher education, defined as more than 15 years of schooling, is observed in 20.7 percent of whites in the sample while only 7.5 percent of blacks had higher education. Thus, white females have some edge in terms of human capital. If this is a possible explanation for the huge wage gap between white and black female workers, much still remains to be seen and we hope to shed some new light on this important social issue.

One important limitation for female labor is motherhood; raising small

Tabela 2: Data Summary

Variable	Female Labor Market		White Female Labor Market		Black Female Labor Market	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Wage ⁽¹⁾	514.946	958.755	759.143	1334.816	396.499	637.159
Educ	8.986	4.447	9.783	4.618	7.715	4.540
Educ17 ⁽²⁾	0.122	0.327	0.207	0.405	0.075	0.263
Exper	21.433	14.631	25.784	13.123	26.248	13.221
SE	0.290	0.454	0.347	0.476	0.245	0.430
N	0.123	0.329	0.066	0.249	0.172	0.378
NE	0.304	0.460	0.188	0.390	0.410	0.492
S	0.173	0.378	0.295	0.456	0.057	0.233
CW	0.110	0.312	0.103	0.304	0.116	0.320
Urban	0.819	0.385	0.124	0.329	0.183	0.387
Maid formal	0.042	0.200	0.038	0.192	0.056	0.231
Maid informal	0.131	0.337	0.091	0.288	0.153	0.360
Formal Labor	0.252	0.434	0.285	0.451	0.193	0.395
Age	36.276	13.333	41.157	11.336	40.584	11.262
Unionized	0.168	0.374	0.209	0.407	0.177	0.382
Children	1.778	1.288	1.622	1.124	1.924	1.410
Indust. Sector	0.125	0.331	0.139	0.345	0.111	0.314
Child<10 ⁽³⁾	0.339	0.473	0.350	0.477	0.397	0.489
Dependent - Partner	0.480	0.500	0.576	0.494	0.541	0.498
Dependent - Son	0.226	0.418	0.046	0.209	0.053	0.224
Dependent - Others	0.068	0.252	0.118	0.322	0.117	0.322
	(N=61,183)		(N=30,489)		(N=30,694)	

⁽¹⁾Wages are measured in Brazilian Currency Reals; the average black market rate for the dollar was 2.7 Reals per dollar

⁽²⁾Indicates individuals with 15 or more years of schooling

⁽³⁾Women having at least one child 10 years old or younger

children demands intense mother care, making it difficult for women to participate in the labor market. As it has been shown by several studies (e.g., Heckman (n.d.), Kassouf & Silvia (2000) and Killingsworth & Heckman (1986)), the presence of small children imposes an upward drift in female reservation wage. If one omits this aspect of selection of sample of female labor supply, one can obtain only biased estimates of the effect of wages on the female labor supply. As it is possible to see in Table 2, 39.7 percent of black female workers have children 10 year old or younger; in case of white female workers, only 35 percent have small children. Another important fact evident from the table is that a high percentage of black women are working as maids. Maid is a typically unskilled labor job and it is a service used widely by Brazilian households. Maids represent 20.9 percent of black female jobs but only 12.9 percent of white female jobs.

This table reinforces what has already been said in the introduction. Black females are in the lower strata of the Brazilian labor market. Their wages, on average, are lower; their education is lower, and they are over represented in low skill, low wage jobs.

4 Results

4.1 Mincer Equation with Heckman Sample Selection Correction

10

The first part of the analysis deals with sample selection. The results of the estimation of the Mincer equation, corrected for sample selection, are presented in Table 3. It is possible to immediately see that the estimation of Mincer equation without correction for sample selection leads to inconsistent OLS estimates. Likelihood-ratio statistics strongly reject independence of the error term of the selection model from the error term of the wage equation for black, as well as for white female workers. The identifying restriction is the presence of young children in the household, preventing women from being employed if the woman is married. Under these circumstances there might be a larger probability for reservation wage to be above the market wage. This result follows previous findings in Brazil, analyzed by Kassouf (1998), and for other economies, as presented by Killingsworth & Heckman (1986), among others.

Presence of young children has a negative and significant impact on participation of black and white female workers in the labor market. Marriage also contributes to lower labor force participation for black as well as for white female laborers.

4.2 Mincer Stochastic Frontier

Estimation of the Mincer equation in a Stochastic Frontier framework has been used as an instrument to analyze underpayment, defined as:

$$\ln(\text{wage}_{ik}^{\text{obs}}) - \ln(\hat{\text{wage}}_{ik}^l) = \hat{u}_{ik}, i_k = 1, \dots, n_k; k = \text{black}, \text{white} \quad (13)$$

where $\ln(\text{wage}_{ik}^{\text{obs}})$ is observed wage and $\ln(\hat{\text{wage}}_{ik}^l)$ is estimated frontier wage, i. e. the maximum wage a female worker can earn given her human capital endowment; \hat{u}_{ik} is a measurement of underpayment, i. e., how much more she would be able to make if she were getting efficient or frontier wage.

Table 4 shows the estimation of Mincer Stochastic Frontier without the correction for sample selection bias for black and white female laborers. Both estimated Mincer frontier present a somewhat larger return to education than the Mincer equation corrected for sample selection bias reported in Table 5¹¹. One year of education, in the Mincer equation of Table 5, leads to a return of 5.9 percent for blacks while without correction for sample selection, the return would be 6.3 percent; for white workers, with the Mincer equation corrected for sample selection bias, one year of education had a return of 9 percent while in the Mincer Stochastic Frontier, the return was 9.2 percent.

¹⁰In a first analysis, Wald testes were carried with the null hypotheses that coefficients are the same for black and white women. For this, pooled regressions were estimated; a dummy for black and interaction dummies with all explanatory variables were created. In all cases, racial equality was rejected.

¹¹The significance levels presented in this Table are corrected standard errors due to the use of estimated regressors as explanatory variable in this case the estimated Inverse Mills Ratio for white and black female workers.. To obtain a standard error asymptotically sound two procedures were carried out: The first was to bootstrap the standard error of all coefficients, and the second was to derive analytically the formula for the standard error, afterward deriving their sample values. The bootstrapped values were derived using 500 replications. For the analytical derivation of the standard errors see Murphy & Topel (2002), Greene (2002).

One additional year of superior education would result in return of 58.9 percent for blacks and 47.2 for whites. For black workers, one year of college education would yield a return that is ten times higher than one additional year of education below college level. For whites, the ratio between return of one year of college education and one year of education before college is around five.

The estimated coefficient for the Inverse Mills Ratio is highly significant for both groups of female laborers. This suggests that sample selection is an important factor that makes up for omission of a latent structure of a Heckman type selection.

Table 6 reports the regression results for underpayment. Underpayment means that workers with equal capital endowment earn wages smaller than frontier or potential wages, due to being separated by race. Inefficiency or underpayment decline with more education for both black and white workers. Formal labor relations improve efficiency, though the impact is stronger for black workers. Unionization is an instrument for white workers to reduce inefficiency and yet it is not an instrument for reduction of inefficiency for black workers. Children of any age are a drawback in terms of efficiency. It is very likely that children are a factor that makes women accept a smaller than potential wage. Maids, when informal, contribute positively to the underpayment of women, either white or black.

4.3 Oaxaca Decomposition Results

The main objective of this paper is to present a measure of pure racial discrimination, defined by the first term of the right hand side of equation 10, as discussed above. To achieve this, we used the estimated coefficients of the second stage Mincer-Heckman equations for white and black female workers. The value of pure discrimination is obtained using average sample values of the explanatory variables for the white workers group multiplied by the difference between the estimated frontiers' coefficients for white and for black female laborers. In other words, after taking into account differences in human capital endowment and frontier's estimated coefficients, an unexplained difference ranging from 19.6 to 14.8 percent still remains, when one goes from the highest efficient or potential wage percentile to the lowest.

In order to understand these differences, as we move along the frontier, we compare only efficient wages while separating the sample by percentiles. Table 7 presents the results. As we can observe in the table, Oaxaca difference unexplained by human capital tends to enlarge as efficient wages increase along the frontier. For the highest 1 percent wages, pure discrimination amounts to a 19.6 percent wage differential, while for the 90 percent percentile, the Oaxaca difference falls to 14.8 percent wage differential. Therefore, results show that even the most efficient black worker is discriminated in the Brazilian female labor market. Worse than that, the more efficient she is, the more discriminated she will be.

One could ask why this study separates black and white workers in labor markets. The answer is very simple: if estimated wage differences were small enough, and only individual human capital endowments were able to explain wage differentials, then one could treat all females, black and white, as a distinct labor market. However, if after taking care of differences in human capital endowments, unexplained difference is still large, then to treat

Tabela 3: Estimated Mincer Equations with Sample Selection

	Black Female Worker		White Female Worker	
	Log Wage	Labor Force	Log Wage	Labor Force
Educ	0.087 (0.002)**	0.097 (0.003)**	0.103 (0.002)**	0.078 (0.003)**
Educ17	0.546 (0.025)**		0.475 (0.019)**	
Age	0.053 (0.003)**	0.070 (0.007)**	0.056 (0.004)**	0.071 (0.007)**
Age sq	-0.001 (<0.001)**	-0.001 (<0.001)**	-0.001 (<0.001)**	-0.001 (<0.001)**
N	-0.032 (0.017)	-0.313 (0.037)**	-0.044 (0.023)	-0.265 (0.051)**
NE	-0.371 (0.014)**	-0.195 (0.031)**	-0.313 (0.016)**	-0.256 (0.036)**
S	0.054 (0.024)*	-0.159 (0.053)**	0.049 (0.014)**	-0.311 (0.032)**
CW	0.107 (0.018)**	0.032 -0.045	0.087 (0.019)**	-0.096 (0.047)*
Married		-0.324 (0.056)**		-0.23 (0.065)**
Number Children at Home		-0.037 (0.011)**		-0.043 (0.014)**
Child < 10		-0.028 (0.031)		-0.063 (0.034)
Urban		1.248 (0.024)**		1.299 (0.028)**
Ind. Sector		0.772 (0.047)**		0.922 (0.049)**
Dependent — Partner		-0.704 (0.053)**		-0.729 (0.062)**
Dependent — Son		-0.426 (0.067)**		-0.414 (0.082)**
Dependent — Others		-0.165 (0.078)*		-0.394 (0.083)**
Constant	3.889 (0.078)**	-0.722 (0.169)**	3.798 (0.081)**	-0.591 (0.184)**
Observations	25545	25545	24432	24432
	LR test of indep. eqns. (rho = 0):		LR test of indep. eqns. (rho = 0):	
	chi2(1) = 123.85		chi2(1) = 53.94	
	Prob > chi2 = 0.0000		Prob > chi2 = 0.0000	

Standard errors in parentheses: * significant at 5%; ** significant at 1%

black and white females as different components of the markets is a correct stand and it is a sign that racial discrimination exists and has to be dealt with in empirical analysis.

In fact, results show that a form of pure discrimination still persists in the Brazilian female labor market. On average, there still remains an unexplained wage gap of 13.3 percent. We also encounter rising pure discrimination, when one moves along the efficient wage frontier. Consistent with these findings, the estimated Mincer-Heckman Stochastic frontiers have shown that education is a factor that improves potential wage for blacks as well as for whites; however, the impact of education is greater for whites than for blacks, such that education improves potential wages at a faster rate for whites than for blacks, which leads to increasing differences in potential wages as human capital increases.

Furthermore, when one looks at inefficiency in women labor markets, one can observe that several factors contribute to underpayment to both white and black female laborers in Brazil; children make women, black or white, accept less than potential wages; however, the impact of underpayment is more severe in case of black women. Unionization contributes positively to reduce underpayment of white women but it does not affect underpayment of

Tabela 4: Mincer-Heckman Stochastic Frontier Estimation Without Sample Selection Correction

	Black Female Worker Log Wage		White Female Worker Log Wage	
	SE	Bootstrapping	SE	Bootstrapping
Educ	0.063 (0.002)**	0.063 (0.002)**	0.092 (0.002)**	0.092 (0.002)**
Educ17	0.579 (0.020)**	0.579 (0.025)**	0.467 (0.017)**	0.467 (0.018)**
Age	0.034 (0.004)**	0.034 (0.004)**	0.040 (0.004)**	0.040 (0.005)**
Age sq	-0.001 (<0.001)**	-0.001 (<0.001)**	-0.001 (<0.001)**	-0.001 (<0.001)**
N	0.018 (0.014)	0.018 (0.014)	-0.029 (0.021)	-0.029 (0.020)
NE	-0.237 (0.011)**	-0.237 (0.011)**	-0.267 (0.015)**	-0.267 (0.016)**
S	0.006 (0.019)	0.006 (0.019)	0.015 (0.012)	0.015 (0.013)
CW	0.116 (0.015)**	0.116 (0.015)**	0.111 (0.017)**	0.111 (0.017)**
Union	0.245 (0.015)**	0.245 (0.016)**	0.183 (0.015)**	0.183 (0.016)**
Urban	0.066 (0.020)**	0.066 (0.019)**	0.076 (0.026)**	0.076 (0.023)**
Constant	4.670 (0.085)**	4.670 (0.084)**	4.362 (0.095)**	4.362 (0.102)**
Observations	20700		20643	

Standard errors in parentheses * significant at 5%; ** significant at 1%

Tabela 5: Mincer-Heckman Stochastic Frontier Estimation Results

	Black Female Worker Log Wage			White Female Worker Log Wage		
	SE	Corrected SE	Bootstrapping	SE	Corrected SE	Bootstrapping
Educ	0.059 (0.002)**	0.059 (<0.001)**	0.059 (0.002)**	0.090 (0.002)**	0.090 (<0.001)**	0.090 (0.002)**
Educ17	0.589 (0.020)**	0.589 (<0.001)**	0.589 (0.023)**	0.472 (0.017)**	0.472 (<0.001)**	0.472 (0.019)**
Age	0.031 (0.004)**	0.031 (<0.001)**	0.031 (0.004)**	0.037 (0.004)**	0.037 (<0.001)**	0.037 (0.005)**
Age sq	< 0.001 (<0.001)**	< 0.001 (<0.001)**	< 0.001 (<0.001)**	< 0.001 (<0.001)**	< 0.001 (<0.001)**	< 0.001 (<0.001)**
N	0.028 (0.014)*	0.028 (<0.001)**	0.028 (0.015)	-0.024 (0.021)	-0.024 (<0.001)**	-0.024 (0.020)
NE	-0.231 (0.012)**	-0.231 (<0.001)**	-0.231 (0.012)**	-0.262 (0.015)**	-0.262 (<0.001)**	-0.262 (0.015)**
S	0.011 (0.019)	0.011 (<0.001)**	0.011 (0.017)	0.022 (0.013)	0.022 (<0.001)**	0.022 (0.012)
CW	0.115 (0.015)**	0.115 (<0.001)**	0.115 (0.015)**	0.112 (0.017)**	0.112 (<0.001)**	0.112 (0.017)**
Union	0.244 (0.015)**	0.244 (<0.001)**	0.244 (0.016)**	0.182 (0.015)**	0.182 (<0.001)**	0.182 (0.016)**
Urban	-0.025 (0.028)	-0.025 (<0.001)**	-0.025 (0.028)	0.010 (0.036)	0.010 (<0.001)**	0.010 (0.033)
Blackmillsratio	-0.171 (0.036)**	-0.171 (<0.001)**	-0.171 (0.034)**	-I	-	-
Whitemillsratio	-	-	-	-0.120 (0.044)**	-0.120 (<0.001)**	-0.120 (0.042)**
Constant	4.862 (0.093)**	4.862 (0.001)**	4.862 (0.090)**	4.500 (0.108)**	4.500 (0.001)**	4.500 (0.117)**
Observations	20700			20643		

Standard errors in parentheses * significant at 5%; ** significant at 1%

Tabela 6: Regressions Results for Inefficiency Wage

	Black Female Worker Inefficiency	White Female Worker Inefficiency
Educ	-0.202 (0.025)**	-0.001 (0.023)
Child<10	0.434 (0.135)**	0.382 (0.203)
Urban	-1.004 (0.165)**	-2.001 (0.382)**
Maid Formal	-22.476 (4.511)**	-12.193 (3.579)**
Maid Informal	0.299 (0.127)*	2.336 (0.441)**
Formal Labor	-17.183 (2.726)**	-11.429 (2.700)**
Age	-0.057 (0.030)	0.018 (0.042)
Age sq	0.001 (0.001)**	0.001 (0.001)
Union	0.802 (0.167)**	-0.725 (0.301)*
Children	0.057 (0.046)	0.132 (0.080)
Ind. Sector	2.429 (0.257)**	3.117 (0.578)**
Constant	-0.972 (0.840)	-5.771 (1.802)**
Observations	20700	20643

Standard errors in parentheses * significant at 5%; ** significant at 1%

Tabela 7: Oaxaca Decomposition by Increasing Efficient Wage

Efficient Wage Quantile	Pure Discrimination (%)
0.01	0.196
0.05	0.158
0.10	0.143
0.20	0.137
0.30	0.143
0.40	0.153
0.50	0.155
0.60	0.153
0.70	0.151
0.80	0.150
0.90	0.148

black women. Education contributes to a reduction in underpayment (in other words, it improves efficiency), but again its impact is stronger for a white worker than for a black worker. Legal employment ties contribute strongly to improve efficiency and to reduce underpayment for both types of laborer, black and white. The occupation of maid without legal ties, contributes strongly to the underpayment of women; however, its impact is stronger in case of black women.

Finally, the paper attempts to contribute to the methodology, to the extent in which it takes into account the selection bias due to the fact that many women do not participate in the labor market if the reservation wage is above the market wage. We used a two-stage Heckman correction in estimation of the Mincer Stochastic Frontier. Oaxaca type analysis was conducted using the Mincer Stochastic Frontier estimated parameters such that we could derive a measure of discrimination, excluding inefficient wages from the analysis.

The end result of our analysis is that reduction of underpayment can be achieved by traditional policies and enforcement of labor laws. However, it might happen that part of the inefficiency could be due to the fact that only part of the workers might be hired under the present labor laws and those excluded from the legal labor markets have to survive by accepting underpaid jobs, such as maids, without formal ties. But our results indicate that even if traditional social economic policies are able to reduce underpayment, the improvement would be of a higher magnitude for white female workers than for black female workers. More importantly, even if wage underpayment is eliminated, there still remains a wage differential that we have called pure discrimination, which can not be eliminated by traditional policies; it is very likely that some form of affirmative action policies, with mandatory enforcement, will be required to eliminate this type of pure discrimination in female labor markets in Brazil.

Referências Bibliográficas

Aigner, D., C., L. & P., S. (1977), 'Formulation and estimation of stochastic frontier reduction models', *Journal of Econometrics* **6**, 21–37.

Arcand, J. & d'Hombres, B. (2004), 'Racial discrimination in the Brazilian labour market: Wage, employment and segregation effects.', *Journal of International Development* **16**, 1053–1066.

Barros, R., R., H. & R., M. A. (2000), 'Estabilidade inaceitável: desigualdade e pobreza no Brasil', *Desigualdade e pobreza no Brasil*. IPEA, Rio de Janeiro.

Blinder, A. S. (1973), 'Wage discrimination: Reduced form and structural estimates', *Journal of Human Resources* **8**, 436–455.

Carvalho, Néri & Nascimento. (2006), 'Diferenciais de salários por raça e gênero no Brasil: Aplicação dos procedimentos de Oaxaca e Heckman', *Pesquisas Amostrais Complexas*. Rio de Janeiro, Brazil: Instituto Brasileiro de Geografia e Estatística. Mimeo.

Dawson, P., T., H. & D., W. (2001), 'German wage underpayment: An investigation into labor market inefficiency and discrimination', *Vierteljahrshefte zur Wirtschaftsforschung* **70**(1), 107–114.

Ferreira, F. H. G. (2000), Os determinantes da desigualdade de renda no brasil: Luta de classes ou heterogeneidade educacional, Technical report, IPEA, Rio de Janeiro.

Freyre, G. (1933), *Casa-Grande & Senzala: Formação da Família Brasileira sob o Regime de Economia Patriarcal*, Maia & Schmidt.

Greene, W. (2002), *Econometric Analysis*, Prentice Hall.

Heckman, J. (n.d.), 'Sample selection bias as a specification error', *Econometrica* 47(1), 153–161.

Henriques, R. (2001), 'Desigualdade racial no brasil: Evolução das condições de vida na década de 90', IPEA. Available in www.ipea.gov.br/pub/td/td_2001/td0807.pdf. Cited 25 July 2007.

Hofler, R. A. & Murphy, K. J. (1992), 'Underpaid and overworked: Measuring the effect of imperfect information on wages', *Economic Inquiry* 30(3), 511–529.

Hofler, R. A. & Murphy, K. J. (1994), 'Estimating reservation wages of employed workers using a stochastic frontier', *Southern Economic Journal* 60(4), 961–976.

Jondrow, J., Lovell, C. A. K., Materov, I. S. & Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* 19, 233–238.

Kassouf, A. L. (1998), 'Wage gender discrimination and segmentation in the brazilian labor market', *Revista de Economia Aplicada* 2(2), 243–269.

Kassouf, A. L. & Silvia, N. D. V. (2000), 'Mercados de trabalho formal e informal: Uma análise da discriminação e da segmentação', *Nova Economia* 10(1), 41–77.

Killingsworth, M. & Heckman, J. (1986), Female labor supply: A survey, in 'Handbook of Labor Economics', Ashenfelter, O. and Layard.

Kumbhakar, S. C. & Lovell, C. A. K. (2003), *Stochastic Frontier Analysis*, University Press.

Lang, G. (2005), 'The difference between wages and wage potentials: Earnings disadvantages of immigrants in germany', *Journal of Economic Inequality* 3(1), 21–42.

Matos, R. & Machado, A. F. (2006), 'Diferencial de rendimentos por cor e sexo no brasil (1987-2001)', *Econômica* 8(1), 5–27.

McClure, K. G., Girma, P. B. & Hofler, R. A. (1998), 'International labor underpayment: a stochastic frontier comparison of canada and the united states', *Canadian Journal of Regional Science* 21(1), 87–91.

Meeusen, W. & Van Den Broeck, J. (1977), 'Efficiency estimation from cobb-douglas production functions with composed error term', *International Economic Review* 18, 435–444.

Mincer, J. (1970), 'The distribution of labor incomes: A survey with special reference to the human capital approach', *Journal of Economic Literature* .

Mincer, J. (1974), *Schooling, Experience and Earnings*, edição edn, Columbia University Press.

Murphy, K. & Topel, R. (2002), 'Estimation and inference in two step econometric models', *Journal of Business and Economics Statistics* **20**(1), 88–97.

Oaxaca, R. (1973), 'Male female wage differentials in urban labor markets', *International Economic Review* **14**(3), 693–709.

Prieto, C. G., Román, A. M. & Domínguez, C. P. (2003), 'Actual and potential returns to schooling in Spain. estudios sobre la economía española, eee 157', Available at SSRN: <http://ssrn.com/abstract=381060>.

Robinson, M. D. & Wunnava, P. V. (1989), 'Measuring direct discrimination in labor markets using a frontier approach: Evidence from cps female earnings data', *Southern Economic Journal* **56**(1), 212–218.

Telles, E. (2004), *Race in Another America: The Significance of Skin Color in Brazil*, Princeton University Press.

Zuchi, J. (2006), 'Desigualdade de rendimentos entre brancos e negros nos setores público e privado', UNESP. Dissertation.

Zuchi, J. & Hoffman, R. (2004), 'Diferenças de renda associadas à cor', *Pesquisa & Debate* **15**(1), 107–129.

ESTABILIDADE DOS PROFESSORES E QUALIDADE DO ENSINO DE ESCOLAS PÚBLICAS

LEANDRO OLIVEIRA COSTA *
RONALDO DE ALBUQUERQUE E ARRAES †
DANIEL BARBOZA GUIMARÃES ‡

Resumo

Propõe-se analisar o efeito da estabilidade dos professores sobre a qualidade do ensino fundamental nas escolas públicas brasileiras, considerando que a decisão de tornar-se docente também envolve fatores não pecuniários relacionados aos benefícios de ingressar no setor público. Com base nos dados mais recentes do exame Prova Brasil (2011) para estudantes do 5º ano do ensino fundamental, aplicou-se a metodologia de Pareamento no Escore de Propensão para múltiplos tratamentos com o objetivo de contornar o possível viés de seleção das escolas decorrente da correlação entre o benefício da estabilidade dos professores com fatores relacionados aos salários e aos critérios de admissão dos alunos. Objetivando prover robustez nos resultados, conduziu-se uma Análise de Sensibilidade proposta por Ichino et al. (2006), que simula a possibilidade da presença de uma variável não observada, a qual invalidaria a suposição de independência condicional necessária para identificar a relação de causalidade. Em comparação com os efeitos ambíguos dispostos na literatura acerca dos salários dos professores sobre o desempenho estudantil, demonstra-se, para todas as regiões do país, que quanto maior a falta de professores com estabilidade menor o desempenho dos estudantes, resultado este que se traduz no aprendizado dos alunos.

Palavras-chave: Qualidade da Educação Pública; Estabilidade dos Professores; Pareamento no Escore de Propensão; Análise de Sensibilidade.

Abstract

This paper is aimed at analyzing the effect of teachers' stability on education quality in Brazilian public elementary schools, considering that the decision to become a teacher also involves non-pecuniary factors. The latest database of the Prova Brazil Examination (2011) was applied to the methodology of propensity score matching in order to overcome the possible selection bias of schools due to the correlation between the teachers'

* Curso de Pós-Graduação em Economia da Universidade Federal do Ceará — CAEN/UFC. E-mail: leandroitap@gmail.com

† Curso de Pós-Graduação em Economia da Universidade Federal do Ceará — CAEN/UFC. E-mail: barbozadan@hotmail.com

‡ Departamento de Administração da Universidade Federal do Ceará. E-mail: barbozadan@hotmail.com

stability benefit factors associated with salaries and the students' admission rules. In addition, sensitivity analysis was performed according to Ichino et al. (2006) to verify the robustness of results over the possibility of the presence of an unobserved variable to invalidate the conditional independence assumption needed to identify the causal relationship. The results indicate, for all regions in the country, that the larger the lack of teachers holding tenure positions the lower the students' performance, vis-à-vis the ambiguous effect found in the research of wages on students' performance. This suggests that schools keeping greater proportion of teacher with tenure tend to provide a better students' scholar achievement.

Keywords: Quality of Public Education; Teachers' Stability; Propensity Score Matching; Sensitivity Analysis.

JEL classification: I21, J38, C14

DOI: <http://dx.doi.org/10.1590/1413-8050/ea63523>

1 Introdução

Embora a condição de estabilidade no emprego seja adquirida apenas no setor público mediante realização de concurso,¹ ela gera várias consequências nas relações entre a escola pública e seus docentes, podendo influenciar indiretamente a qualidade do ensino público brasileiro. De uma forma geral, a estabilidade gera um benefício não pecuniário que é um fator motivador ao ingresso no setor público. A falta de estabilidade, por um lado, pode contribuir para elevar a competitividade entre os professores dessas escolas, bem como, diminuir o custo de oportunidade de um profissional vir a ser um docente. Mas, por outro lado, conforme apontam Bresser-Pereira (1996) e Pires & Macedo (2006), restringir a possibilidade de demissão torna o setor público ineficiente na contratação de profissionais devidamente qualificados e, principalmente, motivados. Nesse sentido, a aquisição do direito de estabilidade pode influenciar negativamente na motivação devido ao excesso de segurança no emprego, que é acompanhado de muitas dificuldades relacionadas ao ensino de estudantes em condições socioeconômicas desfavoráveis.

A rigidez na regulamentação trabalhista do setor público, em especial o educacional, dificulta determinar se os salários ou outros benefícios não pecuniários são responsáveis pela atração de bons professores que melhorem o desempenho dos alunos, pois, conforme argumentam Hanushek & Rivkin (2007), a avaliação qualitativa de algum fator no setor da educação é uma tarefa extremamente complexa em virtude do domínio da prestação pública do ensino. As restrições sobre as operações de mercado e a importância dos fatores não pecuniários na decisão de oferta de professores implicam que as escolas não são, necessariamente, operadas de forma eficiente e não tomam as decisões de contratação dos professores baseadas no desempenho esperado.²

Não há consenso entre os estudos da literatura sobre a relação entre a formação e os salários dos professores e o desempenho dos alunos. Ao fazer um survey sobre os trabalhos até então publicados, Hanushek (1986, 1997) mostra que poucas são as pesquisas que apresentam um efeito positivo dos salários sobre o desempenho dos discentes. Por outro lado, Loeb & Page (2000) afirmam que esses estudos têm falhado por não considerar os atributos não pecuniários e outras oportunidades de salário que afetam o custo de oportunidade no processo de escolha de se tornar professor. Ademais, os autores apontam que os pais normalmente escolhem as escolas com base na qualidade percebida, e não nos salários dos professores, pois aqueles que demandam uma educação de qualidade, provavelmente, gastam mais com educação. Quanto à formação de professores, os resultados são controversos ou inconclusivos. Se por um lado Wobmann (2003), em análise envolvendo 39 países, identifica uma relação positiva definida entre o nível de escolaridade dos professores e o desempenho de alunos do ensino fundamental, por outro lado, Rice (2003) atesta ser inconclusiva tal relação nos Estados Unidos. No Brasil, Menezes Filho (2007) e Louzano (2010) demonstram não haver uma correlação significativa entre a formação de professores com nível superior e o desempenho dos alunos.

¹A Constituição Federal (Art. 41) estabelece garantia de estabilidade para os servidores públicos, após três anos de efetivo exercício, nomeados em virtude de concurso público.

²Estudos relacionados à realidade americana, como Hanushek & Rivkin (2007) e Hanushek et al. (2004), citam outros fatores não pecuniários que influenciam a decisão dos indivíduos em se tornarem professores. Por exemplo, a possibilidade de permanecer na mesma cidade, mesmo sabendo que mudar de cidade pode elevar seu salário.

Nos últimos anos, alguns autores têm se dedicado a pesquisar as características mais relevantes para explicar o desempenho dos alunos nas escolas brasileiras. Vale salientar, todavia, que poucos são os trabalhos que abordam a relação entre o salário do professor e o desempenho dos alunos, além de nenhum deles tratar a questão da estabilidade do docente e sua influência na melhoria do ensino, conseqüentemente, no desempenho dos alunos.

No Brasil, acredita-se que os pais de baixa renda não levam em consideração fatores não pecuniários ou os salários relacionados à carreira dos professores quando decidem matricular seus filhos em escolas públicas. Além do mais, Curi & Menezes Filho (2010) mostram que, em virtude da baixa qualidade do ensino público e de sua limitada oferta de vagas, famílias pobres são induzidas a matricular seus filhos em escolas privadas, mesmo comprometendo boa parte da sua renda.

Desse modo, supõe-se que os pais que demandam escolas públicas não têm uma concepção bem definida acerca dos salários dos professores e da estrutura da escola e não detêm poder de escolha sobre a turma ou dos professores dos seus filhos. Ademais, no ato da matrícula, as famílias de baixa renda levam em consideração, principalmente, as restrições de disponibilidade de vagas e da proximidade da escola às suas residências. No entanto, caso seja possível escolher entre duas escolas próximas, espera-se que a decisão recaia sobre aquela dotada de melhores estruturas e resultados³.

Considerando que a seleção dos professores leva em conta, além dos salários, o benefício não pecuniário da estabilidade, e que há restrições econômicas dos pais na escolha das escolas, este artigo tem como objetivo analisar o efeito da estabilidade dos professores das escolas públicas do ensino fundamental brasileiro sobre o desempenho dos estudantes nos exames de proficiência. Para isso, será utilizada a metodologia de Pareamento no Escore de Propensão, uma técnica semiparamétrica que, sob fortes suposições de identificação do efeito causal, reproduz um experimento aleatório a partir da atribuição de pesos aos estudantes não tratados na tentativa de se determinar um confiável contrafactual. Não obstante, questiona-se a suposição que as características observáveis, que influenciam tanto o desempenho dos alunos como o processo de seleção das escolas, possibilitem determinar esses pesos. Assim, foi realizada a Análise de Sensibilidade proposta por Ichino et al (2006) para verificar a robustez dos resultados diante da presença de uma variável não observada que invalidasse a suposição de independência condicional necessária para identificar a relação de causalidade.

Devido à impossibilidade de identificar todos os respectivos professores de cada um dos alunos no 1º, 2º, 3º e 4º ano e em alguns casos ter sido informado mais de um professor por turma no 5º ano, foi utilizado como variável do tratamento uma variável categórica que ordenada os quartis de professores com estabilidade. Essa variável permitiu testar o grau de associação entre a proporção de professores estáveis nas escolas públicas e o desempenho dos estudantes do 5º ano do Ensino Fundamental nos exames de Língua Portuguesa e Matemática dos estudantes no Prova Brasil de 2011. Com isso, pretende-se investigar se a estabilidade é um importante indicativo para a seleção de bons professores, conseqüentemente, para a melhoria da qualidade do ensino das

³Pode-se supor, ainda, que as famílias mais próximas das escolas têm uma maior oportunidade para conseguir matricular seus filhos, e que pais mais motivados com a educação criam mecanismos que possibilitam a escolha das melhores escolas.

escolas públicas.

O artigo está organizado de modo a compor mais quatro seções. Na próxima, faz-se uma revisão da literatura acerca dos trabalhos mais relevantes que tratam desse assunto. Na terceira seção apresenta-se a base de dados acompanhada de uma análise descritiva dos mesmos. Em seguida, discute-se a metodologia de Pareamento no Escore de Propensão e de Análise de Sensibilidade que darão suporte aos resultados das estimativas, os quais constam na quinta seção. As considerações finais concluem o artigo.

2 Revisão da Literatura

O estudo de Hanushek & Rivkin (2007) defende que a qualidade do ensino é um componente central para aumentar a qualidade da escola, e que, a qualidade dos professores e as características do mercado de trabalho em que estão inseridos, apresentam papel preponderante na determinação do desempenho do aluno. De acordo com os autores, determinadas características observáveis, muitas vezes relacionadas com as decisões de contratação e salário, explicam a variação na qualidade do ensino. Eles fizeram uma comparação entre as políticas que buscam elevar a qualidade exigindo mais qualificações necessárias para entrar no ensino e políticas que visam aumentar a qualidade, relaxando as restrições de entrada e de introdução de incentivos de desempenho para professores e administradores. Concluem que a maioria das pessoas percebe os efeitos potenciais das políticas por meio do impacto aferido pela qualidade dos professores, embora pouco se conheça sobre a potencial interação entre as estruturas institucionais e a qualidade dos professores.

Rivkin et al. (2005) investigam como o salário e a condições de trabalho afetam a qualidade do ensino em escolas públicas dos Estados Unidos, buscando contornar o problema de endogeneidade devido a variáveis omitidas, erro de medida das variáveis e viés de seleção dos estudantes e escolas. O estudo conclui que aumentos salariais para professores seriam, em geral, ineficazes, mas, melhorias na qualidade do ensino poderiam ser alcançadas por meio da implementação de algumas medidas, tais como: 1) reduzir as barreiras para um profissional se tornar um professor, como por exemplo, a exigência de certificação; 2) haver uma relação direta entre salários e progressão na carreira; 3) possibilidade de compensar os professores com maior capacidade em aumentar o desempenho dos estudantes.

Loeb & Page (2000) focam seu estudo na relação entre salário dos professores e aprendizagem dos alunos a partir de dados longitudinais de estados americanos, na tentativa de contestar as conclusões de Hanushek (1986, 1997), que utilizou dados em *cross-section*, e de outros estudos que utilizaram dados longitudinais (Altonji 1988, Loeb & Page 2000, Grogger 1996, Betts 1995), que o salário do professor não afeta o desempenho dos alunos, o que contradiz os estudos que afirmam que o efeito escola é significativo (Altonji 1988, Hanushek et al. 1998). As autoras afirmam que esses estudos não levaram em consideração os atributos morais do trabalho e as oportunidades salariais alternativas, que afetam o custo de oportunidade de escolher lecionar. Elas desenvolveram um modelo com variáveis defasadas para inferir sobre a relação entre os salários dos professores e os resultados dos alunos, medido por níveis acadêmicos dos estudantes, que incorpora atributos não pecuniários associados com o ensino, juntamente com as oportunidades alternativas de mercado

de trabalho. Elas rejeitam os resultados dos autores supracitados e concluem que um aumento de 50% no salário dos professores reduziria a taxa de abandono escolar em 15% e aumentaria a taxa de entrada na faculdade em cerca de 8%.

Hoxby & Leigh (2004) apontam para a possibilidade de que a manutenção de salários relativamente mais baixos no setor educacional dos Estados Unidos gera um problema de seleção adversa, levando os estudantes graduados menos habilitados a optar pela carreira docente. Suas estimativas com variáveis instrumentais sob um painel de dados comprovam que, nos últimos anos, a contratação de professores mais habilitados caiu de 5% para 1%, e os menos habilitados aumentou de 16% para 36%. O crescimento nessa categoria deve-se à contratação de mulheres graduadas com competência acadêmica mais baixa. A maior parte dessa mudança foi atribuída à compressão salarial, que é a redução no tempo das diferenças salariais entre pessoas no mesmo emprego ou entre pessoas em diferentes postos de trabalho em uma hierarquia organizacional, seguida das políticas de paridade salarial e redução no salário médio dos professores.

Trabalhos correlatos no Brasil com abordagem econômica foram introduzidos, inicialmente, por Barros et al. (2001), onde a partir dos dados da Pesquisa Nacional por Amostra de Domicílio (PNAD) de 1996 e da Pesquisa sobre Padrão de Vida (PPV) de 1996/97 aplicados em regressões lineares, verificam que a escolaridade dos professores da segunda etapa do ensino fundamental contribui mais que a dos professores do ensino médio sobre a formação escolar dos brasileiros (número de séries completadas). No entanto, os autores mostram que a escolaridade dos pais, e em particular da mãe, é a característica mais importante para determinar o desempenho educacional dos filhos.

Menezes Filho (2007) utiliza dados do Sistema de Avaliação do Ensino Básico (SAEB) de 2003 para examinar os fatores que afetam o desempenho dos alunos do ensino fundamental e médio nos testes de proficiência em Matemática. A partir do exercício de decomposição da variância, identifica que o desempenho escolar é prioritariamente explicado pelos fatores: educação da mãe, cor, atraso escolar e reprovação prévia, número de livros disponíveis aos alunos, disponibilidade de computador próprio e atividades no mercado de trabalho. Conclui também que os salários dos professores explicam o desempenho dos alunos apenas na rede privada. Esse resultado talvez decorra da concorrência entre essas escolas por profissionais mais qualificados, o que elevaria a variância dos salários, contrariamente do que ocorre com a maior rigidez dos salários dos professores de escolas públicas.

No tocante à remuneração dos professores das redes pública e privada brasileira, cabem destacar outros trabalhos. Inicialmente, Barbosa Filho et al. (2009) utilizam os censos demográficos de 1980, 1991 e 2000 para analisar a evolução dos diferenciais salariais nos níveis de ensino para o Brasil e seus estados, com destaque para São Paulo. Concluem, por um lado, que os diferenciais de salários médios público/privado reduziram sobremaneira ao longo dos vinte anos, à exceção de alguns estados do Norte e Nordeste, além de as variâncias dos diferenciais terem diminuído drasticamente em todos os níveis de ensino. Mas, por outro lado, ao se computar o valor presente do contrato de trabalho (vpct) para 2000 dos rendimentos futuros incorporados os benefícios previdenciários com diferentes taxas de desconto, a situação se reverte. Em nível de Brasil, para uma taxa de desconto de 3%, o vpct para professores da rede pública excede em no mínimo 18% (nível superior) e no máximo 47%

(pré-escola) os da rede privada, e se professoras, os excedentes são de 25% (nível superior) e 40% (ensino fundamental).

Becker & Kassouf (2012) também propõem investigar os diferenciais salariais de professores do ensino fundamental das redes pública e privada por meio de um modelo econométrico composto por duas equações, onde aplicam a decomposição de Oaxaca, as quais objetivam explicar o salário e o *vpct*, este último seguindo Barbosa Filho et al. (2009), tendo como fatores explicativos as características individuais, características do trabalho, produtividades individuais e aspectos regionais. Com base em dados das PNADs de 2004 a 2008, as autoras concluem que os salários dos professores da rede pública são, em média, maiores do que os da rede privada, ditado, principalmente, pela maior experiência. Os benefícios de aposentadorias são ratificados para indicar o maior *vpct* dos professores da rede pública, e a decomposição revela um diferencial de salário em favor dos homens em ambas as redes de ensino, cuja explicação se baseia em fatores não observáveis, dentre os quais, a discriminação.

Com base nos dados do SAEB de 2001 dos alunos de 4ª série do Ensino Fundamental do estado de São Paulo, Felício & Fernandes (2005) decompõem a desigualdade das notas, onde parte é explicada pela escola e outra pelo *background* familiar. Seus resultados indicam que o efeito escola explica até 28,4% da desigualdade total de notas da disciplina de Língua Portuguesa e entre 8,7% e 34,44% das notas de Matemática. Em seguida, incluem variáveis dummies indicativas de escola para estimar, por efeitos fixos, a importância de cada escola no aprendizado dos alunos, e estabelecer um ranking da qualidade das escolas. Em ambas as disciplinas, as simulações revelam um impacto de um desvio-padrão sobre o desempenho dos alunos, o que, comparativamente, significa elevar o desempenho médio verificado na 4ª série com o da 7ª série. Comprovam também que, entre as escolas públicas, seria possível obter um avanço significativo com um ganho de três anos de estudo para Matemática e 2,4 para Língua Portuguesa. Nesta última, isso equivaleria elevar o desempenho médio dos alunos da rede pública ao da rede particular.

Biondi & Felício (2007), utilizando dados do SAEB e Censo Escolar de 2003, estimam o efeito de variáveis escolares no desempenho dos alunos por meio de um painel balanceado de 260 escolas. Constataram que, a experiência média e a baixa rotatividade dos professores durante o ano letivo têm efeito positivo sobre o desempenho dos estudantes da 4ª série do ensino fundamental. Contrariamente, Franco (2009), a partir de um painel não balanceado de dados do SAEB de 1997, 1999, 2001, 2003 e 2005, não observou nenhum efeito das características observáveis dos professores e poucas variáveis das escolas se mostraram relevantes. Cabe ressaltar que, mesmo não tendo sido aplicada a abordagem de avaliação de programas e focado na relação causal entre as características dos professores que afetam o desempenho dos alunos, Biondi & Felício (2007) e Franco (2009) buscaram solucionar o problema de endogeneidade decorrente do problema de variável omitida e viés de seleção por meio de estimações com efeito fixo.

O estudo de Scorzafave & Ferreira (2011) toma o estado de São Paulo como referência para avaliar os determinantes da desigualdade no desempenho escolar entre os alunos da 4ª série (5º ano) do ensino fundamental da rede pública. Com base nos dados do exame Prova Brasil (2005), o índice de desigualdade Theil-L é aplicado para decompor as contribuições de variáveis distribuídas no *background* dos alunos (21 variáveis) e características das esco-

las (20 variáveis, nenhuma que caracterize o *status* empregatício do professor) nas diferenças de notas obtidas nas provas de português e matemática. Dentre suas conclusões, destaca-se o fato de as contribuições das variáveis serem mais acentuadas para a desigualdade dos escores em português do que em matemática. Enquanto as variáveis de *background* do aluno respondem por 22,94% e 26,14% da desigualdade dos exames de matemática e português, respectivamente, as contribuições das variáveis das características escolares reduzem tal desigualdade para 13,55% e 12,13% nos respectivos exames.

Conforme relatado anteriormente nenhum desses artigos abordou o efeito da estabilidade do professor sobre o desempenho dos estudantes. Portanto, este artigo busca contribuir com a literatura ao analisar o efeito da estabilidade dos docentes do ensino fundamental da rede pública no Brasil sobre o desempenho dos estudantes nos exames de proficiência. Ademais, acredita-se que a relação entre fatores não pecuniários no mecanismo de seleção dos professores das escolas públicas pode ter mascarado o efeito dos salários dos professores sobre o desempenho dos estudantes.

3 Descrição dos Dados

Para a consecução do objetivo proposto serão utilizados os microdados do Prova Brasil de 2011, os quais são coletados a partir de questionários socioeconômicos e exames de proficiência nas disciplinas de Língua Portuguesa e Matemática dos alunos da 5ª série do Ensino Fundamental da rede pública urbana, pertencentes aos estabelecimentos com um mínimo de 20 alunos por turma. A opção por essa base de dados decorre de ser a mais recente, além de conceber um vasto detalhamento da avaliação da qualidade da educação no Brasil, uma vez que, ela fornece informações relativas à 2.278.771 alunos matriculados em 40.823 escolas participantes distribuídas nos municípios dos 27 estados da Federação.⁴

Com relação à distribuição espacial, o Prova Brasil insere 10,3% de estudantes do Norte, 26,9% do Nordeste, 43,2% do Sudeste, 12,3% do Sul e 7,2% do Centro-Oeste. Em relação à dependência administrativa das escolas públicas, 75,05% são municipais, 24,91% estaduais e 0,1% federais.⁵ Os critérios mais aplicados nas escolas públicas para a cessão de vagas são a localização da moradia e a ordem de chegada para a matrícula. Ambos restringem a escolha dos pais pela escola pública de seus filhos, haja vista que umas detêm melhor qualidade de ensino do que outras, por conseguinte, o excedente da demanda sobre a oferta de vagas é inevitável.

Devido à impossibilidade de identificar todos os respectivos professores de cada um dos alunos no 1º, 2º, 3º e 4º ano e em alguns casos ter sido informado mais de um professor por turma no 5º ano, a informação acerca do regime de trabalho dos professores que mensura a proporção de professores com estabilidade foi extraída da seguinte questão contida no questionário direcionado aos diretores: “qual é o percentual de professores com vínculo estável nesta

⁴As médias de desempenho nessas avaliações também constituem o cálculo do Índice de Desenvolvimento da Educação Básica (IDEB). Os dados dessas avaliações são comparáveis ao longo do tempo, ou seja, pode-se acompanhar a evolução dos desempenhos das escolas, das redes e do sistema como um todo.

⁵Optou-se por excluir os estudantes das escolas públicas federais pelo fato de essas instituições possuírem critérios de ingresso que são correlacionados com o nível socioeconômico dos seus alunos.

escola?”. Como esse vínculo é adquirido apenas pelos professores associados ao sistema estatutário de servidores públicos, assume-se que aqueles sem vínculo estável são substitutos ou temporários. Justifica-se esta opção tanto pelo fato de os pais não considerarem o status do professor ao matricular o seu filho em determinada escola pública, como também não deterem informação completa sobre as condições de contrato dos professores, sejam trabalhistas ou salariais. Consequentemente, como se objetiva investigar o efeito da estabilidade a partir de uma variável que representa a condição da escola no conjunto de professores com estabilidade, é analisado o efeito da condição de estabilidade dos professores de forma agregada por escola, o que impõe uma suposta homogeneidade do efeito da estabilidade entre alunos da mesma instituição. A Tabela 1 apresenta a distribuição regional de escolas públicas por faixa de professores com estabilidade empregatícia distribuída por quartis.

Os dados dessa tabela revelam que, em todas as regiões, a participação de escolas com professores estáveis cresce a taxas crescentes ao se evoluir do primeiro para o quarto quartil. O percentual acumulado até o terceiro quartil não ultrapassa 43,4% de escolas (região Sudeste), demonstrando, assim, a elevada contribuição marginal do quartil superior sobre o percentual de escolas. Especificamente, observa-se que 57,8% das escolas públicas brasileiras empregam mais de 75% de seus professores no regime estatutário. Comparando as regiões, percebe-se que no Sudeste 7,1% das escolas públicas contratam até 25% de seus professores em regime de estabilidade, enquanto para as demais esse percentual é de no mínimo 8,2%. Do lado oposto, quando se observa o percentual de escolas com mais de 75% dos professores com estabilidade, verifica-se que o Sudeste supera as demais regiões com 58,9% de escolas. Isso pode refletir a forma de contratação dos professores nessa região, que reduz a competitividade entre professores, o que pode ocasionar uma flexibilidade na escolha de docentes menos qualificados ou motivados em detrimento da elevação dos benefícios não pecuniários proporcionados pela estabilidade.

Tabela 1: Distribuição Regional de Escolas Públicas por Intervalo de Professores com Estabilidade (%)

Faixas de Professores com Estabilidade	Norte	Nordeste	Sudeste	Sul	Centro-Oeste	Brasil
(0%, 25%]	12,5	11,5	7,1	8,2	9,6	9,2
(25%, 50%]	13,0	12,1	10,7	14,2	11,9	11,8
(50%, 75%]	16,4	15,1	20,6	18,1	21,9	18,5
(75%, 100%]	55,3	58,5	58,9	57,0	53,7	57,8
Dados Omitidos	2,8	2,7	2,6	2,5	2,9	2,7

Fonte: Elaboração dos autores a partir dos dados do Prova Brasil, 2011.

Dados descritivos sobre as características próprias dos alunos e socioeconômicas de seus pais, por regiões, estão dispostos na Tabela 2.

Percebe-se que há uma associação direta ao se relacionar as proporções de pais sem o ensino fundamental (EF) e as condições de alunos que já foram reprovados, notadamente com percentuais mais elevados nas regiões mais pobres, Norte e Nordeste. No extremo oposto, as regiões Norte e Nordeste detêm o menor percentual de pai e mãe com nível superior.⁶ Isso deve ser um fator

⁶Considerando as regiões, as correlações entre as proporções de mãe e pai sem EF e alunos

Tabela 2: Distribuição Percentual das Características dos Estudantes e seus Pais, e dos Domicílios

	Norte	Nordeste	Sudeste	Sul	Centro-Oeste	Brasil
Escolas Estaduais	31,22	12,18	31,33	26,61	41,24	26,30699
Alunos mulheres	46,08	44,99	47,40	46,94	47,29	46,55136
Alunos pretos	9,06	10,89	9,30	6,12	8,44	9,2463
Pai e mãe morando no domicílio	91,85	91,22	94,76	94,95	93,29	93,42918
Mãe sem EF	13,55	18,74	9,89	11,56	9,96	12,85645
Mãe com Graduação	11,61	7,96	12,60	11,92	14,07	11,2752
Pai sem EF	14,53	18,86	9,17	10,08	10,42	12,52662
Pai com Graduação	7,74	5,41	9,28	9,58	9,23	8,1138
Domicílios com um carro	17,88	16,98	39,16	47,34	39,11	32,00567
Domicílios com dois carros	3,69	2,89	10,15	12,66	8,66	7,73241
Domicílios com três carros	1,54	1,32	3,56	4,37	2,8	2,79282
Domicílios com Internet	23,76	19,1	49,82	46,45	40,85	37,81007
Domicílios com Domésticas	11,44	10,24	9,38	10,02	10,31	9,42402
Alunos com distorção idade-série	29,89	28,91	11,94	16,24	17,86	19,3094
Alunos reprovados	39,07	39,59	21,40	26,87	26,7	29,16976
Alunos com pré-escola	65,15	71,57	77,44	66,29	68,68	72,58606

Fonte: Elaboração dos autores a partir dos dados do Prova Brasil, 2011.

relevante de herança educacional com impacto no desempenho escolar dos filhos e gerações futuras.

No entanto, mesmo apresentando os maiores e menores percentuais de pai e mãe sem ensino fundamental (EF) e com nível superior (graduação), respectivamente, a região Nordeste detém o segundo maior indicador quando se considera os alunos com pré-escola ou educação infantil (71,6%), superada apenas da Região Sudeste, cujo percentual é de 77,4%. Por sua vez, a região Norte apresenta o pior indicador nesse quesito, ou seja, 65,2% dos alunos dessa região concluíram a pré-escola.

Em relação à condição econômica das famílias, medida pela posse de bens, as regiões Sudeste, Sul e Centro-Oeste são dotadas com valores mais de duas vezes superiores aos das regiões mais pobres, Norte e Nordeste, resultado este em concordância com o esperado, haja vista a superioridade de desenvolvimento das primeiras regiões.

Extraem-se dessas informações que as escolas localizadas nas regiões Norte e Nordeste possuem alunos com os piores indicadores de condições socioeconômicas domiciliares e desempenho escolar (distorção idade-série, reprovação), baixo nível de escolaridade dos pais, em contraposição àqueles das demais regiões brasileiras onde os pais apresentam melhores condições socioeconômicas e *status* educacional, além da elevada participação de alunos com frequência à pré-escola, e ter apresentado o mais baixo índice de reprovação.

As distribuições de rendimentos por faixa salarial dos professores estatutários e não estatutários entre as regiões estão sintetizadas na Tabela 3, as quais apresentam heterogeneidade quanto à estabilidade e regionalização. Inicialmente, chama a atenção que, à exceção do Nordeste, as proporções de professores que recebem remunerações até quatro salários mínimos são maiores na categoria não estatutária, a qual dá a liderança para a região Sul com 91%. Dentre os professores nesse nível de remuneração, a região Centro-Oeste de-

com reprovações são de 85% e 92%, respectivamente, enquanto para pais graduados as correlações se reverterem para -70% e -84%.

têm as menores proporções sejam estatutários (56%) ou não (71%), enquanto as maiores proporções são lideradas por professores não estatutários da região Sul (91%) e estatutários da região Nordeste (84%). Tais resultados se revertem na complementaridade das proporções para os maiores salários.

Em termos de variabilidade, extrai-se dessa tabela que há, para as proporções até quatro salários mínimos entre todas as regiões, uma amplitude de variação de 28 p.p. para professores estatutários e de 20 p.p. para os não estatutários, enquanto que para as proporções de maiores salários tais variações são iguais a 20 p.p. Isso reflete certa estabilidade de variações nas classes da distribuição de rendimentos entre os professores do país.

A relação entre escolaridade e situação trabalhista dos professores nas escolas públicas brasileiras é descrita na Tabela 4, a qual mostra que as maiores taxas daqueles sem nível superior, sejam estatutários ou não, encontram-se nas regiões menos desenvolvidas (Norte e Nordeste). Realce-se que as taxas desses professores não estatutários do Nordeste é quase dez vezes à correspondente do Centro-Oeste e mais de duas vezes às das regiões Sul e Sudeste. Em todas as regiões, as proporções de professores sem nível superior e não estatutários excedem em pelo menos 50% às de estatutários.

Dentre os professores com nível superior, há baixa variabilidade na proporção entre as duas categorias de emprego em todas as áreas para cada região, à exceção de Pedagogia no Nordeste e Sul com um diferencial de 9% a favor dos estatutários e não estatutários, respectivamente.

Distinguindo a variabilidade entre todas as regiões para os professores com nível superior, a área de Pedagogia lidera a amplitude de variação tanto para os estatutários (15 p.p.) quanto para os não estatutários (23 p.p.). A área de Matemática é a que apresenta menor variação nas proporções, com um diferencial máximo de 2 p.p. na região Centro-Oeste.

Condicionar pelas características individuais dos alunos (escolaridade de seus pais e informações sobre sua vida estudantil) e pelos critérios que as escolas utilizam para selecioná-los (ordem de chegada e a proximidade entre a residência e a escola), pode eliminar o viés associado às características observadas que influenciam as decisões dos pais sobre qual escola pública matricular seu filho.⁷

As características de estabilidade dos professores que podem mascarar o impacto na aprendizagem dos alunos são outras condicionalidades necessárias para isolar o efeito estudado a partir do pareamento de alunos com professores dotados com características de salário e escolaridade comuns.

As prováveis variáveis não observadas que afetam a escolha da escola pelos pais são: a distância da escola, a disponibilidade de vagas e a ideia que o pai tem sobre a qualidade da estrutura, gestão e do ensino. Caso essas variáveis sejam realmente significativas, pode-se esperar uma forte correlação dessas com a decisão de o pai matricular o filho em uma escola com determinado percentual de professores com estabilidade e que apresente alunos com melhores desempenhos. Isso tornaria inviável a comparação dos resultados dos alunos em escolas com diferentes percentuais de estabilidade e, conseqüentemente, questionáveis as estimativas da metodologia de Pareamento no Escore de Pro-

⁷Restringe-se a análise do efeito da estabilidade dos professores sobre os resultados dos alunos de escolas públicas, pois se supõe que a decisão dos pais que podem matricular seus filhos em escolas privadas é diferente.

Tabela 3: Rendimentos dos Professores por Vínculo Empregatício nas Regiões Nordeste e Sudeste

Classe de Salário (R\$)	Norte		Nordeste		Sudeste		Sul		Centro-Oeste	
	Estatutário	Não Estatutário	Estatutário	Não Estatutário	Estatutário	Não Estatutário	Estatutário	Não Estatutário	Estatutário	Não Estatutário
(0; 545]	2%	3%	2%	9%	2%	2%	1%	3%	1%	2%
(545; 817]	3%	7%	6%	18%	6%	5%	2%	10%	6%	3%
(817; 1.090]	7%	9%	11%	11%	11%	8%	5%	11%	12%	7%
(1.090; 1.362]	12%	12%	14%	7%	14%	15%	11%	7%	13%	16%
1.362; 1.635]	7%	6%	9%	3%	9%	11%	11%	9%	10%	5%
(1.635; 1.907]	5%	3%	7%	3%	7%	9%	9%	10%	7%	2%
(1.907; 2.180]	7%	3%	4%	1%	4%	5%	9%	5%	6%	2%
(2.180; 2.725]	12%	4%	5%	1%	5%	7%	13%	4%	10%	10%
(2.725; 3.815]	7%	2%	4%	1%	4%	3%	9%	2%	10%	5%
(3.815; ,5.450]	2%	0%	1%	0%	1%	1%	3%	0%	2%	1%
Mais de 5.450	1%	1%	1%	1%	1%	1%	1%	1%	1%	2%

Fonte: Elaboração dos autores a partir dos dados do Prova Brasil, 2011.

Tabela 4: Distribuição Percentual da Escolaridade dos Professores por Vínculo Empregatício

Nível de Escolaridade	Norte		Nordeste		Sudeste		Sul		Centro-Oeste	
	Estatutário	Não Estatutário	Estatutário	Não Estatutário	Estatutário	Não Estatutário	Estatutário	Não Estatutário	Estatutário	Não Estatutário
Ensino Médio ou Menos	14%	20%	14%	26%	8%	11%	8%	11%	3%	7%
Ensino Superior — Pedagogia	47%	38%	51%	38%	51%	52%	54%	54%	65%	57%
Ensino Superior — Matemática	4%	3%	5%	4%	4%	3%	4%	3%	6%	3%
Ensino Superior — Letras	6%	5%	10%	8%	10%	7%	11%	4%	10%	7%
Ensino Superior — Outros	27%	17%	20%	15%	26%	21%	22%	13%	17%	11%

Fonte: Elaboração dos autores a partir dos dados do Prova Brasil, 2011.

pensão. A Tabela 5 descreve as variáveis utilizadas na determinação do Escore de Propensão, desconsiderando a influência das variáveis não observadas.

A variável referente à dependência administrativa da escola ser ou não estadual pode desempenhar papel importante na análise, pois caracteriza de forma satisfatória a gestão das escolas. Cabe realçar a inserção de *dummies* para os estados, tendo em vista as diferenças nas características socioeconômicas de cada região e os programas educacionais particulares de cada estado.⁸

Em relação às variáveis referentes aos professores, os salários foram sintetizados em duas variáveis *dummies* para capturar os efeitos dos extremos da distribuição salarial, seguindo a escala disponível no questionário; uma para salários menores, se o professor ganha até R\$817,50 mensais (1,5 salário mínimo de 2011), e outra para salários maiores, se o professor ganha acima de R\$2180,00 reais (4 salários mínimos de 2011). Além destas, foi criada outra variável qualitativa para captar o efeito da escolaridade do professor definida por meio da detenção de nível superior.

4 Metodologia

Inferência sobre a relação de causalidade em estudos observacionais requer suposições sobre como o efeito da intervenção pode diferir do efeito observado em um experimento aleatório. Nesse sentido, a análise do efeito da estabilidade dos professores sobre o desempenho dos estudantes das escolas públicas brasileiras insere-se nesse contexto, pois não é possível controlar o mecanismo de distribuição dos estudantes nas diversas escolas de um município, cada uma das quais com sua dotação de professores com estabilidade, possibilitando, portanto, comparar o desempenho de alunos para verificar o efeito do tratamento. Ademais, o processo de determinação de quais escolas os alunos estão matriculados pode ser regido por fatores não observados nos questionários socioeconômicos dos alunos. Nesse contexto, será realizada uma análise de sensibilidade sobre as hipóteses da metodologia de Pareamento no Escore de Propensão (PEP) para verificar se, utilizando esta metodologia,⁹ é possível obter resultados confiáveis.

A metodologia PEP, introduzida por Rosenbaum & Rubin (1983b) e aplicada por diversos autores, tais como, Heckman et al. (1998), Imbens (2004), Caliendo & Kopeinig (2008),¹⁰ busca determinar um contrafactual a partir da determinação de indivíduos dentro do grupo de controle que sejam semelhantes aos tratados pelas variáveis observáveis inseridas no cálculo do escore de propensão. Essa metodologia é uma técnica semi-paramétrica de verificação dos efeitos de um programa em um quase-experimento, baseada em algoritmos

⁸Os estados do Maranhão e Espírito Santo foram excluídos para evitar multicolineariedade nas estimações do escore de propensão.

⁹Rosenbaum (2002), Morgan & Winship (2007), Imbens & Wooldridge (2008) apresentam a importância da análise de sensibilidade em contextos em que se suspeita da veracidade das suposições de identificação das metodologias para avaliar os programas. Ademais, Todd (2008), Blundell & Costa-Dias (2009), Heckman & Todd (2009), Khandker et al. (2010) são bons apanhados sobre as mais recentes metodologias de avaliação de programas.

¹⁰Um dos mais interessantes debates sobre a metodologia PEP ocorreu quando Dehejia & Wabba (1999), utilizando os dados do estudo de Lalonde (1986), mostrou que estimadores do PEP estão mais próximos aos estimadores de estudos experimentais do que os produzidos pelos métodos tradicionais de avaliação. Entretanto, Smith & Todd (2005) mostraram que os estimadores são melhores em apenas uma subamostra muito específicas dos dados Lalonde, gerando dúvidas sobre a generalização dos resultados de Dehejia & Wabba (1999).

Tabela 5: Estatísticas Descritivas das Variáveis Utilizadas na Obtenção do Escore de Propensão

Variáveis	Nordeste				Sudeste				Norte				Sul				Centro-Oeste							
	Média ¹	DP	Min	Max	Média ¹	DP	Min	Max	Média ¹	DP	Min	Max	Média ¹	DP	Min	Max	Média ¹	DP	Min	Max				
Aluna Mulher	0,45	0,50	0	1	0,47	0,50	0	1	0,46	0,50	0	1	0,47	0,50	0	1	0,47	0,50	0	1				
Aluno Preto	0,11	0,31	0	1	0,09	0,29	0	1	0,09	0,29	0	1	0,06	0,24	0	1	0,08	0,28	0	1				
Mãe sem EF	0,19	0,39	0	1	0,10	0,30	0	1	0,14	0,34	0	1	0,11	0,31	0	1	0,10	0,29	0	1				
Pai sem EF	0,19	0,39	0	1	0,09	0,29	0	1	0,15	0,35	0	1	0,10	0,30	0	1	0,10	0,30	0	1				
Distorção Idade-Série	0,29	0,45	0	1	0,12	0,32	0	1	0,30	0,46	0	1	0,16	0,37	0	1	0,17	0,38	0	1				
Aluno Reprovado	0,40	0,49	0	1	0,21	0,41	0	1	0,39	0,49	0	1	0,27	0,44	0	1	0,26	0,44	0	1				
Aluno com Pré-escola	0,72	0,45	0	1	0,77	0,42	0	1	0,65	0,48	0	1	0,67	0,47	0	1	0,69	0,46	0	1				
Professor sem Graduado	0,19	0,39	0	1	0,09	0,29	0	1	0,17	0,37	0	1	0,09	0,28	0	1	0,04	0,20	0	1				
Professor Salário Baixo	0,16	0,36	0	1	0,04	0,20	0	1	0,07	0,25	0	1	0,09	0,29	0	1	0,03	0,18	0	1				
Professor Salário Alto	0,07	0,25	0	1	0,20	0,40	0	1	0,15	0,36	0	1	0,17	0,37	0	1	0,20	0,40	0	1				
Escola Estadual	0,12	0,33	0	1	0,31	0,46	0	1	0,31	0,46	0	1	0,27	0,44	0	1	0,42	0,49	0	1				
Admissão Local de Moradia	0,14	0,35	0	1	0,49	0,50	0	1	0,12	0,32	0	1	0,48	0,50	0	1	0,19	0,39	0	1				
Admissão Ordem de Chegada	0,22	0,41	0	1	0,09	0,29	0	1	0,30	0,46	0	1	0,12	0,33	0	1	0,29	0,45	0	1				
PI	0,06	0,23	0	1	MG	0,27	0,44	0	1	RO	0,04	0,20	0	1	SC	0,42	0,49	0	1	MT	0,18	0,39	0	1
CE	0,17	0,38	0	1	RJ	0,16	0,36	0	1	AM	0,24	0,43	0	1	RS	0,29	0,45	0	1	GO	0,21	0,41	0	1
RN	0,06	0,24	0	1	SP	0,53	0,50	0	1	RR	0,03	0,17	0	1						DF	0,41	0,49	0	1
PB	0,06	0,23	0	1						PA	0,46	0,50	0	1										
PE	0,14	0,35	0	1						AP	0,05	0,22	0	1										
AL	0,07	0,26	0	1						TO	0,08	0,28	0	1										
SE	0,04	0,19	0	1																				
BA	0,25	0,43	0	1																				

Fonte: Elaboração dos autores a partir dos dados do Prova Brasil, 2011.

¹ Refere-se à proporção do atributo "1" (coluna Max) das variáveis dummies.

Nº de obs: Norte: 235.452; Nordeste: 612.270; Sudeste: 985.467; Sul: 230.062; Centro-Oeste: 144.746.

de paramento de indivíduos pertencentes a grupos distintos com o objetivo de se julgar os efeitos de uma determinada intervenção. As suas premissas de identificação são a Suposição de Independência Condicional (SIC) ou seleção nos observáveis ou unconfundness e a sobreposição ou balanceamento nas variáveis pré-tratamento.

Em decorrência de a variável tratamento ser aqui definida para quatro categorias naturalmente ordenadas do percentual de professores com estabilidade nas escolas, utiliza-se a metodologia de pareamento no escore de propensão com múltiplos tratamentos, a qual é abordada por Rosenbaum & Rubin (1983a), Imbens (2000), Lechner (2001). Sua diferença para a análise do efeito de um único tratamento é que as suposições de identificação são modificadas para incorporar a estrutura não binária do tratamento. Algumas aplicações dessa metodologia foram implementadas por Lechner (1999), Imbens (2000), Blundell et al. (2005). Conforme frisam Becker & Caliendo (2007), verificar a sensibilidade dos resultados estimados em relação aos desvios da hipótese de identificação torna-se uma abordagem cada vez mais importante na literatura de avaliação de programas. Para os estimadores do efeito da estabilidade dos professores sobre o desempenho dos estudantes, essa análise aborda a possibilidade de existir variáveis não observadas que afetam simultaneamente a atribuição para o tratamento e a variável de resultado, criando um viés igualmente não observado que possa tornar os estimadores de pareamento não robustos. Essa abordagem foi introduzida por Rosenbaum & Rubin (1983a) e desenvolvida por Rosenbaum (2002), Imbens (2003), Altonji et al. (2005), Ichino et al. (2006),¹¹ cujas aplicações foram adotadas, dentre as mais citadas, por Aakvik (2001), Diprete & Gangl (2004), Ichino et al. (2006), Caliendo et al. (2007).

4.1 Pareamento no Escore de Propensão — Múltiplos Tratamentos

Seguindo a notação de Imbens (2000) sobre a metodologia PEP, supõe-se que a escola “*i*” possa obter múltiplos valores ordenados de tratamento quanto aos intervalos de percentual de professores com estabilidade, $T_i(t)$, onde o tratamento de interesse $t = 0, 1, 2, \dots, k$. No caso específico de distribuírem-se os professores com estabilidade em quartis, faz-se uma inversão da categoria dos intervalos para que o tratamento $T_i(0)$ seja a categoria que engloba o maior número de casos, qual seja, estudantes em escolas com mais de 75% de professores com estabilidade. Essa inversão também facilita a análise do efeito da falta de professores com estabilidade nas escolas, pois os tratamentos ordenados das outras categorias torna o maior tratamento igual à menor categoria de professores com estabilidade, ou seja $T_i(3)$, correspondente à categoria de estudantes em escolas com até 25% de professores com estabilidade. Os resultados potenciais $Y(t)$ são as performances dos alunos, dado que suas escolas podem apresentar diferentes níveis de tratamento, ou seja, estarem em escolas em diferentes intervalos de professores com estabilidade.

Com o objetivo de estabelecer uma suposição que possibilite a identificação no caso de múltiplos tratamentos, Rosenbaum & Rubin (1983a) definem que a Fraca Suposição de Independência Condicional (FSIC) requer somente a independência de cada tratamento com cada um dos resultados potenciais

¹¹Diprete & Gangl (2004), Becker & Caliendo (2007), Nannicini (2007) desenvolveram rotinas estatísticas que realizam a análise de sensibilidade dos estimadores de pareamento no escore de propensão para o software Stata.

$Y(t)$, em vez da suposição de qualquer tratamento ser independente do conjunto total de resultados potenciais. Ademais, essa suposição somente requer a independência “local” de cada resultado potencial $Y(t)$ e o tratamento de interesse, que é independente do indicador binário do nível de tratamento $D(t)$, em vez do nível de tratamento T . Por exemplo, se a escola se encontra com o tratamento $T_i(3)$, na pior classe em comparação com as outras, $T_i(2)$, $T_i(1)$ e $T_i(0)$, desconsidera-se a classificação e leva-se em conta somente a independência do $T_i(3)$ como se fosse um tratamento binário. Ou seja, dadas as variáveis X de pré-tratamento, a FSIC corresponde a:

$$D(t) \perp Y(t)|X \tag{1}$$

Segundo Imbens (2000), a FSIC está relacionada à interpretação do problema da inferência causal a partir de *missing data*. Para unidades com $D_i(t) = 0$, a variável $Y_i(t)$ sempre é desconhecida. Como o objetivo é estimar a média de $Y_i(t)$ da população, $E[Y(t)]$, deve-se assegurar a representatividade da média dos $Y_i(t)$ na subamostra dos participantes que recebem o tratamento $D_i(t) = 1$. Mesmo não existindo nessa interpretação uma função direta para o nível de tratamento realmente atribuído às unidades com $D_i(t) = 0$, o que importa é que os indivíduos desse grupo de comparação não recebem tratamento “t”. No entanto, em decorrência dos outros resultados potenciais $Y_i(s)$, para $s \neq t$, nunca será observado os mesmos resultados que as unidades com $D_i(t) = 1$. Ou seja, os resultados não podem desempenhar nenhum papel em qualquer processo de ajustamento sem suposições sobre as características dos que não participaram do programa. A definição de FSIC corresponde à falta de relevância desses outros resultados potenciais, dado a condicionalidade das variáveis explicativas X , o que possibilita estimar o resultado médio condicionado pelo status do tratamento, ou seja:

$$E[T(t)|X = x] = E[Y(t)|D(t) = 1, X = x] = E[Y(t)|T = t, X = x] = E[Y|T = t, X = x] \tag{2}$$

Os resultados médios podem ser estimados a partir dessas médias condicionais com:

$$E[Y(t)] = E[E[Y(t)|X]] \tag{3}$$

Com o objetivo de resolver o problema dimensional advindo de X , define-se o Escore de Propensão Generalizado (EPG) como a probabilidade condicional de receber um particular nível de tratamento, dadas as variáveis observáveis, que se iguala à esperança condicional do tratamento binário (Imbens 2000). O EPG pode ser expresso como:

$$e(t, x) = Pr(T = t|X = x) = E[D(t)|X = x] \tag{4}$$

Dado o EPG e a FSIC para múltiplos tratamentos, essa suposição pode ser, assim, representada:

$$D(t) \perp Y(t)|e(t, x) \tag{5}$$

No caso de haver múltiplos tratamentos, a suposição de sobreposição ou balanceamento nas variáveis de pré-tratamento dado o EPG, segue sua representação:

$$D(t) \perp X|e(t,x) \quad (6)$$

Nota-se que o argumento condicionante muda com o nível do tratamento. Para garantir a independência condicional, no caso de múltiplos tratamentos e variáveis de pré-tratamento X , é necessário condicionar por todo o conjunto de $K + 1$ escores, $\{e(t,x)\} \in T$.

Com o objetivo de melhorar o pareamento, escolheu-se de forma *ad hoc* quatro tratamentos, que é a variável categórica, com correspondência inversa aos quartis de professores com estabilidade. Assim, por exemplo, $T(4) = 4$ se refere ao tratamento das escolas com perfis que detêm tais professores no intervalo [0%, 25%]. O último quartil, (75%, 100%), que corresponde ao tratamento $T(1) = 1$, será considerado como intervalo de grupo de controle de todas as outras possibilidades de tratamento. Ao analisar os níveis de tratamento individualmente, $D(t) = 1$ significa que determinada escola tem um percentual no intervalo do tratamento t e $D(t) = 0$ significa que a escola está em outro intervalo.

A implementação da metodologia de pareamento no escore de propensão para múltiplos tratamentos é similar ao caso quando este é binário. Inicialmente, estima-se o escore de propensão, $e(t,x)$, a partir de um modelo que incorpore os múltiplos tratamentos ordenados, que optou-se pela hipótese da distribuição logística. Em seguida, estima-se o efeito médio no nível de tratamento t , $\beta(t) = E\{\beta[t, e(t,x)]\}$, que é estimado com a análise de pareamento de *Kernel* para cada nível de tratamento relativo aos demais, caracterizando-se, desta forma, uma escolha de tratamento binário.

4.2 Análise de Sensibilidade

A análise de Sensibilidade tem o objetivo de simular a inserção de uma suposta variável não observada para testar a possibilidade de invalidar a hipótese que fundamenta o Pareamento no Escore de Propensão. Para isso, considera-se um possível problema denominado por Rosenbaum (2002) como viés oculto. Esse se verifica quando variáveis não observadas afetam tanto a participação no programa quanto os resultados potenciais, contrariando a suposição de independência condicional. Embora essa suposição não seja testável, a sensibilidade dos resultados estimados pelo PEP pode ser verificada no que diz respeito a desvios dessa hipótese de identificação. Ou seja, mesmo que o grau de seleção ou do viés oculto não possa ser estimado, o grau de sensibilidade dos resultados do PEP à suposição de *unconfoundedness* pode ser testado. No caso específico da estimativa referente à estabilidade dos professores, o efeito de variáveis não observadas com a motivação dos professores e a habilidade individual dos estudantes não pode ser mensurado, mas se poderia verificar a robustez das estimativas se fosse possível invalidar a hipótese de ignorabilidade inserindo uma suposta variável não observada.¹²

Em outras palavras, a Análise de Sensibilidade dos estimadores de pareamento testa a robustez dos resultados que são criados a partir da suposição de identificação de independência condicional. A ideia básica é que se supõe que o tratamento, a faixa de professores com estabilidade, não é independente

¹²*Unconfoundedness* e ignorabilidade do tratamento são utilizadas como sinônimos da Suposição de Independência Condicional (SIC) que requer a independência do tratamento em relação aos resultados potenciais.

dos exames dos alunos dado apenas ao conjunto de características observáveis X , mas seria independente se condicionado conjuntamente com outras características não observáveis, U . Mas como não conhecemos a variável U , faz-se simulações considerando diferentes conjuntos de suposições sobre a distribuição de U e sua relação com D e os resultados $Y(0)$ e $Y(1)$, é possível verificar a sensibilidade dos resultados com respeito às variações nesses pressupostos (Caliendo & Kopeinig 2008). Isto é, para verificar o grau de sensibilidade dos estimadores do PEP, supomos algumas variáveis que poderiam representar variáveis não observadas para investigar como os resultados encontrado pelo PEP variaram.

Seguindo a abordagem desenvolvida por Ichino et al. (2006) e Nannicini (2007), a suposição de independência condicional, dadas as características observáveis (X) e não observáveis (U), pode ser representada por:

$$Y(t) \perp D(t) | (X, U) \quad (7)$$

Uma vez que U não é observada, o resultado dos controles não pode ser usado para estimar o resultado contrafactual do tratado, ou seja:

$$E[D(0)|D(t) = 1, X] \neq E[D(0)|D(t)=0, X] \quad (8)$$

Por outro lado, caso U e X fossem conhecidos, então se estabeleceria uma condição suficiente para se estimar consistentemente o efeito médio do tratamento sobre os tratados (ETT), uma vez que:

$$E[D(0)|D(t) = 1, X, U] = E[D(0)|D(t) = 0, X, U] \quad (9)$$

Ichino et al. (2006) sugere a hipótese simplificadora de U possuir a distribuição de uma variável binária a fim de permitir uma análise de sensibilidade, além de ser independente e identicamente distribuída nas células representadas pelo produto cartesiano do tratamento e os valores dos resultados. Por simplicidade, considera-se o caso de resultados potenciais binários, $Y(0), Y(1) \in \{0, 1\}$. Para isso, cria-se uma nova variável binária que assume o valor unitário quando o desempenho dos estudantes for maior que a média, e valor nulo caso contrário. Considera-se também, a partir da equação, $Y(t) = D(t).Y(1) + [1 - D(t)].Y(0)$, os resultados observados para uma determinada unidade, que é igual a um dos dois resultados possíveis, dependendo da atribuição do tratamento. A distribuição do fator binário U é plenamente caracterizada pela escolha de quatro parâmetros:

$$P_{i,j} = Pr(U = 1 | D(t) = i, Y(t) = j) = Pr(U = 1 | D(t) = i, Y(t) = j, W) \quad (10)$$

onde, $i, j \in \{0, 1\}$ e W é o conjunto das variáveis X e U .

Da equação (10) extraem-se as probabilidades da ocorrência $U = 1$ em cada um dos quatro grupos definidos pelo status de tratamento e os valores dos resultados. Por exemplo, para um aluno com desempenho maior que a média, $Y(1)$, e na escola com o determinado intervalo de professores com estabilidade, $D(1)$, o grupo da probabilidade de que $U = 1$ é p_{11} . Consequentemente, dado arbitrariamente os valores dos parâmetros p_{ij} , uma valor de U é atribuído a cada indivíduo. Outra hipótese simplificadora é a independência condicional de U em relação a W . Ichino et al. (2006) mostram que essas hipóteses simplificadoras não afetam os resultados da análise de sensibilidade.

A variável não observada U é então tratada como qualquer outra observada, condição esta que permite ser incluída no conjunto de variáveis utilizadas para estimar o escore de propensão e calcular o efeito médio do tratamento sobre os tratados (ETT), de acordo com o estimador de pareamento escolhido. Usando um determinado conjunto de valores dos parâmetros de sensibilidade, a estimativa do pareamento é repetida várias vezes e uma simulada estimação do ETT é obtida como uma média dos ETTs sobre a distribuição de U . Assim, para qualquer determinada configuração dos parâmetros p_{ij} , a análise de sensibilidade recupera uma estimativa pontual do ETT, que é robusto com o fracasso da SIC implícito naquela particular configuração.

No caso de resultados contínuos, a análise de sensibilidade acima pode ser realizada a partir de parâmetros de simulação p_{ij} com base em D e uma transformação binária de Y . Uma vez que os parâmetros p_{ij} são definidos dessa maneira, pode-se aplicar a análise de sensibilidade como descrito acima.

Nannicini (2007) reconhece que uma variável não observada U poderia causar erros estatísticos na inferência, no sentido que sua existência pode dar origem a uma significativa e positiva estimativa do ETT mesmo na ausência de um verdadeiro efeito causal, se for observado que:

$$Pr(Y(0) = 1|D(t), X, U) \neq Pr(Y(0) = 1|D(t), X) \quad (11)$$

$$Pr(D = 1|X, U) \neq Pr(D = 1|X) \quad (12)$$

Ichino et al. (2006) demonstram que, assumindo $p_{01} > p_{00}$, é possível simular uma variável considerada danosa à inferência que permita a simulação de um efeito positivo significativo de uma variável não observada sobre os resultados dos não tratados. Ademais, assumindo $p_{1j} > p_{0j}, \forall j$, é possível simular uma variável não observada que tem um efeito positivo sobre a seleção do tratamento.

Com o objetivo de caracterizar a simulada variável não observada (U), é possível medir o quanto cada configuração escolhida de p_{ij} é responsável pelo efeito de U sobre $Y(0)$, denominado efeito sobre os resultados, e de U sobre T , efeito sobre a seleção, condicionado em W . Em síntese, estima-se um modelo de escolha discreta binária em cada iteração, $Pr(Y = 1|D = 0, U, X)$ e a razão de probabilidade média de U , considerando todos os efeitos, é obtida:

$$\Gamma = \begin{cases} Pr(Y = 1|D = 0, U = 1, X) \\ Pr(Y = 0|D = 0, U = 1, X) \\ Pr(Y = 1|D = 1, U = 1, X) \\ Pr(Y = 0|D = 1, U = 1, X) \end{cases} \quad (13)$$

Similarmente, o modelo binário, $Pr(D = 1|U, X)$, Logit é estimado em cada interação, e a razão de probabilidade média é considerada como o efeito sobre a seleção da simulada variável não observada, ou seja:

$$\Lambda = \begin{cases} Pr(D = 1|U = 1, X) \\ Pr(D = 0|U = 1, X) \\ Pr(D = 1|U = 0, X) \\ Pr(D = 0|U = 0, X) \end{cases} \quad (14)$$

Nannicini (2007) reconhece que pela simulação de U sob as suposições que $p_{10} > p_{00}$ e $p_{1j} > p_{0j}$, ambos os efeitos, sobre os resultados e sobre a seleção, devem ser positivos. Isto é, na análise de sensibilidade variável U representa um risco de simulação quando, $\Gamma > 1$ e $\Lambda > 1$.

5 Resultados

O escore de propensão foi estimado como a probabilidade prevista de um aluno estar matriculado em uma escola com um percentual de professores com estabilidade. As probabilidades foram estimadas a partir de uma regressão de escolha discreta da variável categórica sobre as características observadas que afetam a escolha da escola, como: a escolaridade da mãe e do pai, a dependência administrativa, as características da escolaridade e salários dos professores, os critérios de local de moradia e ordem de chegada da matrícula dos estudantes na escola, o número de alunos por turma e *dummies* dos estados.¹³ Baseado no pareamento de casos similares do grupo de tratamento e controle, o método de pareamento no escore de propensão, descrito na seção 4, busca eliminar o viés de seleção originado pelas variáveis observáveis e, conseqüentemente, parear a partir do escore de propensão os indivíduos no grupo de tratamento e controle que simulem a escolha dos indivíduos de um experimento aleatório. Os resultados do modelo de regressão de escolha discreta para as regiões brasileiras¹⁴ estão dispostos nas Tabelas 6 e 7. Foi estimado um modelo Logit Ordenado que é o mais indicado para a variável dependente considerada no estudo (Becker & Caliendo 2007).

Os sinais dos coeficientes estimados ditam o efeito positivo ou negativo dos regressores sobre o aumento ou redução da probabilidade de uma variável dependente discreta assumir o valor unitário. Como frisado, os intervalos foram invertidos para que o tratamento $T_i(0)$ seja a categoria que engloba o maior número de casos (mais de 75% dos professores com estabilidade), e, conseqüentemente, facilitar a análise do efeito da falta de professores com estabilidade nas escolas, que se encontra principalmente no tratamento $T_i(3)$ (de 0% a 25% de professores com estabilidade). Nesse sentido, se os coeficientes são positivos, um aumento nas variáveis explicativas diminui (aumenta) a probabilidade de um aluno estar na categoria das escolas com mais de 75% (até 25%) dos professores com estabilidade.¹⁵

Observa-se, portanto, que tanto no quesito significância de uma variável como em relação ao sinal do coeficiente, os resultados apresentados nas Tabelas 6 e 7 seguem o mesmo padrão, ou seja, não há diferença nos impactos que as variáveis exercem sobre a probabilidade das escolas terem menos professores com estabilidade quando é considerado apenas os resultados do exame de

¹³Caliendo & Kopeinig (2008) sintetizam a ideia que há sobre a escolha das variáveis explicativas que devem ser inseridas no cálculo do escore de propensão. Em suma, a quantidade não deve ser muito excessiva ou limitada, mas, conter variáveis relacionadas tanto com o tratamento como com o resultado e, principalmente, serem inseridas visando melhorar o balanceamento das variáveis entre o grupo controle e tratamento.

¹⁴Seguindo sugestão de Long & Freeze (2001), os interceptos foram suprimidos das equações para permitir que os mesmos variem dentre as variáveis *dummies* da equação.

¹⁵No modelo logístico ordenado os resultados ordenados são modelados para aumentar sequencialmente como uma variável latente, y^* , que é dividida progressivamente por maiores cortes. Ou seja, y^* é uma medida não observada do percentual de professores com estabilidade. Veja mais informações sobre esse modelo em Wooldridge (2010).

Tabela 6: Português

Variáveis	Nordeste		Sudeste		Norte		Sul		Centro-Oeste					
	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão				
Aluna Mulher	-0,007	0,006	-0,003	0,004	-0,004	0,009	0,008	0,009	0,005	0,011				
Aluno Preto	0,004	0,010	0,008	0,008	0,068*	0,016	0,064*	0,018	0,044*	0,019				
Mãe sem EF	0,041*	0,009	0,052*	0,008	0,060*	0,015	0,038*	0,015	0,094*	0,019				
Pai sem EF	0,044*	0,008	0,046*	0,008	0,040*	0,015	-0,007	0,016	0,085*	0,018				
Distorção Idade-Série	-0,008	0,008	0,069*	0,008	0,004	0,012	0,033*	0,014	0,054*	0,017				
Aluno Reprovado	-0,003	0,007	0,037*	0,006	0,006	0,011	0,061*	0,012	0,095*	0,014				
Aluno com Pré-escola	-0,016*	0,007	-0,058*	0,005	-0,007	0,010	-0,077*	0,009	-0,121*	0,012				
Professor Sem Graduação	0,238*	0,008	0,195*	0,008	0,251*	0,015	0,130*	0,015	0,306*	0,026				
Professor Salário Baixo	0,514*	0,008	0,444*	0,011	0,329*	0,020	0,257*	0,015	0,564*	0,029				
Professor Salário Alto	-0,420*	0,013	-0,310*	0,006	-0,275*	0,013	-0,171*	0,012	0,002	0,013				
Escola Estadual	1,060*	0,009	0,571*	0,005	0,903*	0,011	1,536*	0,011	1,127*	0,013				
Admissão Local de Moradia	0,039*	0,009	0,063*	0,005	0,107*	0,015	0,079*	0,009	-0,144*	0,014				
Admissão Ordem de Chegada	-0,095*	0,007	0,071*	0,008	0,001	0,010	-0,110*	0,015	-0,170*	0,013				
Dummies dos Estados														
PI	0,116*	0,015	MG	-1,504*	0,011	RO	1,788*	0,026	SC	-0,208*	0,012	MT	0,824*	0,019
CE	-0,271*	0,011	RJ	-1,444*	0,012	AM	0,804*	0,019	RS	1,120*	0,011	GO	1,398*	0,018
RN	-0,607*	0,015	SP	-1,197*	0,010	RR	0,175*	0,033				DF	0,367*	0,019
PB	0,304*	0,014				PA	0,825*	0,020						
PE	0,023*	0,011				AP	-0,736*	0,032						
AL	-0,009	0,014				TO	0,051*	0,024						
SE	-0,441*	0,019												
BA	-0,318*	0,011												
Número de Observações	433556		846883		178865		225138		140675					
Log likelihood	32161,19*		37582,99*		16846,37*		62289,26*		17618,13*					
Pseudo R ²	0,0350		0,0210		0,0414		0,1241		0,0552					

Tabela 6: Matemática (continuação)

Variáveis	Nordeste		Sudeste		Norte		Sul		Centro-Oeste					
	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão	Coefficientes	Erro-Padrão				
Aluna Mulher	-0,008	0,006	-0,006	0,004	-0,004	0,009	0,007	0,009	0,003	0,011				
Aluno Preto	0,005	0,010	0,007	0,008	0,070*	0,016	0,065*	0,018	0,041*	0,019				
Mãe sem EF	0,042*	0,009	0,055*	0,008	0,058*	0,015	0,043*	0,015	0,090*	0,019				
Pai sem EF	0,044*	0,008	0,045*	0,008	0,040*	0,015	-0,002	0,015	0,082*	0,018				
Distorção Idade-Série	-0,006	0,008	0,065*	0,008	0,002	0,012	0,026	0,014	0,049*	0,017				
Aluno Reprovado	-0,005	0,007	0,037*	0,006	0,008	0,011	0,064*	0,012	0,091*	0,014				
Aluno com Pré-escola	-0,015*	0,007	-0,060*	0,005	-0,009	0,010	-0,075*	0,009	-0,124*	0,012				
Professor Sem Graduação	0,236*	0,008	0,203*	0,008	0,279*	0,015	0,127*	0,015	0,322*	0,026				
Professor Salário Baixo	0,520*	0,008	0,431*	0,011	0,332*	0,020	0,265*	0,015	0,519*	0,029				
Professor Salário Alto	-0,406*	0,013	-0,308*	0,006	-0,274*	0,013	-0,178*	0,012	0,008	0,013				
Escola Estadual	1,068*	0,009	0,570*	0,005	0,907*	0,011	1,527*	0,011	1,126*	0,013				
Admissão Local de Moradia	0,036*	0,009	0,064*	0,005	0,088*	0,015	0,073*	0,009	-0,139*	0,014				
Admissão Ordem de Chegada	-0,091*	0,007	0,062*	0,008	0,000	0,010	-0,107*	0,015	-0,174*	0,013				
Dummies dos Estados														
PI	0,106*	0,015	MG	-1,492*	0,011	AC	1,821*	0,0263	PR	1,337*	0,012	MT	0,860*	0,019
CE	-0,266*	0,011	RJ	-1,432*	0,012	AM	0,803*	0,0193	RS	0,212*	0,012	GO	1,423*	0,018
RN	-0,613*	0,015	SP	-1,192*	0,010	RR	0,142*	0,0332				DF	0,400*	0,019
PB	0,305*	0,014				PA	0,831*	0,0196						
PE	0,029*	0,011				AP	-0,744	0,0317						
AL	-0,018	0,014				TO	0,055*	0,0238						
SE	-0,437*	0,019												
BA	-0,320*	0,011												
Número de Observações	435874		850629		179753		225721		140703					
Log likelihood	32642,38*		37160,14*		17056,98*		62389,66*		17572,78*					
Pseudo R ²	0,0348		0,0206		0,0416		0,1238		0,0550					

Tabela 7: Estimativas do Modelo de Pareamento no Escore de Propensão e MQO — Português

Resultado	Amostra	Grupos			Teste <i>t</i>		MQO
		Tratamento	Controle	Diferença	E. P.	<i>t</i>	
Região Nordeste							
(0%, 25%]	Não Pareados	-1,438	-1,414	-0,024	0,009	-2,70	-0,0321*
	ETT	-1,438	-1,421	-0,017	0,009	-1,88	
(25%, 50%]	Não Pareados	-1,455	-1,412	-0,044	0,009	-5,11	-0,0221*
	ETT	-1,455	-1,420	-0,035	0,009	-4,15	
(50%, 75%]	Não Pareados	-1,438	-1,413	-0,025	0,008	-3,25	-0,0186*
	ETT	-1,438	-1,421	-0,017	0,008	-2,23	
Região Sudeste							
(0%, 25%]	Não Pareados	-1,059	-0,944	-0,114	0,011	-10,18	-0,0785*
	ETT	-1,059	-0,957	-0,102	0,011	-9,28	
(25%, 50%]	Não Pareados	-0,998	-0,947	-0,050	0,009	-5,39	-0,0463*
	ETT	-0,998	-0,961	-0,037	0,009	-3,96	
(50%, 75%]	Não Pareados	-0,965	-0,949	-0,016	0,007	-2,22	-0,0174*
	ETT	-0,965	-0,958	-0,007	0,007	-1,01	
Região Norte							
(0%, 25%]	Não Pareados	-1,346	-1,293	-0,053	0,008	-6,44	-0,0688*
	ETT	-1,346	-1,271	-0,075	0,008	-8,99	
(25%, 50%]	Não Pareados	-1,299	-1,299	0,000	0,008	0,04	-0,011*
	ETT	-1,299	-1,296	-0,003	0,008	-0,36	
(50%, 75%]	Não Pareados	-1,281	-1,303	0,022	0,007	3,08	-0,008
	ETT	-1,281	-1,304	0,023	0,007	3,18	
Região Sul							
(0%, 25%]	Não Pareados	-0,997	-0,927	-0,070	0,010	-6,67	-0,104*
	ETT	-0,997	-0,902	-0,095	0,011	-8,88	
(25%, 50%]	Não Pareados	-0,934	-0,933	-0,001	0,008	-0,11	-0,060*
	ETT	-0,934	-0,899	-0,035	0,009	-4,05	
(50%, 75%]	Não Pareados	-0,923	-0,935	0,013	0,007	1,73	-0,041*
	ETT	-0,923	-0,921	-0,002	0,007	-0,27	
Região Centro-Oeste							
(0%, 25%]	Não Pareados	-1,144	-0,986	-0,159	0,009	-16,96	-0,121*
	ETT	-1,144	-1,018	-0,126	0,009	-13,72	
(25%, 50%]	Não Pareados	-1,082	-0,989	-0,094	0,008	-11,38	-0,075*
	ETT	-1,082	-1,030	-0,052	0,008	-6,25	
(50%, 75%]	Não Pareados	-1,001	-0,999	-0,002	0,006	-0,24	-0,027*
	ETT	-1,001	-1,013	0,012	0,007	1,89	

Fonte: Elaboração própria a partir dos dados do Prova Brasil, 2011.

Nota: (*) valor- $p < 1\%$.

Tabela 7: Estimativas do Modelo de Pareamento no Escore de Propensão e MQO — Matemática (continuação)

Resultado	Amostra	Grupos			Teste <i>t</i>		MQO
		Tratamento	Controle	Diferença	E. P.	<i>t</i>	
Região Nordeste							
(0%, 25%]	Não Pareados	-1,130	-1,117	-0,014	0,009	-1,56	-0,030*
	ETT	-1,130	-1,123	-0,007	0,009	-0,78	
(25%, 50%]	Não Pareados	-1,154	-1,113	-0,041	0,008	-4,87	-0,019*
	ETT	-1,154	-1,122	-0,033	0,008	-3,88	
(50%, 75%]	Não Pareados	-1,134	-1,115	-0,019	0,008	-2,47	-0,019*
	ETT	-1,134	-1,123	-0,011	0,008	-1,43	
Região Sudeste							
(0%, 25%]	Não Pareados	-0,676	-0,551	-0,1252	0,011	-11,06	-0,080*
	ETT	-0,676	-0,588	-0,088	0,011	-7,85	
(25%, 50%]	Não Pareados	-0,627	-0,552	-0,0755	0,009	-8,02	-0,045*
	ETT	-0,627	-0,565	-0,062	0,009	-6,67	
(50%, 75%]	Não Pareados	-0,570	-0,557	-0,0131	0,007	-1,84	-0,019*
	ETT	-0,570	-0,562	-0,008	0,007	-1,13	
Região Norte							
(0%, 25%]	Não Pareados	-1,074	-1,006	-0,069	0,008	-8,52	-0,069*
	ETT	-1,075	-0,981	-0,094	0,008	-11,29	
(25%, 50%]	Não Pareados	-1,016	-1,013	-0,002	0,008	-0,27	-0,020*
	ETT	-1,016	-1,008	-0,008	0,008	-0,99	
(50%, 75%]	Não Pareados	-0,992	-1,018	0,027	0,007	3,79	-0,001
	ETT	-0,992	-1,020	0,028	0,007	4,00	
Região Sul							
(0%, 25%]	Não Pareados	-0,630	-0,524	-0,106	0,011	-9,91	-0,105*
	ETT	-0,630	-0,523	-0,107	0,011	-10,00	
(25%, 50%]	Não Pareados	-0,558	-0,528	-0,029	0,008	-3,59	-0,062*
	ETT	-0,558	-0,523	-0,034	0,009	-4,02	
(50%, 75%]	Não Pareados	-0,540	-0,531	-0,009	0,007	-1,19	-0,042*
	ETT	-0,540	-0,531	-0,009	0,008	-1,23	
Região Centro-Oeste							
(0%, 25%]	Não Pareados	-0,853	-0,651	-0,202	0,009	-21,64	-0,122*
	ETT	-0,853	-0,683	-0,169	0,009	-18,43	
(25%, 50%]	Não Pareados	-0,760	-0,657	-0,103	0,008	-12,43	-0,074*
	ETT	-0,760	-0,699	-0,060	0,008	-7,14	
(50%, 75%]	Não Pareados	-0,660	-0,671	0,010	0,006	1,63	-0,023*
	ETT	-0,660	-0,682	0,021	0,007	3,30	

Fonte: Elaboração própria a partir dos dados do Prova Brasil, 2011.

Nota: (*) valor-*p* < 1%.

português ou de matemática. Não obstante, observam-se variações nas magnitudes dos coeficientes entre as regiões brasileiras, evidenciando dessa forma a heterogeneidade presente no território brasileiro. Por exemplo, em relação às escolas estaduais da região Sul, seu efeito marginal, é cerca de 3 (razão dos coeficientes) vezes superior ao do Sudeste sobre a probabilidade de as escolas terem menos professores com estabilidade.

Neste estudo, utilizou-se o algoritmo de pareamento de *Kernel* no escore de propensão desenvolvido em Heckman et al. (1998), que faz combinações usando todos os indivíduos na amostra de comparação ponderando com pesos menores as observações mais distantes.¹⁶ A metodologia de pareamento pondera a amostra do grupo controle a fim de aumentar a semelhança com os indivíduos do grupo tratamento. Essa é verificada a partir de uma análise de balanceamento das variáveis utilizadas na estimação do escore de propensão. Uma importante ferramenta são os histogramas dos escores de propensão para os indivíduos do grupo controle e tratamento, Gráficos A e B do Apêndice. Visualmente, verifica-se que as distribuições são em torno dos mesmos valores de escore de propensão, ou seja, os indivíduos apresentam escores de propensão que possibilitam o pareamento no escore de propensão.¹⁷

Em decorrência de o pareamento no escore de propensão ter removido a maior parte do viés atribuído às variáveis observadas, pode-se usar a diferença nos resultados médios encontrados nas amostras para obter uma estimativa do Efeito Médio do Tratamento sobre os Trados (ETT).¹⁸ As estimações do ETT a partir da metodologia de Pareamento no Escore de Propensão, para os exames de Português e Matemática estão expostas nas Tabelas 8 e 9, respectivamente.

Os resultados apontam para os significativos efeitos negativos das escolas em todos os intervalos de proporção de professores com estabilidade em todas as regiões. Na região Nordeste, o efeito médio do aluno estar matriculado em uma escola com menos de 25% de professores estáveis é um desempenho negativo nos exames de português 0,017 do desvio-padrão, e na região Centro-Oeste um efeito médio negativo nos exames de matemática é 0,169 do desvio-padrão. Esses resultados são os maiores comparados com as outras regiões. Ainda, verifica-se que na maioria das regiões, em ambos os exames, que os efeitos negativos nas escolas com mais de 25% de professores com estabilidade são maiores que nas demais regiões.

Para o segundo intervalo, (25%,50%], os impactos negativos também são significantes, no entanto verifica-se que o efeito se reduz em menos da metade, como pode ser observado para as regiões Sudeste, Sul e Centro-Oeste e chegando a ser praticamente nulo na região Norte, relativo ao primeiro intervalo. Os efeitos negativos não são expressivos para o terceiro intervalo, (50%,75%], principalmente nas regiões Sudeste e Sul onde não atingem sequer o nível de 0,001 do desvio-padrão. Ou seja, supondo que a ignorabilidade forte se con-

¹⁶Segundo Caliendo & Kopeinig (2008) o algoritmo de pareamento de *Kernel* é indicado quando há muitas observações para fazer o pareamento. Essa escolha obtém menores variâncias e estimações mais precisas, mas não se exime a possibilidade de haver pareamentos ruins. Os autores não encontraram diferenças significativas na estimação com o algoritmo de raio e a de vizinhos mais próximos. Tais resultados estarão prontamente disponíveis à solicitação do leitor.

¹⁷Os autores podem disponibilizar a análise de balanceamento das variáveis, proposta por Rosenbaum (2002) e Caliendo & Kopeinig (2008), utilizadas na estimação do escore de propensão.

¹⁸Nessa análise considera-se que o principal efeito a ser investigado é o efeito da falta de professores com estabilidade sobre os alunos que são afetados diretamente por essa ingerência dos municípios e estados. Nesse sentido, o efeito a ser identificado é o Efeito Médio do Tratamento sobre os Trados em vez do Efeito Médio do Tratamento.

Tabela 8: Probabilidades p_{ij} das Variáveis Utilizadas como Modelo — Português

Região	Local de Moradia			Salário Baixo		
	(0%, 25%]	(25%, 50%]	(50%, 75%]	(0%, 25%]	(25%, 50%]	(50%, 75%]
NORDESTE						
p_{11}	0,16	0,17	0,17	0,19	0,19	0,18
p_{10}	0,13	0,15	0,15	0,23	0,22	0,17
p_{01}	0,16	0,16	0,16	0,13	0,13	0,13
p_{00}	0,13	0,13	0,13	0,14	0,14	0,15
$p_{1.}$	0,15	0,16	0,16	0,21	0,21	0,17
$p_{0.}$	0,15	0,14	0,14	0,14	0,14	0,14
SUDESTE						
p_{11}	0,46	0,55	0,52	0,08	0,05	0,04
p_{10}	0,52	0,49	0,50	0,04	0,03	0,05
p_{01}	0,50	0,49	0,50	0,04	0,04	0,04
p_{00}	0,48	0,48	0,49	0,04	0,04	0,03
$p_{1.}$	0,49	0,52	0,51	0,06	0,03	0,05
$p_{0.}$	0,49	0,49	0,49	0,04	0,04	0,04
NORTE						
p_{11}	0,13	0,11	0,15	0,08	0,07	0,05
p_{10}	0,13	0,10	0,16	0,11	0,06	0,05
p_{01}	0,12	0,12	0,11	0,06	0,06	0,07
p_{00}	0,12	0,12	0,11	0,07	0,07	0,07
$p_{1.}$	0,13	0,10	0,15	0,10	0,06	0,05
$p_{0.}$	0,12	0,12	0,11	0,06	0,07	0,07
CENTRO OESTE						
p_{11}	0,15	0,13	0,22	0,06	0,06	0,03
p_{10}	0,21	0,13	0,20	0,06	0,04	0,03
p_{01}	0,20	0,20	0,19	0,03	0,03	0,04
p_{00}	0,17	0,18	0,17	0,04	0,04	0,04
$p_{1.}$	0,18	0,13	0,21	0,06	0,05	0,03
$p_{0.}$	0,19	0,19	0,18	0,03	0,03	0,04
SUL						
p_{11}	0,51	0,48	0,52	0,12	0,11	0,05
p_{10}	0,48	0,53	0,47	0,10	0,11	0,11
p_{01}	0,48	0,48	0,48	0,08	0,08	0,09
p_{00}	0,45	0,44	0,45	0,11	0,11	0,11
$p_{1.}$	0,49	0,51	0,49	0,11	0,11	0,08
$p_{0.}$	0,47	0,46	0,46	0,10	0,09	0,10

Fonte: Elaboração própria a partir dos dados do Prova Brasil, 2011.

Tabela 8: Probabilidades p_{ij} das Variáveis Utilizadas como Modelo — Matemática (continuação)

Região	Local de Moradia			Salário Baixo		
	(0%, 25%]	(25%, 50%]	(50%, 75%]	(0%, 25%]	(25%, 50%]	(50%, 75%]
NORDESTE						
p_{11}	0,13	0,16	0,17	0,20	0,19	0,17
p_{10}	0,16	0,15	0,16	0,22	0,22	0,18
p_{01}	0,15	0,15	0,15	0,13	0,13	0,13
p_{00}	0,14	0,14	0,14	0,15	0,14	0,15
$p_{1.}$	0,14	0,16	0,16	0,21	0,21	0,17
$p_{0.}$	0,15	0,15	0,14	0,14	0,14	0,14
SUDESTE						
p_{11}	0,49	0,55	0,52	0,07	0,04	0,03
p_{10}	0,49	0,48	0,48	0,04	0,03	0,05
p_{01}	0,50	0,49	0,49	0,03	0,04	0,03
p_{00}	0,48	0,48	0,48	0,04	0,04	0,04
$p_{1.}$	0,49	0,51	0,50	0,06	0,04	0,04
$p_{0.}$	0,49	0,49	0,48	0,04	0,04	0,04
NORTE						
p_{11}	0,11	0,11	0,14	0,09	0,07	0,05
p_{10}	0,13	0,10	0,16	0,10	0,06	0,05
p_{01}	0,12	0,12	0,11	0,06	0,06	0,07
p_{00}	0,12	0,12	0,11	0,06	0,07	0,07
$p_{1.}$	0,13	0,10	0,15	0,09	0,06	0,05
$p_{0.}$	0,12	0,12	0,11	0,06	0,07	0,07
CENTRO OESTE						
p_{11}	0,19	0,13	0,23	0,06	0,05	0,02
p_{10}	0,17	0,13	0,19	0,06	0,05	0,04
p_{01}	0,20	0,21	0,19	0,03	0,03	0,04
p_{00}	0,17	0,18	0,17	0,04	0,04	0,04
$p_{1.}$	0,18	0,13	0,21	0,06	0,05	0,03
$p_{0.}$	0,19	0,19	0,18	0,03	0,04	0,04
SUL						
p_{11}	0,50	0,51	0,52	0,11	0,11	0,06
p_{10}	0,49	0,50	0,46	0,11	0,11	0,08
p_{01}	0,49	0,48	0,48	0,09	0,08	0,09
p_{00}	0,45	0,44	0,45	0,10	0,10	0,11
$p_{1.}$	0,49	0,50	0,49	0,11	0,11	0,07
$p_{0.}$	0,47	0,46	0,46	0,09	0,09	0,10

Fonte: Elaboração própria a partir dos dados do Prova Brasil, 2011.

Tabela 9: Análise de Sensibilidade do Efeito Médio do Tratamento sobre os Tratados — Português

Nordeste	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,029	-0,029	1,221	0,992	0%
	(25%, 50%]	-0,014	-0,014	1,235	1,114	0%
	(50%, 75%]	-0,032	-0,032	1,245	1,129	0%
Salário Baixo	(0%, 25%]	-0,029	-0,028	0,934	1,702	2%
	(25%, 50%]	-0,014	-0,013	0,930	1,639	-4%
	(50%, 75%]	-0,032	-0,032	0,881	1,317	0%
Sudeste	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,044	-0,044	1,085	1,017	0%
	(25%, 50%]	-0,056	-0,056	1,041	1,132	0%
	(50%, 75%]	-0,008	-0,008	1,036	1,079	0%
Salário Baixo	(0%, 25%]	-0,056	-0,056	1,057	1,501	0%
	(25%, 50%]	-0,058	-0,058	1,084	0,881	0%
	(50%, 75%]	-0,010	-0,010	1,302	1,222	0%
Norte	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,028	-0,028	1,001	1,126	0%
	(25%, 50%]	-0,017	-0,018	0,971	0,859	-3%
	(50%, 75%]	0,022	0,021	1,005	1,439	2%
Salário Baixo	(0%, 25%]	-0,028	-0,027	0,924	1,616	2%
	(25%, 50%]	-0,017	-0,018	0,853	0,950	-3%
	(50%, 75%]	0,022	0,021	0,889	0,719	2%
Centro-Oeste	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,050	-0,050	1,202	1,000	0%
	(25%, 50%]	-0,051	-0,050	1,145	0,630	0%
	(50%, 75%]	-0,002	-0,001	1,129	1,217	1%
Salário Baixo	(0%, 25%]	-0,050	-0,049	0,941	1,936	1%
	(25%, 50%]	-0,051	-0,051	0,886	1,438	0%
	(50%, 75%]	-0,002	-0,001	0,958	0,831	1%
Sul	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,035	-0,037	1,121	1,119	-3%
	(25%, 50%]	-0,042	-0,042	1,192	1,191	0%
	(50%, 75%]	0,019	0,018	1,111	1,132	-2%
Salário Baixo	(0%, 25%]	-0,035	-0,035	0,745	1,119	0%
	(25%, 50%]	-0,042	-0,041	0,743	1,225	1%
	(50%, 75%]	0,019	0,018	0,874	0,737	-2%

Fonte: Elaboração própria a partir dos dados do Prova Brasil, 2011.

Tabela 9: Análise de Sensibilidade do Efeito Médio do Tratamento sobre os Tratados — Matemática (continuação)

Nordeste	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,016	-0,016	1,108	0,971	0%
	(25%, 50%]	0,039	0,039	1,074	1,104	0%
	(50%, 75%]	-0,025	-0,025	1,075	1,145	0%
Salário Baixo	(0%, 25%]	-0,016	-0,014	0,855	1,751	7%
	(25%, 50%]	0,039	0,041	0,859	1,656	-3%
	(50%, 75%]	-0,025	-0,025	0,849	1,296	0%
Sudeste	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,016	-0,016	1,078	1,032	0%
	(25%, 50%]	-0,040	-0,039	1,043	1,106	1%
	(50%, 75%]	-0,017	-0,017	1,050	1,064	0%
Salário Baixo	(0%, 25%]	-0,016	-0,016	0,824	1,459	0%
	(25%, 50%]	-0,040	-0,040	0,843	0,917	0%
	(50%, 75%]	-0,002	-0,002	0,860	1,152	0%
Norte	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,051	-0,051	0,990	1,082	0%
	(25%, 50%]	-0,015	-0,015	0,944	0,825	0%
	(50%, 75%]	0,022	0,021	0,999	1,434	2%
Salário Baixo	(0%, 25%]	-0,051	-0,051	0,996	1,613	0%
	(25%, 50%]	-0,015	-0,015	0,943	0,976	0%
	(50%, 75%]	0,022	0,021	0,970	0,732	2%
Centro-Oeste	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,098	-0,098	1,215	0,955	0%
	(25%, 50%]	-0,065	-0,063	1,217	0,631	2%
	(50%, 75%]	-0,006	-0,005	1,187	1,180	9%
Salário Baixo	(0%, 25%]	-0,098	-0,097	0,842	1,826	1%
	(25%, 50%]	-0,065	-0,064	0,803	1,425	1%
	(50%, 75%]	-0,006	-0,005	0,970	0,846	9%
Sul	Percentual com Estabilidade	ETT	ETT Simulado	Efeito Resultado	Efeito Seleção	Diferença
Local de Moradia	(0%, 25%]	-0,056	-0,055	1,178	1,130	1%
	(25%, 50%]	-0,012	-0,012	1,188	1,167	0%
	(50%, 75%]	0,004	0,004	1,143	1,117	0%
Salário Baixo	(0%, 25%]	-0,056	-0,055	0,837	1,161	1%
	(25%, 50%]	-0,012	-0,011	0,821	1,263	4%
	(50%, 75%]	0,004	0,004	0,864	0,706	0%

Fonte: Elaboração própria a partir dos dados do Prova Brasil, 2011.

firme, escolas com altos percentuais de professores com estabilidade, mais de 50%, o efeito negativo da falta de professores com estabilidade é menor ou insignificante, principalmente, comparando com os resultados dos intervalos inferiores.

Em comparação com os resultados das estimativas dos efeitos por Mínimos Quadrados Ordinários (MQO), o efeito negativo das faixas de percentuais de professores sem estabilidade são maiores no Sudeste, Norte, Centro-Oeste e Sul (apenas para Português), menor no Nordeste e praticamente igual no Sul (para Matemática). Esse resultado havia sido confirmado por Webbink (2005), o qual aponta que inúmeros resultados encontrados na literatura da função de produção educacional que utilizam MQO sobre o efeito dos fatores escolares podem não ser estatisticamente confiáveis em virtude de o processo educacional ser a princípio endógeno.

Após verificar o efeito negativo maior nas escolas com menor percentual de professores com estabilidade sobre os desempenhos dos estudantes, pretende-se realizar uma análise de sensibilidade das estimativas dos efeitos para atestar a robustez dos resultados encontrados sob a Suposição de Independência Condicional. Nesse sentido, para implementar a análise proposta por Ichino et al. (2006), faz-se necessário selecionar outra variável que represente uma variável não observada para simular os resultados encontrados. Essa escolha deve ser fundamentada sobre quais os valores dos parâmetros de referência da probabilidade da variável binária simulada, p_{ij} , apresentam alguma informação adicional sobre o efeito de uma variável não observada U . Ou seja, substitua satisfatoriamente uma variável não observada que poderia existir, porém, causar danos na inferência.

Duas variáveis observadas foram selecionadas como modelos para se determinar os valores de p_{ij} na simulação, as quais são: a escola contrata professores com baixos salários; a escola possui critério de admissão por local de moradia. A escolha dessas variáveis tem o objetivo de substituir variáveis não observadas que estão relacionadas ao mecanismo de escolha da escola pelos pais e as restrições impostas pela escola a essa escolha. Ademais, essas variáveis foram escolhidas porque em várias simulações apresentam características de uma variável não observada gerar $p_{10} > p_{00}$ e $p_{1j} > p_{0j}$, pois, como já foi frisado, variáveis com essas probabilidades poderiam levar a uma significativa e positiva estimativa do ETT mesmo na ausência de um verdadeiro efeito causal.

A Tabela 8 expõe as probabilidades p_{ij} das variáveis utilizadas como modelo, o critério de admissão por localização da moradia dos alunos e o indicador de salários baixos dos professores. Para facilitar a análise da tabela, por exemplo, pode-se interpretar p_{10} como sendo as probabilidades das escolas utilizarem como critério de admissão o local de moradia, dado que a escola tem de 0 a 25% de professores estáveis e os alunos em média estão abaixo da média, tendo como referência os exames de Português e Matemática, respectivamente. Já em relação à variável salário baixo, observa-se que p_1 é interpretado como sendo igual à probabilidade de a escola ter salário baixo dado que a escola tem de 0 a 25% de professores com estabilidade.

Verifica-se que a variável local de moradia para as regiões Nordeste, Norte e Sul apresenta percentuais entre 0,1 e 0,23 e para as regiões Sudeste e Sul, tais percentuais variam entre 0,44 a 0,55. Isto é, as probabilidades do critério de admissão por local de moradia dado o tratamento e o resultado observado são maiores nas regiões mais desenvolvidas do Brasil. Em relação à variável

salário baixo dos professores, as probabilidades não apresentam uma variabilidade tão acentuada, visto que variam entre 0,02 e 0,23. Dessa forma, pode-se afirmar que as probabilidades do critério de professores com salários baixos não apresentam certa variabilidade entre as regiões brasileiras, no entanto as maiores probabilidades são verificadas na região Nordeste, onde observa-se uma variação de 0,13 a 0,23, já para as demais regiões esses valores transitam na faixa de 0,02 a 0,12.

Os resultados da análise de sensibilidade são apresentados na Tabela 9, onde constam as estimações de ETT pelo Pareamento no Escore de Propensão, os simulados ETT, o Efeito Resultado e o Efeito Seleção da simulação e a diferença percentual dos dois efeitos. Devido ao elevado número de observações individuais constantes na base de dados do Prova Brasil e a utilização do programa SENSATT para a análise de sensibilidade, gerou-se uma amostra aleatória de 5% do total de alunos para as todas as regiões.

Analisando os efeitos resultado e seleção da análise de sensibilidade, todos os efeitos são positivos, mas a relação entre os efeitos varia em vários casos. Por outro lado, várias das simulações apresentaram efeitos maiores do que um, significando que, dadas as variáveis utilizadas como não observadas, ocorrem casos em que os efeitos possam sinalizar que a inserção das variáveis pode ter implicado numa estimativa significativa do ETT, mesmo quando não há uma relação causal. O único caso que em as variáveis escolhidas permitem uma inferência sobre a relação causal entre a estabilidade e o ETT simulado é para a variável local de moradia e salário baixo no Sudeste, que a diferença no efeito é zero na maioria dos tratamentos.

Em relação à comparação do ETT e ETT simulado, este último é nulo ou negativo na maioria dos casos. Isso implica que as variáveis não observadas simuladas enfraquecem apenas uma pequena quantidade das estimativas iniciais, em conjunto com a análise do efeito resultado e seleção, contribuindo para a robustez do impacto negativo da falta de professores com estabilidade. Inspeccionando toda a coluna das diferenças nas estimativas, extrai-se da análise de sensibilidade que a existência de um *confounder*, comportando-se como a variável local de moradia e salário baixo dos professores, pode responder entre 0 e 9% da estimativa inicial do efeito dos percentuais de professores serem maior que a média sobre o desempenho dos estudantes.

6 Considerações Finais

Este artigo se propôs a investigar o efeito da estabilidade dos professores sobre a qualidade do ensino das escolas públicas brasileiras, pois não se sabe se essa característica do serviço público é um fator motivador ou um entrave burocrático na contratação de professores de qualidade. Por essa razão, a comparação entre as regiões teve a pretensão de apresentar essa discussão analisando diferentes mercados de trabalho dos professores, que são marcados por diferenças de competitividade e no conjunto de restrições socioeconômicas, tanto das famílias como dos empregadores.

A partir da metodologia de Pareamento no Escore de Propensão aplicada aos dados de uma avaliação nacional da qualidade da educação, Prova Brasil de 2011, estimou-se o efeito de diferentes faixas de percentual de professores com estabilidade sobre o desempenho no exame de Língua Portuguesa dos alunos do 5º ano do Ensino Fundamental. A aplicação de uma Análise de Sen-

sibilidade possibilitou investigar a sensibilidade desta estimativa com relação à possível existência de uma variável não observável que afeta tanto os resultados potenciais como a seleção no tratamento. Ou seja, a partir de um exercício de simulação buscou-se revelar se as estimativas são robustas a desvios da Suposição de Independência Condicional que fundamenta a identificação da estimativa obtida pela estratégia de Pareamento.

Os resultados revelam que, em todas as regiões, quanto maior a falta de professores com estabilidade, menor é o desempenho dos alunos. Isso pode estar relacionado ao mercado de trabalho dos professores que, supostamente, pode ser menos ou mais competitivo, bem como indicar que o benefício não pecuniário da estabilidade é um importante benefício na seleção de profissionais que influenciam a melhoria da qualidade do ensino. Complementarmente, os resultados derivados da análise de sensibilidade, com suas especificidades e limitações, apontam que os efeitos negativos são robustos, dada a presença de uma variável simulada não observada que poderia invalidar a suposição de independência condicional, embora alguns resultados revelem o enfraquecimento do efeito em até 9%.

A pretensão deste artigo é propor um redirecionamento na discussão sobre as principais variáveis que influenciam a decisão de profissionais qualificados a ingressarem nas escolas públicas brasileiras. Nesse sentido, acredita-se que os resultados aqui obtidos contribuam com políticas educacionais voltadas para fatores escolares que influenciam a qualidade do ensino ofertado para os estudantes que tem restrições socioeconômicas à liberdade de escolha.

Referências Bibliográficas

Aakvik, A. (2001), 'Bounding a matching estimator: the case of a norwegian training program', *Oxford Bulletin of Economics and Statistics* **63**(1), 115–143.

Altonji, J. G. (1988), 'The effects of family background and school characteristics on education and labor market outcomes', Mimeo, Northwestern University.

Altonji, J. G., Elder, T. E. & Taber, C. R. (2005), 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools', *Journal of Political Economy* **113**(1), 151–184.

Barbosa Filho, F. H., Pessôa, S. A. & Afonso, L. E. (2009), 'Um estudo sobre os diferenciais de remuneração entre os professores das redes pública e privada de ensino', *Estudos Econômicos* **39**(3), 597–628.

Barros, R. P., Mendonça, R., dos Santos, D. D. & Quintaes, G. (2001), Determinantes do desempenho educacional no brasil, Technical report, IPEA.

URL: Disponível em: http://www.ipea.gov.br/pub/td/td_2001/td_0834.pdf. Acesso em: 26 out. 2011.

Becker, K. L. & Kassouf, A. L. (2012), 'Diferença salarial e aposentadoria dos professores do ensino fundamental', *Economia Aplicada* **16**(1), 77–104.

Becker, S. O. & Caliendo, M. (2007), 'Sensitivity analysis for average treatment effect', *Stata Journal* **7**(1), 71–83.

- Betts, J. (1995), 'Does school quality matter? evidence from the national longitudinal survey of youth', *The Review of Economics and Statistics* 77, 231–247.
- Biondi, R. & Felicio, F. (2007), Atributos escolares e o desempenho dos estudantes: uma análise em painel dos dados do saeb, Texto para discussão, INEP.
- Blundell, R. & Costa-Dias, M. (2009), 'Alternative approaches to evaluation in empirical microeconomics', *Journal of Human Resources* 44(3).
- Blundell, R., Dearden, L. & Sianesi, B. (2005), 'Evaluating the impact of education on earnings in the uk: models, methods and results from the ncds', *Journal of the Royal Statistical Society, Series A* 168(3), 473–512.
- Bresser-Pereira, L. C. (1996), 'Da administração pública burocrática à gerencial', *Revista do Serviço Público* 47(1).
- Caliendo, M., Hujer, R. & Thomsen, S. (2007), The employment effects of job creation schemes in germany — a microeconometric evaluation, Discussion Paper 1512, IZA.
- Caliendo, M. & Kopeinig, S. (2008), 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys* 22(1), 31–72.
- Curi, A. & Menezes Filho, N. A. (2010), 'Os determinantes dos gastos com educação no brasil', *Pesquisa e Planejamento Econômico* 40, 1–39.
- Dehejia, R. & Wabba, S. (1999), 'Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs', *Journal of the American Statistical Association* 94, 1053–1062.
- Diprete, T. & Gangl, M. (2004), 'Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments', *Sociological Methodology* 34(1), 271–310.
- Felicio, F. & Fernandes, F. (2005), O efeito da qualidade da escola sobre o desempenho escolar: Uma avaliação no ensino fundamental no estado de são paulo, Technical report.
URL: Disponível em: <http://ideas.repec.org/p/anp/en2005/157.html>. Acesso em: 21 out. 2011.
- Franco, A. M. P. (2009), Os determinantes da qualidade da educação no Brasil, Tese de doutorado, Universidade de São Paulo.
- Grogger, J. (1996), 'School expenditures and post-schooling earnings: Evidence from high school and beyond', *The review of Economics and Statistics* 78, 628–637.
- Hanushek, E. A. (1986), 'The economics of schooling: Production and efficiency in public schools', *Journal of Economic Literature* 29, 1141–1147.

- Hanushek, E. A. (1997), 'Assessing the effects of school resources on student performance: An update, educational', *Evaluation & Policy Analysis* **19**, 141–164.
- Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (1998), Teachers, schools and academic achievement, NBER Working Paper 6691.
- Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (2004), 'Why public schools lose teachers', *Journal of Human Resources* **39**(2), 326–354.
- Hanushek, E. A. & Rivkin, S. G. (2007), 'Pay, working conditions and teacher quality', *Future of Children* **17**(1), 68–86.
- Heckman, J., Ichimura, H. & Todd, P. (1998), 'Matching as an econometric evaluation estimator', *Review of Economic Studies* **65**(2), 261–294.
- Heckman, J. & Todd, P. E. (2009), 'A note on adapting propensity score matching and selection models to choice based samples', *Econometrics Journal* **12**, S230–S34.
- Hoxby, C. M. & Leigh, A. (2004), 'Pulled away or pushed out? explaining the decline of teacher aptitude in the united states', *American Economic Review* **94**(2), 236–240.
- Ichino, A., Mealli, F. & Nannicini, T. (2006), From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity, Discussion Paper 2149, IZA, Bonn.
- Imbens, G. (2000), 'The role of the propensity score in estimating dose–response functions', *Biometrika* **87**(3), 706–710.
- Imbens, G. (2003), 'Sensitivity to exogeneity assumptions in program evaluation', *American Economic Review* **93**(2), 126–132.
- Imbens, G. (2004), 'Nonparametric estimation of average treatment effects under exogeneity: a review', *Review of Economics and Statistics* **86**(1), 4–29.
- Imbens, G. M. & Wooldridge, J. M. (2008), Recent developments in the econometrics of program evaluation, NBER Working Paper 14251.
- Khandker, S., Koolwal, G. & Samad, H. (2010), Handbook on impact evaluation, Technical report, World Bank, Washington DC.
- Lalonde, R. J. (1986), 'Evaluating the econometric evaluations of training programs with experimental data', *The American Economic Review* **76**(4), 604–620.
- Lechner, M. (1999), 'Earnings and employment effects of continuous off-the-job training in east germany after unification', *Journal of Business Economic Statistics* **17**(1), 74–90.
- Lechner, M. (2001), 'Identification and estimation of causal effects of multiple treatments under the conditional independence assumption', *Econometric Evaluation of Labour Market Policies, Heidelberg* p. 1–18.

Loeb, S. & Page, M. E. (2000), 'Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation.'

URL: Disponível em: <http://www.econ.ucdavis.edu/faculty/mepage/econ250b/examining%20the%20link%20betwen%20wages%20and%20quality%20in%20the%20teacher%20workforce.pdf>. Acesso em: 26 out. 2010.

Long, J. S. & Freeze, J. (2001), *Regression Models for Categorical Dependent Variables Using Stata*, Sata Press.

Louzano, P. (2010), 'Quem quer ser professor? atratividade, seleção e formação docente no brasil', *Estudos em Avaliação Educacional* 21(47), 543–568.

Menezes Filho, N. (2007), *Os Determinantes do Desempenho Escolar no Brasil*, Instituto Futuro Brasil/IBMEC, São Paulo.

Morgan, S. L. & Winship, C. W. (2007), 'Counterfactuals and causal inference methods and principles for social research', *Cambridge University Press, New York*.

Nannicini, T. (2007), 'A simulation-based sensitivity analysis for matching estimators', *The Stata Journal* 7(3), 334–350.

Pires, J. C. S. & Macedo, K. B. (2006), 'Cultura organizacional em organizações públicas no brasil', *Revista de Administração Pública, Rio de Janeiro* 40(1), 81–104.

Rice, J. K. (2003), *Teacher quality: understanding the effectiveness of teacher attributes*, Economic Policy Institute.

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *Econometrica* 73(2), 417–458.

Rosenbaum, P. (2002), *Observational Studies*, Springer, New York.

Rosenbaum, P. & Rubin, D. (1983a), 'Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome', *Journal of the Royal Statistical Society, Series B* 45, 212–218.

Rosenbaum, P. & Rubin, D. (1983b), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* 70(1), 41–50.

Scorzafave, L. G. & Ferreira, R. A. (2011), 'Desigualdade de proficiência no ensino fundamental público brasileiro: Uma análise de decomposição', *Economia* 12(2), 337–360.

Smith, J. & Todd, P. (2005), 'Does matching overcome lalonde's critique of nonexperimental estimators?', *Journal of Econometrics* 125(1-2), 305–353.

Todd, P. (2008), Evaluating social programs with endogenous program placement and selection of the treated, in 'Handbook of Development Economics', Vol. 60, pp. 384–389.

Webbink, D. (2005), 'Causal effects in education', *Journal of Economic Surveys* 19, 535–560.

Wobmann, L. (2003), 'Schooling resources, educational institutions and student performance: the international evidence', *Oxford Bulletin of Economics and Statistics* **65**(2), 117–170.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

Apêndice A

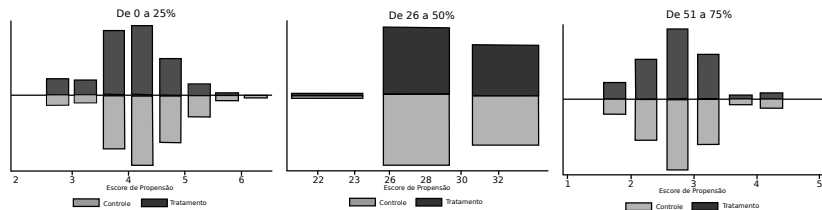


Figura A.1: Sobreposição do Escore de Propensão para a região Nordeste

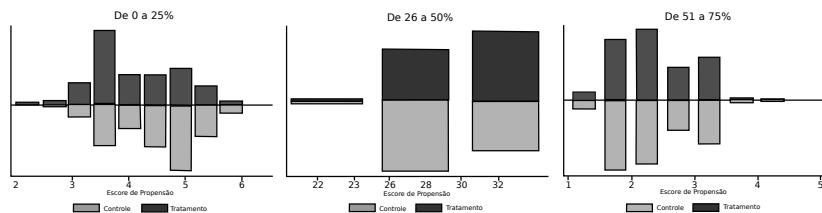


Figura A.2: Sobreposição do Escore de Propensão para a região Sudeste

DETERMINANTES SOCIOECONÔMICOS, ESTRUTURAS PRODUTIVAS REGIONAIS E CONDIÇÃO OCUPACIONAL NO BRASIL, 2000-2010

ANA CAROLINA DA CRUZ LIMA *

RODRIGO SIMÕES †

ANA MARIA HERMETO ‡

Resumo

O objetivo do artigo é analisar os principais determinantes da condição de ocupação individual. Foram utilizados os microdados dos Censos Demográficos 2000 e 2010 (IBGE) para estimar modelos multinomiais e, assim, identificar em que medida varia a probabilidade de ocupação de indivíduos com características semelhantes, porém, residentes em microrregiões distintas. Os resultados dos modelos demonstram que quanto mais desenvolvida a estrutura produtiva regional, mais complexa é a inserção individual no mercado de trabalho, pois apesar de existirem mais oportunidades emprego, há maior concorrência pelos postos gerados, que tendem a ser intensivos em capital.

Palavras-chave: Desenvolvimento Local; Emprego; Modelo Multinomial

Abstract

The main goal of this paper is to analyze the determinants of the individual employment condition. Using data from the demographic censuses of 2000 and 2010 (IBGE) we estimate multinomial models to verify to what extent the probability of employment of similar individuals varies from a region to another. The results of the models show that the more developed is a region, the more complex is the labor insertion. Despite of existing more labor opportunities in big cities, there are also more competition between individuals in these regions, especially because the new jobs are capital intensive.

Keywords: Local Development; Employment; Multinomial Model.

JEL classification: J20, J24, R23

DOI: <http://dx.doi.org/10.1590/1413-8050/ea97903>

* PPGCE/FCE/UERJ. Rio de Janeiro, Rio de Janeiro, Brasil. E-mail: ana.lima@uerj.br

† Cedeplar/UFMG. Belo Horizonte, Minas Gerais, Brasil. E-mail: limoes@cedeplar.ufmg.br

‡ Cedeplar/UFMG. Belo Horizonte, Minas Gerais, Brasil. E-mail: ahermeto@cedeplar.ufmg.br

1 Introdução

A análise da dinâmica dos mercados de trabalho brasileiros é bastante discutida no meio acadêmico, principalmente em relação ao seu comportamento nas macrorregiões do país, dadas as diferenças em seus níveis de desenvolvimento. Estudos realizados nessa área procuram demonstrar como as disparidades regionais de renda, originadas no processo de desenvolvimento da economia brasileira e intensificadas em seu período de industrialização recente, refletem-se em diferenciais de salários entre as regiões (ocupações e trabalhadores com características similares são remunerados de forma diferenciada no espaço) e como as características individuais dos trabalhadores, não necessariamente produtivas, influenciam seus níveis de remuneração.

Nesse último caso, destacam-se os trabalhos de Valenzuela (1999), Barros et al. (2007), Soares (2000). Esses autores analisam as diferenças nas oportunidades de trabalho e sua remuneração ao longo do território brasileiro provocadas por questões de gênero e raça. Essas características seriam elementos centrais nesse tipo de análise e, como destacado por Valenzuela (1999, p. 150), “estão por trás da construção social da desigualdade e da expressão que esta adquire no mercado de trabalho”. Em torno do gênero é estruturada a diferença entre trabalho remunerado (produtivo) e doméstico (reprodutivo), bem como as diferenças dentro do trabalho produtivo entre ocupações e posições de maior prestígio social (que favorece os homens). Ao mesmo tempo, as diferenças no mercado de trabalho baseadas na questão racial refletem-se principalmente nas ocupações: os negros concentram-se nas atividades fundamentalmente manuais e de menor valorização social e remuneração, enquanto os brancos concentram-se, preferencialmente, em atividades mais elaboradas, de maior remuneração e valorização social. Os autores demonstram que desde meados da década de 1980, a posição dos indivíduos negros no mercado de trabalho melhorou, tanto em termos de participação quanto de remuneração, mas essas características ainda implicam piores condições de inserção no mercado. Destacam também que as diferenças de gênero e raça tornam-se mais significativas em funções que exigem maior nível de escolaridade e que no Brasil há forte sensibilidade dos salários em relação ao nível educacional. Assim, a educação torna-se um fator tradicionalmente associado às possibilidades de inserção ocupacional e de mobilidade social. Os autores evidenciam o importante papel que o nível de qualificação pessoal desempenha sobre os diferenciais de salários, entretanto, enfatizam que estas diferenças são insuficientes para explicar esse fenômeno, sendo de fundamental importância considerar as características não produtivas dos indivíduos.

Ramos & Ferreira (2005), Ramos (2007), Barros et al. (2007), Fontes et al. (2010) avançam nessa análise ao considerarem que a ocupação e os diferenciais de remuneração entre os mercados de trabalho brasileiros também são segmentados no espaço. Ramos & Ferreira (2005), Ramos (2007) utilizam dados da PNAD para analisar a dinâmica espacial dos mercados de trabalho brasileiros no período subsequente à implantação do Plano Real e na primeira metade da década de 2000. Os autores evidenciam que a realocação espacial dos indivíduos ocupados e desocupados nos mercados de trabalho brasileiros foi bastante heterogênea, especialmente quando são comparadas as dinâmicas das áreas metropolitanas e não metropolitanas, bem como o desempenho das macrorregiões e Unidades da Federação do país. A análise da distribuição do emprego no território permite identificar que a proporção de indivíduos

ocupados aumentou de forma mais expressiva nas áreas urbanas não metropolitanas do país, bem como nas regiões norte, sudeste e centro-oeste. Em contrapartida, o volume de indivíduos desocupados cresceu de forma mais expressiva nas regiões metropolitanas, especialmente nas regiões sudeste e nordeste.

Barros et al. (2007), por sua vez, utilizam dados da PNAD (1995- 2005) para evidenciar que as diferenças salariais de trabalhadores com características idênticas e empregados em atividades similares chegam a 40% entre as regiões brasileiras. Essas diferenças existem porque homens e mulheres, negros e brancos, são remunerados de forma diferenciada, inclusive no espaço, apesar de serem perfeitamente substituíveis no processo de produção. Em análise convergente, Fontes et al. (2010) estimam modelos hierárquicos a partir de dados censitários (1991 e 2000) para mensurar a influência das aglomerações urbanas e das estruturas produtivas locais sobre os salários nas cidades médias e regiões metropolitanas brasileiras. Os autores identificam a manutenção de disparidades no rendimento médio do trabalho de indivíduos com características produtivas e não produtivas similares, todavia, residentes em centros urbanos localizados em diferentes regiões do Brasil. Os resultados dos modelos indicam que as disparidades regionais (estruturas produtivas, nível de escolaridade, etc.) se refletem em dinâmicas diferenciadas dos mercados de trabalho e, consequentemente, em diferentes níveis salariais e ocupacionais.

Uma justificativa para os resultados encontrados por todos esses autores (Ramos & Ferreira 2005, Ramos 2007, Barros et al. 2007, Fontes et al. 2010) reside na desigual distribuição das atividades produtivas no país e nos impactos das mudanças estruturais pós-1990 (abertura comercial e reestruturação produtiva) sobre as mesmas. Por esse motivo é importante compreender os fatores que levam as atividades a se aglomerarem em determinadas regiões, tornando-as mais dinâmicas e, consequentemente, estimulando seus níveis salariais médios e taxas de ocupação e desemprego. A perspectiva adotada neste artigo para explicar essa dinâmica foi elaborada pelos teóricos da Economia Urbana, mais especificamente por Henderson (1988), Glaeser (1999). Esses estudiosos utilizam o conceito de economias externas, relacionadas a *spillovers* tecnológicos, para explicar os fatores que podem favorecer a concentração espacial das atividades produtivas. Essas externalidades são classificadas como economias de localização e de urbanização. No primeiro caso, as firmas são beneficiadas pela concentração de firmas da mesma indústria ou indústrias similares em determinado centro urbano, como destacado por Marshall (1982) — economias externas às firmas, mas internas à aglomeração. No segundo caso, as firmas são beneficiadas pela diversificação produtiva, pela própria escala urbana e pela oferta de serviços sofisticados à produção, como evidenciado por Jacobs (1969) — economias internas à indústria e à localidade, mas externas às firmas. Assim, nos grandes centros urbanos há uma quantidade elevada de firmas e trabalhadores e uma oferta diversificada de serviços produtivos, o que cria um ambiente propício à inovação e favorece o surgimento de novas atividades. Estas, por sua vez, incentivam o desenvolvimento do sistema de transportes, das comunicações, de atividades financeiras, etc., o que torna esses centros extremamente atraentes do ponto de vista da produção e do consumo. A diversidade implica maior flexibilidade e adequação às mudanças, o que torna a economia menos vulnerável¹.

¹A aglomeração é acompanhada por deseconomias, relacionadas aos custos de deslocamento

Nesse contexto, o artigo procura avaliar a influência das características locais sobre a dinâmica de seus respectivos mercados de trabalho, sem desconsiderar as particularidades dos indivíduos. A ideia subjacente é que maiores níveis de desenvolvimento implicam alterações nas estruturas produtivas em prol de atividades de alto teor tecnológico e do setor terciário, o que provavelmente afeta sua estrutura ocupacional. O objetivo é analisar a probabilidade dos indivíduos estarem ou não ocupados de acordo com suas características, produtivas e não produtivas, considerando-os no espaço – representado por variáveis indicadoras do nível de desenvolvimento socioeconômico da localidade de residência (microrregiões brasileiras). Para sua consecução, serão estimados modelos multinomiais para avaliar a influência desses fatores sobre a probabilidade de ocupação, desocupação e inatividade individual. As informações utilizadas para a análise foram obtidas a partir dos microdados dos Censos Demográficos de 2000 e 2010, elaborados pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A análise dos resultados dos modelos evidencia as diferenças existentes entre homens e mulheres e a importância das estruturas regionais para a situação ocupacional individual: quanto mais desenvolvida uma região, mais complexa é a inserção no mercado de trabalho, pois apesar de existirem maiores oportunidades emprego, há maior concorrência pelos postos gerados (excesso de mão de obra provocado pela atratividade dos grandes centros urbanos), que tendem a ser intensivos em capital. A qualificação individual (educação) continua a ser a medida mais eficaz para contrabalançar essa tendência, especialmente para as mulheres.

O trabalho possui mais três seções além desta introdução. A metodologia e a base de dados utilizadas são descritas na segunda seção. Os resultados dos modelos são analisados na terceira seção. Em seguida, são realizadas as considerações finais.

2 Metodologia e Base de Dados

A problemática que o artigo se propõe a analisar — a condição de ocupação individual — é representada por uma variável categórica que assume três possibilidades mutuamente excludentes: 00 (zero) para indivíduos inativos, ou seja, que não tinham ocupação e não haviam procurado trabalho na data de referência dos censos; 01 (um) para indivíduos ativos e ocupados; e 02 (dois) para indivíduos ativos e desocupados, ou seja, que haviam procurado emprego na data de referência dos censos, mas não possuíam ocupação. O procedimento econométrico escolhido para analisar essa condição é o modelo de regressão logística multinomial, cujas principais propriedades são descritas detalhadamente por Cameron & Trivedi (2005). O método de estimação adotado é a Máxima Verossimilhança e os resultados do modelo são interpretados a partir de suas razões relativas de risco (*odds ratio*). As *odds ratio* demonstram a mudança relativa nas probabilidades das categorias analisadas, ou seja, elas informam como a probabilidade de escolher a categoria de contraste relativamente à alternativa de referência muda quando determinada variável independente

e às desamenidades urbanas (poluição, criminalidade, exclusão social, elevação da renda fundiária, aceleração da migração interna, aumento do desemprego e subemprego urbanos etc.), que contrabalançam a concentração. Todavia, estas não impedem o desenvolvimento urbano; apenas reduzem seu ritmo de expansão (Henderson 1988). A cidade é o elemento chave deste processo, pois nela são criadas as melhores condições para a troca de informações e o fluxo de produtos e serviços. A cidade representa o equilíbrio entre as forças aglomerativas e de dispersão.

umenta uma unidade. Assim, a variável X aumenta (diminui) a probabilidade de que a alternativa j seja escolhida ao invés da alternativa de referência se a *odds ratio* for maior (menor) que 01. O objetivo é identificar a probabilidade de ocupação, desocupação e inatividade dos indivíduos de acordo com suas características e as características do local de residência.

2.1 Fonte de dados

A base de dados utilizada para a estimação dos modelos foi construída a partir dos microdados dos Censos Demográficos 2000 e 2010, fornecidos pelo IBGE. As informações presentes nos censos demográficos englobam diversas características socioeconômicas da população brasileira e de seus respectivos locais de residência, incluindo aspectos demográficos, educacionais, de trabalho e rendimento, migração, habitação e fecundidade. Em relação à abrangência geográfica, os censos demográficos possuem elevado grau de desagregação, tornando possíveis análises em pequenas escalas espaciais.

A amostra selecionada para o trabalho é composta por indivíduos agrupados em 557 unidades espaciais, referentes às microrregiões geográficas brasileiras (exceto Fernando de Noronha). Considerando o objetivo proposto, foram realizadas algumas restrições amostrais para garantir uma homogeneidade mínima ao perfil dos indivíduos analisados e, conseqüentemente, facilitar a interpretação dos impactos da estrutura econômica local sobre a condição de ocupação dos indivíduos.

Em primeiro lugar, foram excluídos os indivíduos com idade inferior a 25 anos e superior a 64 anos. Esse recorte etário tem duas justificativas: a primeira refere-se à necessidade de excluir da amostra indivíduos em idade escolar, que ainda não completaram o seu ciclo educacional e que possuem comportamentos específicos relacionados ao ingresso inicial no mercado de trabalho; a segunda refere-se à necessidade de excluir indivíduos que possuem maior probabilidade de inatividade, por já terem ultrapassado o limite superior do tempo de contribuição previdenciária.

Com o intuito de minimizar comportamentos muito heterogêneos, também foram excluídos da análise indivíduos residentes em áreas rurais. A importância de excluir o setor rural da análise é explicada pela elevada disparidade entre os mercados de trabalho rurais e urbanos no Brasil em termos de taxas de ocupação, desocupação e informalidade, bem como pela ênfase teórica adotada no artigo. Ao supor que o desenvolvimento local e a diversificação produtiva das cidades têm impactos significativos sobre as estruturas ocupacionais, os autores optam por priorizar a análise dos mercados de trabalho urbanos no Brasil.

Ajustes adicionais, como a exclusão de indivíduos que se autodeclararam indígenas e/ou amarelos e residentes em domicílios coletivos, improvisados ou sem chefe, foram realizados para minimizar possíveis erros de estimação.

Adicionalmente aos recortes realizados, a amostra foi dividida em dois grupos: homens e mulheres. A razão para essa divisão refere-se às diferenças de gênero que existem nos mercados de trabalho brasileiros, destacadas na introdução, que tendem a favorecer indivíduos do sexo masculino. A decisão de estimar modelos específicos para cada grupo tem como objetivo aumentar o poder explicativo das variáveis baseadas em características da estrutura produtiva local, auxiliando a identificação da segmentação regional dos mercados de trabalho.

Os fatores que influenciam a inserção de um indivíduo no mercado de trabalho podem ser classificados, *a priori*, em duas grandes categorias. A primeira está associada às características dos próprios indivíduos (condições socioeconômicas) e a segunda às características do ambiente em que os mesmos estão inseridos (infraestrutura econômica, institucional, composição da força de trabalho, etc.). Além disso, ressalta-se que a condição de migração individual é inserida na análise com o intuito de captar um possível viés de seleção nos mercados de trabalho brasileiros, uma vez que os migrantes são positivamente selecionados (mais jovens e mais escolarizados) e tendem a possuir maior probabilidade de inserção ocupacional, conforme evidenciam Lima et al. (2011)². As duas primeiras tabelas descrevem as variáveis indicadoras de atributos pessoais incluídas na análise:

Tabela 1: Variável Dependente

Variável	Nome	Descrição
sitocup	Situação Ocupacional	Variável categórica que assume o valor 0 para indivíduos inativos, 1 para indivíduos ativos e ocupados e 2 para indivíduos ativos e desocupados.

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

Foram criados 08 (oito) indicadores para evidenciar a dinâmica socioeconômica das microrregiões brasileiras entre 2000 e 2010, cujas descrições podem ser observadas na tabela 03 abaixo. A construção desses indicadores procurou cobrir aspectos essenciais do processo de desenvolvimento econômico, enfatizando a crescente importância das aglomerações urbanas. A proporção de domicílios com esgotamento sanitário ligado à rede geral é utilizada como *proxy* da infraestrutura urbana; quanto mais elevada essa proporção, maior a capacidade regional de minimizar os custos de congestionamento das grandes cidades. O valor médio do rendimento domiciliar *per capita* e a média de anos de estudo evidenciam o nível de crescimento regional, pois valores mais expressivos desses indicadores equivalem a regiões mais atrativas do ponto de vista econômico (há trabalhadores mais qualificados e consumidores com maior potencial de renda disponível).

O grau de industrialização, a oferta de serviços modernos, a proporção da força de trabalho qualificada e a taxa de formalização das ocupações evidenciam o grau de diversidade regional e o aumento da complexidade inerente aos grandes centros urbanos; eles representam o padrão de crescimento das atividades produtivas e o estágio de desenvolvimento urbano. Esses indicadores ajudam a identificar as vantagens e desvantagens da especialização e da

²Além disso, se a existência de endogeneidade do processo migratório não é considerada na análise, os estimadores obtidos são inconsistentes e possuem viés, pois existe uma relação significativa entre a condição de ocupação individual, a dinâmica produtiva e dos mercados de trabalho locais e a decisão de migrar. Por esse motivo, antes de analisar a probabilidade de um indivíduo estar ou não ocupado é preciso considerar a probabilidade de o mesmo ser migrante, o que é realizado por meio do modelo de Seleção Amostral de Heckman, conforme evidenciam Cameron & Trivedi (2005). Os resultados do modelo evidenciam que os homens, brancos, jovens, mais escolarizados e que vivem em situação conjugal estável são os indivíduos que apresentam maior probabilidade de efetivar um deslocamento. Ou seja, indivíduos com essas características constituem um grupo que possui maior probabilidade de migrar em comparação aos demais.

Tabela 2: Variáveis explicativas indicadoras de características individuais

Variável	Nome	Descrição
branco	Branco	Dummy com valor igual a 1 se o indivíduo é branco e 0 caso contrário.
idade	Idade	Idade em anos.
idade2	Experiência profissional	Proxy para experiência profissional, que é igual ao quadrado da idade do indivíduo.
respdom	Responsável pelo domicílio	Dummy com valor igual a 1 se o indivíduo é o responsável pelo domicílio e 0 caso contrário.
pmigrante*	Probabilidade de ser migrante	Probabilidade de um indivíduo ser migrante de acordo com suas características produtivas (anos de estudo e experiência profissional) e não produtivas (sexo, idade, cor e área de residência - rural, urbana ou metropolitana). Os valores desta variável foram obtidos a partir da estimação de um modelo probit para a condição de migração individual.
anosest	Anos de estudo	Quantidade de anos de estudo do indivíduo.
metropolitana	Área Metropolitana	Dummy com valor igual a 1 se o indivíduo reside em área metropolitana e 0 caso contrário.

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

(*) O critério de migração refere-se à data fixa: migrante é o indivíduo que mora há menos de 05 anos na microrregião de recenseamento, ainda que tenha nascido na mesma. Não migrante é o indivíduo que nasceu e sempre morou na localidade ou que mora há pelo menos 05 anos na mesma.

diversificação urbanas destacadas por Duraton & Puga (2002) e sua influência sobre a inserção ocupacional. Mais especificamente, a oferta de serviços modernos é fundamental para explicar as mudanças econômicas recentes, visto que a mesma tornou-se elemento essencial para o crescimento econômico (a integração entre indústria e serviços reflete os ajustes realizados para manter a competitividade no cenário econômico mundial em constantes transformações). Segundo Marshall & Wood (1995), os serviços sofisticados concentram-se justamente nas regiões mais desenvolvidas, todavia, as ocupações desse ramo tendem a ser cada vez mais seletivas e intensivas em capital, tornando mais complexa a inserção dos trabalhadores.

A taxa de pobreza foi calculada de acordo com as recomendações de Foster et al. (1984) e seu objetivo é identificar como a estagnação econômica de uma região pode afetar a condição de ocupação individual.

Por fim, foram criadas *dummies* regionais (nordeste, norte, centro-oeste, sul, sudeste e São Paulo), para verificar se há alguma influência macrorregional sobre as probabilidades de ocupação, desocupação e inatividade.

Tabela 3: Variáveis explicativas indicadoras das características microrregionais

Variável	Nome	Descrição
<i>tx_esgot</i>	Taxa de esgotamento sanitário	Proporção de domicílios da microrregião com esgotamento sanitário ligado à rede geral.
<i>med_anosest</i>	Média de anos de estudo	Média de anos de estudo da microrregião.
<i>grau_ind</i>	Grau de industrialização	Proporção de pessoas ocupadas nos setores industriais da microrregião.
<i>of_servmoder*</i>	Oferta de serviços modernos	Proporção de pessoas ocupadas em atividades modernas do setor de serviços da microrregião.
<i>ft_qualif</i>	Força de trabalho qualificada	Proporção de pessoas ocupadas na microrregião que possuem mais de 11 anos de estudo.
<i>tx_formal</i>	Taxa de trabalho formal	Proporção de pessoas ocupadas em atividades formais na microrregião.
<i>renddompc_{medio}</i>	Rendimento domiciliar per capita médio	Rendimento domiciliar per capita médio da microrregião.
<i>pobreza**</i>	Taxa de pobreza	Indicador de pobreza da microrregião

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

(*) Os seguintes setores foram classificados como serviços modernos: serviços de telecomunicações, serviços financeiros e seguros, serviços jurídicos, de contabilidade e auditoria, consultorias, publicidade, engenharia e arquitetura, desenho e outros serviços técnico-profissionais prestados a empresas, serviços auxiliares da indústria e instituições científicas e tecnológicas.

(**) A linha de pobreza equivale a 1/4 do salário mínimo vigente em cada ano censitário. A adoção deste critério é justificada por esta ser a principal referência para a inclusão de indivíduos em programas sociais do Governo Federal.

3 Situação ocupacional no Brasil urbano: os principais determinantes da escolha individual

Serão estimados três modelos empíricos de regressão logística multinomial para cada grupo de análise utilizando as variáveis descritas na seção 3.1. O primeiro modelo (nº1A para homens e nº1B para mulheres) inclui apenas as variáveis explicativas indicadoras das características individuais; o segundo modelo (nº2A para homens e nº2B para mulheres) incorpora as variáveis dos atributos regionais, controlando os resultados do modelo numa perspectiva espacial; o terceiro modelo (nº3A para homens e nº3B para mulheres) incorpora *dummies* macrorregionais com o objetivo de identificar as principais tendências dos mercados de trabalho brasileiros.

3.1 Análise descritiva das variáveis

Para contextualizar a discussão é interessante observar algumas características gerais da amostra selecionada para o estudo. A maior parte dos indivíduos que a compõem encontra-se ocupada, e essa proporção aumentou no período

analisado (de 64,4% para 70,5%) todavia, há uma diferença significativa entre o percentual de homens e mulheres que possuem trabalho. Conforme pode ser observado na tabela 04, a taxa de ocupação masculina é muito superior à feminina e o oposto é válido para a taxa de inatividade. É interessante observar que a proporção de desocupados é relativamente baixa e diminuiu no período para ambos os grupos, o que parece indicar que as principais escolhas individuais transitam entre a inatividade e a atividade. É importante ressaltar que os dados evidenciam o aumento da participação feminina nos mercados de trabalho nos anos mais recentes, corroborando estudos sobre o tema.

A maior parte da amostra é composta por indivíduos brancos (mas esse percentual diminuiu de 59,5% em 2000 para 51,7% em 2010), responsáveis pelo domicílio (55%) e não migrantes (90%). A tabela 5 demonstra a distribuição por sexo de cada uma dessas categorias. Observa-se que os homens continuam a ser os principais responsáveis pelos domicílios brasileiros e que a mobilidade individual parece não ter um recorte por gênero.

Tabela 4: Composição amostral por categoria ocupacional

	2000						
	Inativos		Ativos		Total	Ativos	
	Ocupados	Desocupados	Ocupados	Desocupados			
Homens	311310	2235432	205168	2751910	11,31%	81,23%	7,46%
Mulheres	1245908	1418225	255888	2920021	42,67%	48,57%	8,76%
Total	1557218	3653657	461056	5671931	-	-	-
	2010						
	Inativos		Ativos		Total	Ativos	
	Ocupados	Desocupados	Ocupados	Desocupados			
Homens	421431	2651098	99344	3171873	13,29%	83,58%	3,13%
Mulheres	1256619	1994496	164063	3415178	36,80%	58,40%	4,80%
Total	1678050	4645594	263407	6587051	-	-	-

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

A análise descritiva dos indicadores de desenvolvimento local evidencia a problemática regional brasileira. Os desequilíbrios regionais originados no processo de desenvolvimento da economia nacional persistem e os dados utilizados são capazes de mostrar a complexidade desse processo, que ultrapassa a discussão macrorregional. As disparidades de renda, emprego e oportunidades no país diferem não apenas entre suas macrorregiões, mas também em uma escala microrregional. A tabela 6 mostra os valores médios, mínimos e máximos dos indicadores regionais calculados a partir dos microdados dos censos e é possível observar uma grande variabilidade entre os mesmos.

Em todos os casos, observa-se que as microrregiões com melhor desempenho estão localizadas na região centro-sul do país e nas proximidades das capitais das regiões Norte e Nordeste. Foram raríssimos os casos em que os indicadores das microrregiões do Norte-Nordeste ultrapassaram a média nacional. Por outro lado, as microrregiões mais dinâmicas, localizadas especialmente no interior de São Paulo, algumas regiões de Minas Gerais (Metropolitana de Belo Horizonte, Sul e Triângulo), do Paraná (Curitiba, Maringá), de Santa Catarina (Florianópolis, Joinville, Blumenau), do Rio Grande do Sul (Porto Alegre) e do Mato Grosso do Sul, obtiveram valores superiores à média nacional.

Tabela 5: Distribuição amostral por cor, condição de migração e posição no domicílio, Brasil (2000-2010)

(%)	Branco	Pretos ou pardos	Migrantes	Não migrantes	Responsável pelo domicílio	Outros
2000						
Homens	47,62	49,84	50,46	48,32	75,73	14,93
Mulheres	52,38	50,16	49,54	51,68	24,27	85,07
Total	100	100	100	100	100	100
2010						
Homens	47,58	48,76	50,14	47,98	62,21	30,55
Mulheres	52,42	51,24	49,86	52,02	37,79	69,45
Total	100	100	100	100	100	100

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

Tabela 6: Valores médios, máximos e mínimos dos indicadores regionais (Brasil, 2000-2010)

	2000			
	Média	Desvio padrão	Mínimo	Máximo
Taxa de esgoto	0,4807	0,2958	0,0000	0,9407
Grau de industrialização	0,1304	0,0588	0,0188	0,4262
Oferta de serviços modernos	0,1031	0,0437	0,0000	0,2110
Força de trabalho qualificada	0,2884	0,0978	0,0309	0,5163
Taxa de formalização	0,3818	0,1258	0,0287	0,5659
Pobreza	0,3267	0,1725	0,0851	0,8146
Média de anos de estudo	6,33	1,34	2,39	8,52
Rendimento domiciliar per capita médio*	440,12	189,42	99,57	881,73

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

(*) R\$ 2010.

Nº de microrregiões: 557

Esse recorte macrorregional não excluiu a existência de áreas estagnadas no centro-sul do país, como, por exemplo, o norte de Minas Gerais, microrregiões do Rio de Janeiro, Espírito Santo, Goiás, Mato Grosso e Mato Grosso do Sul (exceto aquelas nucleadas pelas capitais) e localidades do interior do Rio Grande do Sul e do Paraná.

3.2 Análise dos resultados dos modelos

A comparação entre as condições de ocupação dos indivíduos é realizada por intermédio das seguintes alternativas:

- $y_i = 0$ se o indivíduo é inativo;
- $y_i = 1$ se o indivíduo é ativo e possui ocupação;
- $y_i = 2$ se o indivíduo é inativo e está desocupado.

A categoria de referência (base) no problema analisado é a inatividade. O objetivo da estimação é avaliar como as características individuais e do local de residência influenciam a condição de ocupação dos indivíduos, visto que as mesmas podem afetar significativamente a inserção no mercado de trabalho.

A escolha ocupacional masculina

Os resultados da estimação dos modelos multinomiais (razões de riscos relativos) podem ser observados na tabela 7. Como destacado anteriormente, a categoria de referência é inativo, logo, a interpretação dos resultados dos modelos deve considerar a probabilidade de um indivíduo estar ocupado (ou desocupado) em detrimento da inatividade.

Nos três modelos estimados, observa-se que a probabilidade de ocorrência da alternativa ativo e ocupado aumenta, quando comparada à categoria inativo, de acordo com as variáveis idade, responsável pelo domicílio, probabilidade de ser migrante e anos de estudo. Esses resultados indicam que características individuais tradicionalmente beneficiadas pelos mercados de trabalhos, como escolaridade, facilitam a inserção ocupacional. Além disso, a migração tende a favorecer essa inserção, o que parece confirmar as hipóteses de seletividade do processo migratório — os migrantes, em geral, são

mais qualificados e empreendedores do que a média da população (Lima et al. 2011). Para visualizar essas considerações observe as estimativas do modelo 03A (completo). É possível perceber que a probabilidade de um indivíduo responsável pelo domicílio estar ocupado é maior do que a de um indivíduo que não ocupa essa posição (aproximadamente 90% em 2000 e 30% em 2010); para indivíduos mais velhos os percentuais são, respectivamente, 36% e 55%; para homens com elevada probabilidade de migração os percentuais são, respectivamente, 52% e 58%. Nota-se, todavia, que a probabilidade de ocupação não aumenta para indivíduos brancos, o que provavelmente pode ser justificado pela correlação existente entre essa variável e a variável *pmigrante*, assim como esse resultado pode refletir, em alguma medida, a redução da segmentação dos mercados de trabalho brasileiros em relação à cor dos indivíduos. Além disso, os modelos evidenciam que residir em áreas metropolitanas não implica necessariamente maior facilidade de obter trabalho. Na realidade, o elevado grau de diversificação dos grandes centros urbanos metropolitanos torna mais complexa a dinâmica laboral, exige maior qualificação individual e aumenta a concorrência entre a mão de obra, dificultando o ingresso no mercado de trabalho (*odds ratio* menores do que 01).

A estimação dos modelos 2A e 3A evidencia que o processo de inserção ocupacional é bastante complexo e depende não apenas dos níveis de qualificação individual, mas também das características das estruturas produtivas regionais. Seus resultados indicam, inclusive, que a migração *per se* não é capaz de aumentar a inserção individual nos mercados de trabalho, pois a obtenção de novo emprego dependerá não apenas dos atributos individuais, mas também das características produtivas e dos mercados de trabalho das regiões de destino dos migrantes. Em outras palavras, a dinâmica econômica das regiões de destino dos migrantes é essencial para a determinação de seu processo de inserção ocupacional.

Os resultados do modelo 2A evidenciam que a residência em áreas mais dinâmicas, com elevados níveis educacionais médios e expressiva densidade urbana, torna mais difícil o processo de inserção ocupacional dos indivíduos. Nessas microrregiões, os mercados de trabalho são caracterizados por expressivos graus de formalização e de participação das atividades do setor de serviços modernos (intensivos em tecnologia e mão de obra especializada), o que implica elevada competição pelas vagas geradas de emprego e, consequentemente, maior dificuldade para a inserção ocupacional dos trabalhadores. As razões de riscos relativos para as variáveis indicadoras dessas características (taxa de esgoto, média de anos de estudo, oferta de serviços modernos e taxa de formalização) são menores do que 01 em todo o período analisado, indicando que a residência nestas localidades diminui a probabilidade de ocupação individual. Esse resultado, que aparentemente é contraditório, evidencia a complexidade inerente aos mercados de trabalho urbanos (já destacada pela variável indicadora de áreas metropolitanas), pois quanto mais densa a localidade, maior a competitividade e apenas indivíduos altamente qualificados obterão sucesso no processo de inserção ocupacional. Ainda em relação às variáveis indicadoras da densidade urbana, observa-se que a razão de riscos relativos para a proporção da força de trabalho qualificada é maior do que 01 em 2000, indicando elevada probabilidade de ocupação para indivíduos residentes em microrregiões com maior parcela da força de trabalho desempenhando atividades técnico-científicas em comparação aos não residentes. Em 2010, todavia, essa variável passa a corroborar os resultados obtidos para os

Tabela 7: Razão de riscos relativos de regressão logística multinomial para a condição de ocupação individual homens (Brasil, 2000-10)

	Categoria 1 = ativo e ocupado						Categoria 2 = ativo e desocupado					
	Modelo 1A		Modelo 2A		Modelo 3A		Modelo 1A		Modelo 2A		Modelo 3A	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
branco	0,9481*** (0,0000)	0,8576*** (0,0000)	0,9022*** (0,0000)	0,7892*** (0,0000)	0,9061*** (0,0000)	0,7983*** (0,0000)	0,7772*** (0,0000)	0,7276*** (0,0000)	0,7694*** (0,0000)	0,7824*** (0,0000)	0,7683*** (0,0000)	0,7958*** (0,0000)
idade	1,3580*** (0,0000)	1,5505*** (0,0000)	1,3597*** (0,0000)	1,5510*** (0,0000)	1,3600*** (0,0000)	1,5516*** (0,0000)	1,2098*** (0,0000)	1,2064*** (0,0000)	1,2091*** (0,0000)	1,2062*** (0,0000)	1,2094*** (0,0000)	1,2069*** (0,0000)
idade2	0,9965*** (0,0000)	0,9954*** (0,0000)	0,9965*** (0,0000)	0,9954*** (0,0000)	0,9965*** (0,0000)	0,9954*** (0,0000)	0,9974*** (0,0000)	0,9974*** (0,0000)	0,9974*** (0,0000)	0,9975*** (0,0000)	0,9974*** (0,0000)	0,9974*** (0,0000)
respdom	1,9366*** (0,0000)	1,3124*** (0,0000)	1,9119*** (0,0000)	1,3017*** (0,0000)	1,9127*** (0,0000)	1,3014*** (0,0000)	1,0125*** (0,0000)	0,9126*** (0,0000)	1,0159*** (0,0000)	0,9158*** (0,0000)	1,0144*** (0,0000)	0,9156*** (0,0000)
pmigrante	1,5236*** (0,0000)	1,5785*** (0,0000)	1,5239*** (0,0000)	1,5820*** (0,0000)	1,5234*** (0,0000)	1,5828*** (0,0000)	1,2619*** (0,0000)	1,1895*** (0,0000)	1,2627*** (0,0000)	1,1898*** (0,0000)	1,2618*** (0,0000)	1,1906*** (0,0000)
anosest	1,0277*** (0,0000)	1,0082*** (0,0000)	1,0296*** (0,0000)	1,0067*** (0,0000)	1,0300*** (0,0000)	1,0066*** (0,0000)	0,9567*** (0,0000)	0,9994* (0,0870)	0,9544*** (0,0000)	0,9995 (0,1290)	0,9546*** (0,0000)	0,9993** (0,0440)
metropolitana	0,8199*** (0,0000)	0,9109*** (0,0000)	0,9492*** (0,0000)	0,9542*** (0,0000)	0,9708*** (0,0000)	0,9482*** (0,0000)	1,2328*** (0,0000)	1,0348*** (0,0000)	1,0226*** (0,0000)	0,9734*** (0,0000)	1,0788** (0,0000)	0,9681*** (0,0000)
tx_esgot			0,8810*** (0,0000)	0,8816*** (0,0000)	1,2815*** (0,0000)	1,0213*** (0,0000)			0,9501*** (0,0000)	0,9812*** (0,0120)	1,1400*** (0,0000)	1,0729*** (0,0000)
med_anosest			0,8084** (0,0000)	0,9941 (0,1500)	0,8444** (0,0000)	1,1013*** (0,0000)			1,0681*** (0,0000)	0,7196*** (0,0000)	1,1231*** (0,0000)	0,8516*** (0,0000)
grau_ind			0,8210*** (0,0000)	1,3765*** (0,0000)	0,6616*** (0,0000)	1,2726*** (0,0000)			0,4144*** (0,0000)	0,2472*** (0,0000)	0,2274*** (0,0000)	0,2216*** (0,0000)
of_servmoder			0,4590*** (0,0000)	0,2272*** (0,0000)	0,5017*** (0,0000)	0,2698*** (0,0000)			3,4998*** (0,0000)	0,5699*** (0,0000)	2,1306*** (0,0000)	0,6641*** (0,0000)
ft_qualif			4,6691*** (0,0000)	0,5068*** (0,0000)	1,4825*** (0,0000)	0,1160*** (0,0000)			0,4919*** (0,0000)	6,1546*** (0,0000)	0,4677*** (0,0000)	5,0093*** (0,0000)
tx_formal			0,1548*** (0,0000)	0,4901*** (0,0000)	0,1614*** (0,0000)	0,5467*** (0,0000)			1,9823*** (0,0000)	2,0520*** (0,0000)	2,2014*** (0,0000)	2,3054*** (0,0000)
renddompc_medio			1,0003*** (0,0000)	1,0003*** (0,0000)	1,0004*** (0,0000)	1,0004*** (0,0000)			1,0002*** (0,0000)	0,9998*** (0,0000)	1,0000 (0,5670)	1,0000*** (0,0060)
pobreza			0,1453*** (0,0000)	0,1172*** (0,0000)	0,1389*** (0,0000)	0,1080*** (0,0000)			1,3719*** (0,0000)	1,1617*** (0,0000)	2,3335*** (0,0000)	0,7924*** (0,0000)
nordeste					0,9229*** (0,0000)	0,9894*** (0,0000)					0,7853*** (0,0000)	1,0198*** (0,0000)
norte					=	=					=	=
centro					1,0530*** (0,0000)	1,0196*** (0,0000)					1,0751*** (0,0000)	0,8874*** (0,0000)
sul					0,8620*** (0,0000)	0,8134*** (0,0000)					0,9740*** (0,0000)	0,6991*** (0,0000)
sudeste					0,7163*** (0,0000)	0,8162*** (0,0000)					0,7790*** (0,0000)	0,7674*** (0,0000)
sp					0,7150*** (0,0000)	0,8339*** (0,0000)					1,0183** (0,0210)	0,8030*** (0,0000)
Pseudo-R2	0,1174	0,0982	0,1204	0,1029	0,1212	0,1032	0,1174	0,0982	0,1204	0,1029	0,1212	0,1032

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

(***) Variável significativa a 1%; (**) Variável significativa a 5%; (*) Variável significativa a 10%.

A categoria de referência é igual a inativo (sitocup==0). N°observações 2000 = 2751910; n°observações 2010 = 3171873.

Tabela 8: Probabilidades para as equações de inatividade, ocupação e desocupação dos homens (Brasil 2000-2010)

2000								
Modelo 1A			Modelo 2A			Modelo 3A		
Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado
0,0765	0,8512	0,0723	0,0762	0,8516	0,0722	0,0760	0,8521	0,0719
2010								
Modelo 1A			Modelo 2A			Modelo 3A		
Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado
0,0989	0,8717	0,0294	0,0983	0,8732	0,0285	0,0982	0,8734	0,0284

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.
A categoria de referência é igual a inativo (sitocup==0).

demais indicadores da densidade urbana, conforme descrito anteriormente. Ou seja, em 2010, o elevado grau de competição dos mercados de trabalho dos grandes centros urbanos metropolitanos minimiza eventuais aumentos de probabilidade de ocupação que poderiam se estimulados pelo desenvolvimento de atividades intensivas em conhecimento nas microrregiões brasileiras (o aumento do emprego nesse setor não necessariamente implica aumento da probabilidade de obtenção de emprego devido aos expressivos níveis de qualificação e treinamento que essas ocupações exigem).

Observa-se que as estimações para o grau de industrialização corroboram esses resultados para 2000, todavia, em 2010, a residência em uma microrregião com elevado grau de industrialização volta a contribuir para o aumento da probabilidade de ocupação individual (aproximadamente 37%), o que provavelmente está relacionado à recuperação do crescimento do emprego nesse setor, conforme evidenciam dados da Pesquisa Industrial Mensal de Emprego e Salário do IBGE. Em outras palavras, as estimações para a variável *grau_ind* parecem refletir que a importância desse setor para a inserção ocupacional dos trabalhadores brasileiros continua a ser expressiva, especialmente em seus momentos de expansão. Salienta-se ainda, que a contribuição da renda domiciliar *per capita* média para a obtenção de emprego é positiva, mas pouco expressiva. Por fim, o indicador de pobreza evidencia que em regiões estagnadas é mais difícil obter trabalho, muito provavelmente pela falta de dinamismo local (escassez de empregos).

A estimação do modelo completo (modelo 3A) evidencia que indivíduos residentes na região Centro-Oeste possuem probabilidade de ocupação ligeiramente superior a dos residentes em outras macrorregiões do país (5% em 2000 e 2% em 2010). As demais variáveis indicadoras das macrorregiões (nordeste, sul, sudeste e São Paulo) apresentam razões de riscos relativos menores do que um, indicando maior dificuldade de inserção ocupacional para indivíduos residentes nas mesmas em comparação a não residentes³. As *odds ratio* das regiões sudeste, sul e de São Paulo são as menos expressivas, indicando, provavelmente, a dificuldade de obter emprego em áreas de intenso dinamismo (elevada competição por postos de trabalho). Vale salientar que nesse modelo, a única alteração das razões de riscos relativos em comparação aos modelos anteriores (1B e 2B) ocorre para a variável taxa de esgotamento sanitário. Ou seja, ao inserir as variáveis indicadoras das macrorregiões de residência no modelo fica evidente o *trade off* existente no processo de decisão ocupacional dos trabalhadores residentes em grandes áreas urbanas metropolitanas: por um lado, a expressiva escala urbana pode contribuir para a obtenção de emprego devido à capacidade local de gerar novos postos de trabalho; por outro lado, a complexidade dos mercados de trabalhos urbanos, que exigem elevados níveis de qualificação e treinamento da mão de obra, e a expressiva competição pelos postos de trabalho dificultam a inserção ocupacional dos indivíduos.

Ao comparar a categoria ativo e desocupado com a alternativa de inatividade, percebe-se que a probabilidade de desocupação aumenta, principalmente, com a probabilidade de migração individual e com a idade. Um aumento percentual na probabilidade de ser migrante eleva em aproximadamente 26% a probabilidade de desocupação individual em 2000 e 19% em

³A variável “norte” não foi incluída na análise, pois devido às suas características produtivas (voltadas principalmente para atividades agrícolas) foi escolhida para se a *baseline* regional.

2010; para indivíduos mais velhos, esse aumento é de 20% nos dois anos analisados. As explicações para essas probabilidades residem, provavelmente, na dificuldade de adaptação dos migrantes aos novos contextos sociais e a incapacidade dos mercados de trabalho em absorver indivíduos mais velhos, especialmente os idosos, devido ao seu elevado custo de treinamento (possuem vícios trabalhistas que dificilmente são revertidos). Pessoas com maiores níveis de escolaridade possuem uma probabilidade de desocupação ligeiramente inferior a dos não escolarizados, provavelmente porque esses indivíduos são mais competitivos nos mercados de trabalho. Indivíduos que residiam em áreas metropolitanas em 2000 possuíam maior probabilidade de desocupação, situação que não é mais observada em 2010, principalmente quando são incorporadas variáveis regionais aos modelos. Muito provavelmente, a dificuldade de obtenção de trabalho nas grandes regiões metropolitanas brasileiras (altamente competitivas) pode ter estimulado a saída de alguns indivíduos da população economicamente ativa (desalento no mercado de trabalho). Ou seja, as razões de riscos relativos da variável “*metropolitana*” nos modelos que incorporam as características regionais retratam a dificuldade de obter emprego nos grandes centros urbanos brasileiros devido à elevada oferta de mão de obra e aos níveis de exigência dos empregadores. O desalento também pode ser utilizado para explicar porque a probabilidade de desocupação dos indivíduos responsáveis pelo domicílio caiu entre 2000 e 2010 em comparação aos demais indivíduos.

Nessa comparação (desocupados *versus* inativos), os indicadores regionais possuem impactos bastante diversificados. As razões de riscos relativos da taxa de esgotamento sanitário (*proxy* para infraestrutura urbana) evidenciam que indivíduos residentes em microrregiões com níveis mais elevados de urbanização tendem a possuir menor probabilidade de desocupação (modelo 2A) – preferem a inatividade devido às dificuldades de obtenção de emprego nos grandes centros urbanos metropolitanos. Contudo, ao inserir na análise as variáveis indicadoras das macrorregiões de residência (modelo 3A), essas *odds ratio* tornam-se maiores do que 01, destacando que, dadas as particularidades macrorregionais, a escala urbana tende a estimular a permanência individual na população economicamente ativa, ainda que o sucesso (inserção) não seja obtido no curto prazo. A residência em microrregiões com elevados níveis médios de escolaridade e expressiva oferta de serviços modernos aumentava a probabilidade de desocupação em 2000, oposto do resultado obtido para 2010. Essa mudança provavelmente é justificada pelo desalento dos trabalhadores urbanos ao longo do período analisado: o elevado nível de exigência dos mercados de trabalho das regiões caracterizadas por alta escolaridade e atividades intensivas em conhecimento/tecnologia dificulta a inserção ocupacional dos indivíduos, estimulando assim a saída dos mesmos da população economicamente ativa (*trade off* entre desocupação e inatividade). Relação similar é observada para a variável força de trabalho qualificada, porém, invertendo os valores das *odds ratio* para cada ano analisado. Esse resultado (presença de força de trabalho qualificada passa a estimular a permanência na população economicamente ativa, ainda que o indivíduo não obtenha emprego) pode estar relacionado ao aumento geral dos níveis médios de escolaridade observados na economia brasileira ao longo dos últimos anos (expectativa de que o aumento individual da escolaridade poderá facilitar a inserção nos mercados de trabalho, que valorizam a educação formal).

O nível de formalização dos mercados de trabalho funciona durante todo

o período analisado como um fator de estímulo à permanência no grupo populacional economicamente ativo. Muito provavelmente, as expectativas de condições de trabalho e emprego mais favoráveis em microrregiões com elevadas taxas de formalização têm estimulado a continuidade da busca por um novo posto de trabalho, aumentando, assim, a probabilidade de desocupação em comparação à inatividade. A forte presença do setor industrial, por sua vez, diminui a probabilidade de desocupação durante todo o período analisado. Ou seja, indivíduos residentes em microrregiões com elevados níveis de industrialização possuem menor probabilidade de ficar sem trabalho quando comparados a indivíduos residentes em regiões com indicadores menos expressivos, pois as primeiras regiões provavelmente possuem maiores oportunidades de emprego.

Assim como é observado para a probabilidade de ocupação, a contribuição da renda domiciliar *per capita* média para desocupação é positiva, mas pouco expressiva. Por fim, o indicador de pobreza evidencia que em regiões estagnadas é mais difícil obter trabalho devido ao baixo dinamismo local, o que aumenta a probabilidade de desocupação (exceto para o modelo completo em 2010, ou seja, mais recentemente, indivíduos residentes em localidades com elevadas taxas de pobreza e de desemprego têm optado pela inatividade).

A influência da localização macrorregional sobre a probabilidade de estar desocupado em comparação à inatividade é relativamente homogênea, mas as chances de estar desempregado em 2010 são maiores na região Nordeste. Ressalta-se que em 2000, a residência nessa região refletia-se em menor probabilidade de desocupação (maior inatividade). Essa alteração nas *odds ratio* da região pode ser explicada pelo aumento da competitividade nos mercados de trabalho locais nos anos mais recentes, estimulada inclusive pela migração de retorno. Salienta-se também que a probabilidade de desocupação na região Centro-Oeste e no estado de São Paulo eram mais elevadas em 2000 (7,5% e 1,8%, respectivamente), situação oposta à vigente em 2010.

De forma geral, os resultados dos modelos para a categoria ativo e desocupado indicam que os indivíduos brancos, responsáveis pelo domicílio, com maior nível educacional, residentes em microrregiões com elevados níveis de escolaridade e de industrialização possuem menores chances de permanecer fora dos mercados de trabalho. Essa probabilidade é ainda menor se a microrregião de residência estiver localizada na região Sul do país. Em contrapartida, indivíduos que residem em microrregiões com elevadas taxas de qualificação da mão de obra e formalização possuem, maior probabilidade de desocupação.

As informações da tabela 7 também evidenciam que os indivíduos negros, não responsáveis pelo domicílio, com menor escolaridade, residentes em áreas metropolitanas e com elevadas taxas de diversificação produtiva possuem maior probabilidade de inatividade, pois a inadequação dos mesmos aos requerimentos dos mercados de trabalho funciona como um estímulo à sua saída dos setores produtivos (os indivíduos optam por não desempenhar funções econômicas).

Por fim, as probabilidades médias de ocupação, desocupação e inatividade do sexo masculino podem ser observadas na tabela 8. Não há variabilidade significativa dessas probabilidades entre os modelos estimados, contudo, ao inserir as características microrregionais nota-se um pequeno aumento nas chances de obter um emprego. De modo geral, os homens brasileiros residentes em áreas urbanas possuem uma elevada probabilidade de ocupação (entre

0,85 e 0,87), especialmente quando comparados às mulheres, o que poderá ser comprovado a partir da análise dos modelos estimados para este grupo.

A escolha ocupacional feminina

As mulheres possuem características não produtivas significativamente diferentes dos homens, o que afeta seu ingresso e permanência nos mercados de trabalho, bem como sua respectiva remuneração. Essas diferenças evidenciam a importância de separar esses grupos para tornar a análise dos determinantes da escolha ocupacional mais homogênea. Os resultados gerais dos modelos estimados para as mulheres evidenciam a influência dessas diferenças sobre a escolha ocupacional feminina. Os dados expostos na tabela 09 demonstram que a probabilidade média de uma mulher estar ocupada é significativamente inferior àquela observada para os homens (0,50 *versus* 0,85 em 2000 e 0,61 *versus* 0,87 em 2010, respectivamente). A probabilidade de inatividade feminina era aproximadamente cinco vezes mais elevada que a masculina em 2000 e o triplo da masculina em 2010. Essa elevada inatividade está relacionada principalmente ao *trade off* que as mulheres continuam a enfrentar entre desempenhar atividades produtivas ou dedicar-se a tarefas não produtivas, como a administração do lar, apesar das melhorias obtidas no período analisado.

Assim como na análise masculina, a categoria de referência é inativo, logo, a interpretação dos resultados dos modelos considera a probabilidade de uma mulher estar ocupada (ou desocupada) em detrimento da inatividade. Os resultados da estimação dos modelos multinomiais estão descritos na tabela 10.

As estimativas dos modelos para as mulheres mantêm, de forma geral, as mesmas características dos modelos para os homens. No caso da alternativa ativa e ocupada, as principais diferenças provavelmente referem-se ao aumento da probabilidade de conseguir um emprego em comparação à inatividade para mulheres brancas no final do período e à incapacidade da migração aumentar essa probabilidade. A não contribuição da migração para a obtenção de uma ocupação pode estar relacionada ao fato de a migração feminina ser majoritariamente de acompanhamento familiar, todavia, os dados analisados neste artigo são insuficientes para realizar essa afirmação. Mulheres mais velhas possuem chances mais elevadas de ter um emprego (a probabilidade é aproximadamente 6% maior do que a observada para mulheres mais jovens em 2000 e 13% em 2010), assim como aquelas que são responsáveis pelos seus respectivos domicílios (aproximadamente 90% em 2000 e 40% em 2010) e que possuem maiores níveis de escolaridade (entre 15% e 16% em todo o período analisado).

Mulheres que residiam em microrregiões com elevados níveis de escolaridade média possuíam maior dificuldade de inserção ocupacional em 2000, todavia, em 2010 a residência nessas localidades gera para as mesmas uma probabilidade de ocupação 17% mais elevada em comparação às mulheres que residem em áreas de baixo nível educacional (modelo completo 3B). A moradia em microrregiões industrializadas aumenta essa probabilidade significativamente, provavelmente porque há maiores oportunidades de trabalho geradas pelas intensas relações setoriais dessas atividades, com expressiva incorporação da mão de obra feminina nos últimos anos. Regiões com elevados níveis médios de rendimento domiciliar *per capita* também favorecem suas moradoras em comparação às residentes em áreas estagnadas, mas de forma bastante sutil (*odds ratio* próximas à unidade); o oposto ocorre em microrre-

Tabela 9: Razão de riscos relativos de regressão logística multinomial para a condição de ocupação individual mulheres (Brasil, 2000-10)

	Categoria 1 = ativo e ocupado						Categoria 2 = ativo e desocupado					
	Modelo 1B		Modelo 2B		Modelo 3B		Modelo 1B		Modelo 2B		Modelo 3B	
branco	0,9726*** (0,0000)	1,1435*** (0,0000)	0,8956*** (0,0000)	1,0131*** (0,0000)	0,8806*** (0,0000)	1,0059*** (0,0000)	0,8108*** (0,0000)	0,8425*** (0,0000)	0,7789*** (0,0000)	0,8452*** (0,0000)	0,7676*** (0,0000)	0,8412*** (0,0000)
idade	1,0622*** (0,0000)	1,1325*** (0,0000)	1,0583*** (0,0000)	1,1275*** (0,0000)	1,0583*** (0,0000)	1,1276*** (0,0000)	0,9716*** (0,0000)	1,0074*** (0,0000)	0,9694*** (0,0000)	1,0073*** (0,0000)	0,9692*** (0,0000)	1,0074*** (0,0000)
idade2	0,9980*** (0,0000)	0,9978*** (0,0000)	0,9981*** (0,0000)	0,9979*** (0,0000)	0,9981*** (0,0000)	0,9979*** (0,0000)	0,9987*** (0,0000)	0,9985*** (0,0000)	0,9987*** (0,0000)	0,9985*** (0,0000)	0,9987*** (0,0000)	0,9985*** (0,0000)
respdom	1,9299*** (0,0000)	1,4013*** (0,0000)	1,9460*** (0,0000)	1,4143*** (0,0000)	1,9460*** (0,0000)	1,4130*** (0,0000)	1,4360*** (0,0000)	1,2352*** (0,0000)	1,4355*** (0,0000)	1,2446*** (0,0000)	1,4364*** (0,0000)	1,2442*** (0,0000)
pmigrante	0,7545*** (0,0000)	0,8688*** (0,0000)	0,7508*** (0,0000)	0,8632*** (0,0000)	0,7507*** (0,0000)	0,8632*** (0,0000)	0,7379*** (0,0000)	0,8014*** (0,0000)	0,7371*** (0,0000)	0,8042*** (0,0000)	0,7367*** (0,0000)	0,8042*** (0,0000)
anosest	1,1630*** (0,0000)	1,1500*** (0,0000)	1,1655*** (0,0000)	1,1519*** (0,0000)	1,1659*** (0,0000)	1,1520*** (0,0000)	1,0636*** (0,0000)	1,0867*** (0,0000)	1,0606*** (0,0000)	1,0834*** (0,0000)	1,0609*** (0,0000)	1,0834*** (0,0000)
metropolitana	0,9334*** (0,0000)	1,0121*** (0,0000)	0,9896*** (0,0000)	1,0272*** (0,0000)	1,0036*** (0,2220)	1,0087*** (0,0000)	1,2999*** (0,0000)	1,1778*** (0,0000)	1,0057*** (0,0230)	1,0274*** (0,0000)	1,0500*** (0,0000)	1,0394*** (0,0000)
tx_esgot			0,8735*** (0,0000)	0,9562*** (0,0000)	1,1613*** (0,0000)	1,0724*** (0,0000)			1,0002 (0,9560)	1,2377*** (0,0000)	1,0804*** (0,0000)	1,2936*** (0,0000)
med_anosest			0,9065*** (0,0000)	1,1320*** (0,0000)	0,9293*** (0,0000)	1,1788*** (0,0000)			1,1348*** (0,0000)	0,8712*** (0,0000)	1,1802*** (0,0000)	0,9309*** (0,0000)
grau_ind			1,4365*** (0,0000)	2,7600*** (0,0000)	1,0224* (0,0720)	2,2647*** (0,0000)			0,5265*** (0,0000)	0,2714*** (0,0000)	0,2539*** (0,0000)	0,1893*** (0,0000)
of_servmoder			1,1803*** (0,0000)	1,0434 (0,1380)	0,5416*** (0,0000)	0,9476* (0,0750)			7,3273*** (0,0000)	0,7580*** (0,0000)	2,4104*** (0,0000)	0,4492*** (0,0000)
ft_qualif			1,4594*** (0,0000)	0,0945*** (0,0000)	0,7245*** (0,0000)	0,0665*** (0,0000)			0,3256*** (0,0000)	9,0362*** (0,0000)	0,2931*** (0,0000)	3,4461*** (0,0000)
tx_formal			0,3480*** (0,0000)	0,4831*** (0,0000)	0,3068*** (0,0000)	0,4771*** (0,0000)			2,7703*** (0,0000)	1,3020*** (0,0000)	2,7346*** (0,0000)	1,2766*** (0,0000)
renddompc_medio			0,9999*** (0,0000)	1,0004*** (0,0000)	1,0002*** (0,0000)	1,0004*** (0,0000)			0,9997*** (0,0000)	0,9998*** (0,0000)	0,9997*** (0,0000)	0,9998*** (0,0000)
pobreza			0,1875*** (0,0000)	0,1888*** (0,0000)	0,1649*** (0,0000)	0,2039*** (0,0000)			1,0643*** (0,0040)	0,4400*** (0,0000)	1,4260*** (0,0000)	0,3424*** (0,0000)
nordeste					1,1408*** (0,0000)	1,0522*** (0,0000)					0,9083*** (0,0000)	1,0104** (0,0120)
norte					-	-					-	-
centro					0,9582*** (0,0000)	1,0322*** (0,0000)					0,8798*** (0,0000)	0,8767*** (0,0000)
sul					1,1579*** (0,0000)	1,0860*** (0,0000)					1,1043*** (0,0000)	0,9416*** (0,0000)
sudeste					0,9329*** (0,0000)	0,9628*** (0,0000)					0,9044*** (0,0000)	0,8915*** (0,0000)
sp					0,8901*** (0,0000)	0,9862*** (0,0000)					1,1086*** (0,0000)	0,9821*** (0,0010)
Pseudo-R2	0,0958	0,0829	0,0984	0,0886	0,0991	0,0887	0,0958	0,0829	0,0984	0,0886	0,0991	0,0887

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.

(***) Variável significativa a 1%; (**) Variável significativa a 5%; (*) Variável significativa a 10%.

A categoria de referência é igual a inativo (sitocup==0). N°observações 2000 = 2920021; n°observações 2010 = 3415178.

Tabela 10: Probabilidades para as equações de inatividade, ocupação e desocupação das mulheres (Brasil 2000-2010)

2000								
Modelo 1B			Modelo 2B			Modelo 3B		
Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado
0,4139	0,4992	0,0869	0,4142	0,4994	0,0864	0,4142	0,4996	0,0862
2010								
Modelo 1B			Modelo 2B			Modelo 3B		
Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado	Inativo	Ativo e ocupado	Ativo e desocupado
0,3428	0,6121	0,0451	0,3423	0,6132	0,0445	0,3424	0,6131	0,0445

Fonte: elaboração própria a partir dos microdados dos Censos Demográficos 2000 e 2010.
A categoria de referência é igual a inativo (sitocup==0).

giões com elevadas taxas de pobreza. As razões de riscos relativos das variáveis regionais indicadoras das estruturas trabalhistas (proporção de mão de obra qualificada e taxa de formalização) evidenciam a dificuldade de inserção ocupacional em mercados dinâmicos e sofisticados, pois residir em regiões com elevados valores dessas variáveis reduz as chances de obter um trabalho. É interessante observar que a contribuição da variável indicadora da densidade urbana — oferta de serviços modernos — varia significativamente entre os modelos 2B e 3B, evidenciando que a incorporação das particularidades macrorregionais na análise (*dummies* regionais) tende a corroborar a complexidade inerente aos mercados de trabalho dos grandes centros urbanos metropolitanos (maior concorrência pelos postos de trabalho, que exigem níveis mais elevados de treinamento e qualificação da mão de obra). Simultaneamente, as estimações demonstram que as chances de ocupação feminina em áreas metropolitanas tendem a ser ligeiramente mais elevadas em comparação à probabilidade ocupacional de mulheres residentes em áreas não metropolitanas nos anos mais recentes, um provável reflexo da maior capacidade de diversificação dos mercados de trabalho metropolitanos. As *dummies* regionais indicam que o ingresso feminino no mercado de trabalho é mais fácil nas regiões Sul e Nordeste do Brasil e mais recentemente na região Centro-Oeste.

Essas tendências evidenciam que a residência em microrregiões com maiores rendimentos, grau de industrialização e média de anos de estudo aumenta a probabilidade de ocupação feminina; a residência em microrregiões com elevadas taxas de formalização, mão de obra qualificada e pobreza diminui essa probabilidade por motivos diametralmente opostos. O aumento da escolaridade e a emancipação feminina (leia-se chefia do domicílio) são as formas mais factíveis de aumentar sua probabilidade de ocupação. A efetivação da migração reduz essa probabilidade.

Por fim, os fatores que determinam a escolha pela atividade e desocupação em detrimento da inatividade estão relacionados principalmente à chefia domiciliar, elevada escolaridade, residência em áreas metropolitanas (*odds ratio* das variáveis metropolitana e taxa de esgoto são maiores do que 01 em todos os modelos estimados) e a elevadas taxas de formalização. Mulheres que apresentam essas características tendem a elevar suas chances de desocupação em relação aos seus respectivos grupos (chefes de domicílios *versus* não chefes, residentes em áreas metropolitanas *versus* não residentes, etc.), pois preferem exercer algum tipo de atividade produtiva, mesmo que o custo de oportunidade de permanecer temporariamente fora do mercado de trabalho seja significativo. Por exemplo, as mulheres mais escolarizadas possuem chances entre 6% e 8% mais elevadas de estarem desocupadas quando comparadas às demais. Mulheres chefes do domicílio também possuem maior probabilidade de desocupação (aproximadamente 43% mais elevada em 2000 e 24% em 2010), pois tendem a se manter ativas mesmo em condições desfavoráveis devido à necessidade de prover seus respectivos domicílios.

É interessante observar que quanto mais elevada a probabilidade de uma mulher ser migrante, maior a probabilidade de inatividade da mesma. Esse resultado, assim como o obtido para a categoria 01 (ocupação), provavelmente está relacionado à importância da migração familiar para a efetivação do deslocamento feminino no país. Sob a hipótese do deslocamento ter sido realizado para acompanhar o cônjuge, a probabilidade de atividade feminina diminui, uma vez que a busca por novos postos de trabalho e emprego é limitada pela necessidade de desempenhar atividades não diretamente relacionadas à

dinâmica dos mercados de trabalho brasileiros.

As características microrregionais que reforçam a escolha pela busca de trabalho ao invés da inatividade são basicamente as mesmas observadas para os homens. De forma geral, mulheres que residem em regiões mais urbanizadas aumentam em 29% as chances de desocupação quando comparadas às não residentes; se essas regiões possuem elevada proporção da força de trabalho ocupada em atividades formais esse percentual é igual a 27% em 2010, ligeiramente mais favorável do que o observado em 2000. Essas informações evidenciam que nos centros urbanos, que oferecem maiores oportunidades de trabalho, as mulheres tendem a procurar emprego com mais frequência. As *dummies* macrorregionais apresentam razões de risco relativas menores do que 01, exceto nas regiões Sul e São Paulo em 2000 e na região Nordeste em 2010. Ou seja, mais recentemente, a região que apresenta maiores estímulos à continuidade da busca pelo emprego, apesar das dificuldades de inserção, é a região Nordeste (a mesma tem crescido a taxas mais expressivas do que as demais regiões do país, apesar de ainda apresentar indicadores de renda e desenvolvimento menos favoráveis).

Assim, a probabilidade de desocupação é menor para mulheres brancas, escolarizadas e experientes, que residem em microrregiões com elevado rendimento domiciliar *per capita* e grau de industrialização. Por outro lado, mulheres que residem em áreas mais urbanizadas e com percentuais expressivos de formalização e de qualificação da mão de obra têm maiores dificuldades de inserção; se essas mulheres são responsáveis pelo domicílio, essa probabilidade é ainda maior, pois as mesmas continuam a buscar o ingresso no mercado de trabalho para suprir suas necessidades econômicas.

Os mesmos fatores que contribuem para a desocupação feminina também estimulam a sua escolha pela inatividade econômica. Quanto maior o tempo de permanência na condição ativa e desocupada, maior a probabilidade de essa situação se reverter para a inatividade, especialmente no caso de mulheres que apresentam características não favorecidas pelos mercados de trabalho (negras e baixa escolaridade). Residir em microrregiões mais desenvolvidas (elevados níveis de renda e escolaridade média) pode diminuir essa probabilidade, todavia, quanto mais qualificada a mão de obra local e o nível de formalização dos mercados locais, maior a dificuldade de inserção e, consequentemente, maior a probabilidade de se tornar inativa. A situação econômica de mulheres que moram em regiões pobres também é bastante desfavorável, pois o aumento da taxa de pobreza eleva a probabilidade de inatividade. Cabe ressaltar, por fim, que as mulheres migrantes possuem elevada probabilidade de optar pela inatividade, pois as mesmas podem ter dificuldades de adaptação no destino ou podem estar acompanhando a migração familiar.

4 Considerações finais

A dinâmica dos mercados de trabalho brasileiros é bastante diferenciada entre suas regiões, seja em termos salariais, de participação ou de ocupação. Essas diferenças são provocadas pela composição dos próprios mercados de trabalho, cujos participantes possuem características específicas de seus respectivos grupos (homens *versus* mulheres, negros *versus* brancos), bem como pelas estruturas produtivas regionais.

O desenvolvimento recente de tecnologias, especialmente de comunica-

ções e transportes, e o processo de urbanização desempenham papel central na nova organização espacial da produção e têm implicações diretas sobre a condição de ocupação dos trabalhadores, pois em regiões mais populosas, cujas atividades principais são mais intensivas em capital e em serviços sofisticados, a inserção dos mesmos torna-se mais complexa.

Considerada a importância dessas diferenças, este trabalho procurou mensurar a probabilidade de um indivíduo estar ocupado, desocupado ou inativo, realizando controles por intermédio de características individuais e regionais. Além disso, para captar a influência das estruturas regionais, a amostra foi restrita às áreas urbanas e dividida por sexo, para garantir maior homogeneidade ao perfil dos grupos analisados, facilitando assim a identificação da segmentação regional. O procedimento metodológico escolhido para a análise foi a estimação de um modelo de regressão logística multinomial e a base de dados utilizada foi construída a partir dos microdados dos Censos Demográficos 2000 e 2010, fornecidos pelo IBGE. As variáveis incorporadas ao modelo referem-se a atributos individuais produtivos e não produtivos (cor, idade, probabilidade de migração, chefia domiciliar, escolaridade e residência em área metropolitana) e das respectivas microrregiões de residência (média de anos de estudo, rendimento domiciliar *per capita* médio, taxa de esgotamento sanitário, grau de industrialização, oferta de serviços modernos, proporção da mão de obra qualificada, taxa de formalização, taxa de pobreza e *dummies* regionais).

Os resultados dos modelos evidenciam a heterogeneidade existente entre homens e mulheres, pois a probabilidade de ocupação masculina é significativamente superior à feminina em todo o período analisado: em 2000, os homens possuíam, em média, 85% de chances de estarem ocupados, enquanto as mulheres possuíam 50%; em 2010, esses percentuais se alteraram para 87% e 61% respectivamente. As probabilidades de desocupação são similares entre os grupos, mas um pouco mais baixas para os homens (7,2% *versus* 8,6% em 2000 e 2,8% *versus* 4,5% em 2010, respectivamente) e as mulheres possuem maiores chances de permanecer na inatividade (41,4% *versus* 7,6% em 2000 e 34,4% *versus* 9,87% em 2010, respectivamente), o que provavelmente é um reflexo de sua dificuldade de inserção (tempo de permanência na desocupação) e/ou de sua escolha entre desempenhar atividades produtivas ou não produtivas (administração do lar). É importante ressaltar que ao longo do período analisado, a participação feminina na população economicamente ativa brasileira cresceu consideravelmente, o que provavelmente é um reflexo da diminuição da segmentação dos mercados de trabalho nacionais por sexo, conforme evidenciam Barros et al. (2007).

Em relação aos atributos pessoais, observa-se que indivíduos que possuem características não produtivas tradicionalmente beneficiadas pelo mercado de trabalho, como a idade e cor, possuem maior probabilidade de ocupação, sejam homens ou mulheres. A qualificação ainda é uma das principais formas de aumentar as chances de conseguir um emprego e a efetivação da migração possui efeitos diversos. No caso dos homens, os indivíduos com elevada probabilidade de migração possuem maior probabilidade de ocupação e de desocupação, o que evidencia a importância da região de destino (a incapacidade de se adaptar ao novo contexto socioeconômico pode dificultar ou atrasar o ingresso no mercado de trabalho). Para as mulheres, a migração diminui a probabilidade de ocupação e aumenta as chances de inatividade. A justificativa para esse resultado provavelmente está atrelada ao fato da migração

feminina ocorrer no contexto familiar e/ou a sua dificuldade de adaptação ou inadequação ao local de destino. Indivíduos responsáveis pelo domicílio possuem maiores probabilidades de ocupação e desocupação (para prover as necessidades do lar, os indivíduos procuram emprego mesmo em contextos desfavoráveis).

A importância das estruturas regionais é similar entre os grupos. As estimativas da variável indicadora de residência em área metropolitana são o primeiro indício da complexidade inerente aos mercados de trabalho urbanos: elas evidenciam que a probabilidade de ocupação é menor em áreas metropolitanas, enquanto as probabilidades de desocupação e inatividade são mais elevadas, tanto para homens quanto para mulheres. A análise das variáveis indicadoras do nível de desenvolvimento microrregional viabiliza a interpretação desses resultados. Áreas mais dinâmicas em termos educacionais, industriais e de rendimento tendem a ofertar maiores e melhores oportunidades de emprego e renda, tornando mais provável o ingresso no mercado de trabalho. Todavia, à medida que uma região cresce e se desenvolve transformações ocorrem em suas estruturas socioeconômicas, inclusive nos mercados de trabalho, tornando mais complexo o processo de inserção produtiva. Microrregiões mais desenvolvidas e urbanizadas tendem a apresentar maior percentual da força de trabalho com elevada qualificação, maior proporção de trabalhos formais e de ocupações intensivas em capital (serviços modernos). Esses fatores aumentam significativamente a competição por postos de trabalho, diminuindo, assim, a probabilidade de ocupação, especialmente de indivíduos pouco qualificados. Mais uma vez, a educação torna-se a principal medida para minimizar as desvantagens ocupacionais no Brasil.

Nesse contexto, é plausível afirmar que as diferenças entre as estruturas produtivas (micro)regionais têm influências significativas sobre a probabilidade de um indivíduo estar ocupado, desocupado ou inativo, pois a heterogeneidade na composição não produtiva da força de trabalho é insuficiente para explicar essa condição. O nível de desenvolvimento de uma região tem implicações diretas sobre as atividades que nela são desenvolvidas, pois quanto mais denso o centro urbano, mais diversificada é a interação entre os agentes econômicos e, conseqüentemente, mais facilmente são geradas inovações tecnológicas e sociais (Fujita & Thisse 2000).

A análise evidencia que os mercados de trabalho podem estar se tornando cada vez mais seletivos e excludentes, o que pode agravar a situação de alguns grupos populacionais, já que os mesmos não conseguem atender os pré-requisitos necessários para ocupar a maior parcela dos postos de trabalho que são gerados. É evidente que mesmo nos grandes centros urbanos continuam a existir atividades informais e que não necessitam de qualificação, o que contrabalança em alguma medida essa tendência principal. Todavia, parece haver uma incapacidade do setor urbano de criar empregos em número suficiente para absorver todo o contingente populacional que chega às cidades, atraído por suas facilidades e oportunidades.

Referências Bibliográficas

- Barros, R. P., Franco, S. & Mendonça, R. (2007), Discriminação e segmentação no mercado de trabalho e desigualdade de renda no Brasil, Texto para Discussão 1288, IPEA, Rio de Janeiro.

Cameron, A. & Trivedi, P. (2005), *Microeconometrics — methods and applications*, Cambridge University Press.

Duraton, G. & Puga, D. (2002), Diversity and specialization in cities: why, where and when does it matter?, in P. McCann, ed., 'Industrial location economics', Cheltenham/Northampton: Edward Elgar.

Fontes, G. G., Simões, R. F. & Oliveira, A. (2010), 'Urban attributes and wage disparities in Brazil: A multilevel hierarchical model', *Regional Studies* 44(5), 595–607.

Foster, J., Greer, J. & Thorbecke, E. (1984), 'A class of decomposable poverty measures', *Econometrica* 52(3), 761–766.

Fujita, M. & Thisse, J.-F. (2000), The formation of economic agglomerations: old problems and new perspectives, in J.-M. Huriot & J.-F. Thisse, eds, 'Economics of cities: theoretical perspectives', Cambridge: Cambridge University.

Glaeser, E. L. (1999), 'Learning in cities', *Journal of Urban Economics* 46(2), 254–277.

Henderson, J. V. (1988), *Urban development: theory, fact and illusion*, New York, Oxford: Oxford University.

Jacobs, J. (1969), *The economy of cities*, Middlesex: Penguin Books.

Lima, A., Oliveira, A. & Simões, R. (2011), Migração e inserção no mercado de trabalho: uma abordagem multinomial para a população economicamente ativa do Brasil, in 'XXXIX Encontro Nacional de Economia'.

Marshall, A. (1982), *Princípios de Economia*, Coleção os Economistas — Nova Cultural, São Paulo.

Marshall, A. & Wood, P. (1995), *Services and space: key aspects of urban and regional development*, Longman, London.

Ramos, L. (2007), O desempenho recente do mercado de trabalho brasileiro: tendências, fatos estilizados e padrões espaciais, Texto para Discussão 1255, IPEA, Rio de Janeiro.

Ramos, L. & Ferreira, V. (2005), 'Geração de empregos e realocação espacial no mercado de trabalho brasileiro — 1992-2002', *Pesquisa e Planejamento Econômico (PPE)* 35(1), 1–31.

Soares, S. S. D. (2000), O perfil da discriminação no mercado de trabalho — homens negros, mulheres brancas e mulheres negras, Texto para Discussão 769, IPEA, Brasília.

Valenzuela, M. E. (1999), Igualdade de oportunidades e discriminação de raça e gênero no mercado de trabalho no Brasil, in A. C. Posthuma, ed., 'Abertura e ajuste do mercado de trabalho no Brasil: políticas para conciliar os desafios de emprego e competitividade', São Paulo: Ed. 34.

DESIGUALDADE DE OPORTUNIDADES NA FORMAÇÃO DE HABILIDADES: UMA ANÁLISE COM DADOS LONGITUDINAIS

IGOR VIEIRA PROCÓPIO [†]
RICARDO DA SILVA FREGUGLIA [‡]
FLÁVIA CHEIN [§]

Resumo

Este artigo analisa a evolução da desigualdade educacional e das oportunidades ao longo dos quatro primeiros anos do ensino fundamental por intermédio do desempenho nas disciplinas de língua portuguesa e matemática daquelas crianças que participaram do GERES no período de 2005 a 2008. Os principais resultados obtidos mostram um aumento na desigualdade educacional em matemática e um comportamento praticamente estável em português. No que tange à desigualdade de oportunidades, no fim do período analisado, os valores são de 8% para matemática e 11% para português. Percebe-se um aumento nesses valores ao longo do período, sendo que português apresenta um crescimento superior ao de matemática, 109% e 80%, respectivamente. A característica que mais influencia na desigualdade de aprendizado é o nível socioeconômico.

Palavras-chave: Desigualdade de Oportunidades; Formação de Habilidades.

Abstract

The aim of this paper is to analyze the evolution of education inequality and the inequality of opportunities over the first four years of elementary school in Brazil from 2005 to 2008. Using proficiency performance information of pupils in mathematics and language (Portuguese) from GERES database. The main results show an increase in educational inequality in mathematics and stable behavior in Portuguese over the years under review. With regard to inequality of opportunity, the percentages

* Agradecemos o apoio financeiro da FAPEMIG, CNPq e Observatório da Educação 2010/INEP/Capes

[†] Universidade Federal de Juiz de Fora. Juiz de Fora, Minas Gerais, Brasil. E-mail: igor.procopio@ufjf.edu.br

[‡] Professor adjunto Faculdade de Economia e do Programa de pós-Graduação em Economia da Universidade Federal de Juiz de Fora (UFJF), pesquisador da FAPEMIG e bolsista de Produtividade em Pesquisa do CNPq. Juiz de Fora, Minas Gerais, Brasil. E-mail: ricardo.freguglia@ufjf.edu.br

[§] Professora adjunto Faculdade de Economia e do Programa de pós-Graduação em Economia da Universidade Federal de Juiz de Fora (UFJF), pesquisadora da FAPEMIG e bolsista de Produtividade em Pesquisa do CNPq. Juiz de Fora, Minas Gerais, Brasil. E-mail: flavia.chein@ufjf.edu.br

are 8% for math and 11% for Portuguese. The Portuguese inequality presents a higher increase in these percentages (109%) than the math inequality (80%) over the four years. The feature that most influences the inequality in proficiency performance is the socioeconomic status of the children family.

Keywords: Inequality of Opportunity; Skill Formation.

JEL classification: D31, I21, I24.

DOI: <http://dx.doi.org/10.1590/1413-8050/ea110460>

1 Introdução

A teoria de Igualdade de Oportunidades, sintetizada no livro de Roemer (1998), vem ganhando espaço nas discussões sobre justiça distributiva e na formulação de políticas públicas voltadas para a redução das desigualdades. De forma sucinta, essa teoria divide os determinantes dos resultados individuais em dois grupos. Um grupo considerado injusto, que são fatores fora do controle do indivíduo, como gênero, raça, origem familiar, local de nascimento, e outro grupo de fatores considerados justos, que engloba os fatores que estão sob controle do indivíduo, como as escolhas autônomas e o nível de esforço. De acordo com Roemer (1998), as desigualdades que são geradas pelos fatores injustos deveriam ser eliminadas, enquanto que as desigualdades de cunho justo deveriam ser mantidas, pois servem de incentivos para o esforço individual.

O Brasil apresenta uma das maiores desigualdades de renda do mundo, fenômeno vastamente documentado na literatura. Desde os anos de 1960, quando se iniciou o levantamento sistemático de informações censitárias no país, os índices de desigualdade encontrados se situam entre os mais altos do mundo, com valores do índice de Gini partindo de 0,535 em 1960, chegando ao ápice em 1990, com 0,607 e reduzindo a 0,527 em 2011¹. Apesar da queda nesses índices na última década, o Brasil continua figurando entre os países com maior desigualdade de renda². Essa desigualdade é, em grande medida, relacionada à origem familiar das pessoas, o que indica uma desigualdade de oportunidades.

A evidente associação entre rendimentos e origem familiar faz com que o uso do arcabouço teórico da igualdade de oportunidades seja adequado para a análise das desigualdades no Brasil. O ideal dessa teoria é um mundo em que as características predeterminadas das pessoas não tenham influência em seus resultados. É uma teoria que se encontra no meio do caminho entre as teorias liberais, em que toda a responsabilidade do sucesso individual é atribuída ao próprio indivíduo e, portanto, não deve haver políticas de redistribuição de renda, e teorias de cunho redistributivista, que consideram a sociedade a responsável pelas conquistas individuais, advogando em causa da redistribuição.

Além dos argumentos normativos em favor da igualdade de oportunidades, pode-se argumentar que políticas públicas voltadas para equalizar as oportunidades individuais terão impactos em termos de eficiência econômica. No campo microeconômico, o fato de a teoria da igualdade de oportunidades responsabilizar os indivíduos por parte de seus resultados mantém os incentivos para o esforço individual. Quando o resultado individual é determinado em parte por suas circunstâncias predeterminadas o indivíduo sabe que mesmo se exercer um alto nível de esforço, a probabilidade de ele alcançar o topo da distribuição de renda é baixa, o que pode levar esse indivíduo a não se esforçar o tanto que ele se esforçaria se soubesse que o principal determinante de sua renda seria seu esforço.

Já no campo macroeconômico, diversos autores, como Bourguignon (1981), Barro (2000), Rebelo (1991), Deininger & Squire (1998), Galor & Moav (2004), tentam relacionar desigualdade de renda com crescimento econômico, con-

¹Ver, entre outros, Fishlow (1972), Langoni (2005), Bonelli & Sedlacek (1988), Hoffman (1989), Lam & Levison (1990), Ramos (1993), Barros & Mendonça (1995), Deininger & Squire (1996), Ferreira & Litchfield (1996, 1999), Ferreira (2000).

²Ver Barros et al. (2006) sobre a queda recente da desigualdade de renda brasileira.

tudo, não conseguem definir precisamente a influência da desigualdade no crescimento. Um recente estudo de Marrero & Rodríguez (2013) traz uma nova luz à discussão. Os autores argumentam que o efeito da desigualdade no crescimento varia em função de sua natureza. Se a desigualdade é em função de diferentes níveis de esforços empreendidos pelos indivíduos, ela exerce um efeito positivo no crescimento. Entretanto, quando é relacionada às características pré-determinadas dos indivíduos, exerce um efeito negativo no crescimento. Portanto, quanto maior a desigualdade de oportunidades em um país, maiores serão os entraves ao crescimento. A parcela da desigualdade resultante de circunstâncias fora do controle do indivíduo leva a uma acumulação de capital humano subótima e conseqüente baixo crescimento, ao passo que a parcela em função de diferentes escolhas e nível de esforço incentivam o investimento do indivíduo em capital humano (Ferreira et al. 2005).

Todo o arcabouço teórico da igualdade de oportunidades, comumente aplicado à desigualdade de renda, pode ser aplicado à desigualdade educacional. O estudo das desigualdades educacionais pode ser justificado pela importância da educação por si própria, como promotora do desenvolvimento humano e das capacidades individuais (Sen, 2000), mas também em função de sua relação com a desigualdade de renda. No Brasil, a partir do trabalho de Langoni (2005), alguns estudos começaram a apresentar evidências de que o principal determinante da alta desigualdade é a disparidade educacional de sua população (Barros & Mendonça 1995, Ferreira 2000). Esses estudos apontam evidências de que a desigualdade de salários apresentada no mercado de trabalho é um reflexo da diferença de produtividade dos trabalhadores, e que essa diferença não é gerada no mercado de trabalho. Barros & Mendonça (1995) desenvolvem um modelo dividindo o ciclo de vida do trabalhador em dois períodos, um período pré-mercado de trabalho, em que as habilidades são formadas, e um segundo período correspondendo ao mercado de trabalho, no qual as habilidades são transformadas em produtividade e remuneradas pelos salários. Os autores concluem que a maior parcela da desigualdade é gerada no primeiro período do ciclo de vida, ou seja, na fase de formação das habilidades. Portanto, os trabalhadores entram no mercado de trabalho em desigualdade de condições.

Do mesmo modo que a desigualdade de renda no Brasil é fortemente relacionada com a origem social dos indivíduos, as desigualdades educacionais também o são. Diversos fatores contribuem para a formação de habilidades. Alguns desses fatores estão sob o controle dos indivíduos, como motivação e esforço, enquanto outros não podem ser considerados de responsabilidade da criança, como local de residência e origem familiar. Evidências apontam que grande parte da desigualdade de desempenho dos alunos se deve às desigualdades de nível socioeconômico dos alunos. Além da questão socioeconômica, os insumos escolares também exercem influência nas desigualdades educacionais (Albernaz et al. 2002). Em um ambiente com elevada desigualdade de oportunidades educacionais, maiores investimentos podem ser realizados em pessoas com melhor *background* social em detrimento de pessoas com maior talento inato, podendo ocasionar perda de eficiência no sistema educacional.

Recentes estudos vêm aplicando o arcabouço teórico da igualdade de oportunidades para o campo da educação, entre os quais estão Albernaz et al. (2002), Waltenberg & Vandenberghe (2007), Gamboa & Waltenberg (2012), Ferreira & Gignoux (2014), Diaz (2012). Esses estudos encontram evidências de que parte das desigualdades educacionais é determinada por característi-

cas relacionadas com a origem familiar. No entanto, tem-se apenas um quadro de momento, pois utilizam dados em *cross-section*. Não é possível analisar o comportamento da desigualdade e nem da influência das circunstâncias ao longo do período escolar.

O objetivo deste artigo é analisar o comportamento da desigualdade educacional ao longo do período escolar, bem como a desigualdade de oportunidades por intermédio do desempenho nas disciplinas de língua portuguesa e matemática daquelas crianças que participaram do GERES (Projeto Geração Escolar), no qual foi acompanhada uma amostra de crianças durante os quatro primeiros anos do ensino fundamental. Essa base permite avaliar o grau de desigualdade com que as crianças iniciam o período escolar obrigatório bem como sua evolução nos primeiros anos. Dessa maneira, é possível verificar o quanto da desigualdade educacional é gerada antes da entrada na escola e qual o papel da escola nesse processo, ampliação ou redução.

O artigo está estruturado em mais cinco seções além desta introdução. A próxima seção apresenta a teoria da igualdade de oportunidades e sua aplicação no campo educacional. A seção 3 traz uma descrição do banco de dados utilizado. A terceira seção apresenta a estratégia empírica, a quarta as estatísticas descritivas e resultados e a quinta seção apresenta as considerações finais.

2 Igualdade de Oportunidades Educacionais

O debate sobre a responsabilidade individual nos resultados econômicos vem ganhando destaque nas discussões de políticas públicas voltadas para a redistribuição de renda. A *Theory of Justice* (Rawls 1971) pode ser considerado o primeiro trabalho sobre justiça distributiva que separa os determinantes dos resultados em fatores de responsabilidade do indivíduo e fatores fora do controle do indivíduo. Seguindo Rawls, destacam-se os trabalhos de Dworkin (1981a,b), Arneson (1989, 1990) e Cohen (1989). Entre os economistas, Sen (1979), em seu trabalho *Equality of What?*, parte do pressuposto que toda teoria normativa sobre ordenamento social associa justiça à igualdade de algo. O que diferencia essas teorias é o que se deve igualar. Sen defende que o essencial é a formação das capacidades do indivíduo e que a atenção das políticas públicas deve ser na distribuição dessas capacidades. Roemer (1998) apresenta uma definição formal do conceito de igualdade de oportunidades, estabelecendo de maneira abstrata a distinção entre os fatores sob responsabilidade do indivíduo e as circunstâncias fora do seu controle.

Roemer (1998) formulou seu modelo teórico de forma bastante flexível, tanto na definição dos fatores injustos (que na linguagem da teoria são chamados de circunstâncias) quanto na definição de qual resultado deve ser objeto de equalização das oportunidades. Os primeiros artigos empíricos, com destaque para os estudos de Betts & Roemer (2005), Roemer et al. (2003), Checchi & Peragine (2010), focaram a análise na desigualdade de renda.

Apesar de a flexibilidade do modelo desenvolvido em Roemer (1998) permitir a análise de outros resultados individuais além dos rendimentos, a aplicação ao campo educacional requer alguns cuidados. Dois pontos merecem atenção, um relacionado à definição de fatores sob responsabilidade do indivíduo e fatores fora do controle individual, e outro ponto referente à estratégia empírica para a mensuração das desigualdades.

O principal problema na definição de fatores justos e injustos na aplicação da teoria de igualdade de oportunidades ao campo da educação se refere à idade dos indivíduos. Enquanto que na análise da desigualdade de renda os indivíduos são todos adultos, na avaliação das desigualdades educacionais os indivíduos podem ser adolescentes e até mesmo crianças. A atribuição de responsabilidades a crianças e adolescentes é mais delicada e controversa. Em que medida se pode atribuir a esses indivíduos a responsabilidade por suas escolhas, em especial o nível de esforço e dedicação nos estudos? Será que essa responsabilidade deve ser atribuída à criança e ao adolescente ou aos seus pais e professores? Caso não se possa atribuir responsabilidade a estes indivíduos, toda a desigualdade seria injusta e não haveria a necessidade de decompor a desigualdade em fatores justos e injustos.

Essa discussão é levantada nos artigos de Waltenberg & Vandenberghe (2007), Peragine & Serlenga (2007), Gamboa & Waltenberg (2012). Peragine & Serlenga (2007) preferem não aprofundar no problema e focam a análise na educação superior, em que os indivíduos já são adultos e, portanto passíveis de serem reponsabilizados por suas decisões. Waltenberg & Vandenberghe (2007) e Gamboa & Waltenberg (2012) analisam a desigualdade educacional para adolescentes entre 14 e 15 anos e argumentam que, apesar de não serem plenamente capazes de serem responsabilizados por suas decisões, já são pelo menos parcialmente capazes. Os autores ressaltam que em vários países adolescentes de 16 anos já podem votar e dirigir.

No caso do presente estudo essa questão é mais delicada, pois avaliamos a desigualdade para crianças que têm 6 ou 7 anos no início do período analisado e 10 ou 11 ao final do período. Apesar da dificuldade de se atribuir responsabilidade às escolhas que crianças nessa faixa etária realizam, o arcabouço teórico da igualdade de oportunidades ainda pode ser utilizado. Dentre os principais determinantes do sucesso educacional estão fatores que não são passíveis de interferência de políticas públicas, como habilidades inatas e até mesmo sorte. Com isso, na fração da desigualdade chamada de esforço na linguagem da teoria da igualdade de oportunidades, estão fatores genéticos, sorte e outros fatores não disponíveis no banco de dados, como preferência dos pais pela educação, disponibilidades de recursos didáticos em casa, entre outros. Todo esse grupo de fatores pode ser interpretado como sendo a desigualdade residual. Portanto, a fração atribuída à desigualdade de oportunidades neste estudo é inferior à verdadeira desigualdade de oportunidades.

Quanto às limitações em relação à estratégia empírica, a análise da desigualdade educacional pode ser realizada em termos de anos de estudo ou distorção idade-série. No entanto, não são critérios adequados, pois não incorporam a qualidade do ensino. Uma pessoa com 10 anos de estudo em um país ou município com um bom sistema educacional não pode ser diretamente comparada com uma pessoa com os mesmos 10 anos de estudo em outra localidade com um sistema de ensino inferior. Em função disso, as avaliações educacionais estão sendo realizadas através dos resultados de testes padronizados aplicados em larga escala. Os resultados desses testes podem ser diretamente comparados entre pessoas de diferentes localidades e ao longo do tempo.

A análise das desigualdades com o uso de testes padronizados não pode ser realizada com as mesmas técnicas utilizadas para a análise da desigualdade de renda. Os resultados dos testes são divulgados não em suas notas absolutas, mas com o uso da Teoria da Resposta ao Item, que dentre algumas peculiaridades, realiza uma normalização das notas. Essa característica faz com que as

medidas usuais de desigualdade, como índice de Gini e Theil, sejam inapropriadas para aplicação na avaliação das desigualdades educacionais (Ferreira & Gignoux 2014). A seção 4 traz esta discussão em maiores detalhes.

A escolha das circunstâncias a serem consideradas na análise envolve sempre certo grau de arbitrariedade e limitação em função das informações disponíveis. No caso de crianças, tal definição é ainda mais complicada, como mencionado anteriormente. Se considerarmos que uma criança de 7 anos não pode ser responsabilizada por suas escolhas, todos os fatores que determinam o aprendizado seriam considerados circunstâncias. Os principais determinantes para o aprendizado das crianças são os fatores genéticos, as características familiares e características escolares. Neste estudo, utilizamos como circunstâncias o nível socioeconômico da família, a raça e o gênero da criança e o município da escola.

O nível socioeconômico das famílias é apontado por diversos estudos como o principal fator determinante do sucesso educacional. Desde o relatório Coleman (1966) que a importância do nível socioeconômico é destacada. O nível educacional dos pais influencia no aprendizado das crianças, tanto por transferir uma maior preferência pela educação quanto pela maior capacidade de auxiliar as crianças nas tarefas educacionais levadas para casa. A riqueza da família permite o acesso a melhores escolas e a um maior número de bens que contribuem para o aprendizado, como livros, revistas, computador e acesso à internet. Utilizamos uma variável composta que engloba ao mesmo tempo informações sobre o nível de escolaridade da mãe, a ocupação do pai e a presença de certos bens materiais na casa da criança. A escolha dessa variável em detrimento de se utilizar cada informação separadamente se deu em função de limitações do banco de dados. A próxima seção contém uma explicação da forma de construção desta variável.

Outra característica que aparece como um forte determinante das desigualdades, tanto de renda quanto educacional, é a raça do indivíduo. O Brasil, historicamente, apresenta uma forte desigualdade racial que é facilmente notada no mercado de trabalho. Em média, a população branca possui salários superiores aos recebidos pela população negra. Um dos fatores explicativos desta desigualdade é a discriminação que sofrem os negros em relação aos brancos. Outra explicação reside na menor escolaridade da população negra, tornando-os menos produtivos. Portanto, se faz necessário entender em que momento se inicia a desigualdade entre as raças. Soares et al. (2005) trazem diversos textos investigando a questão da desigualdade educacional entre as raças. Existe uma forte correlação entre raça e nível socioeconômico, portanto, parte da desigualdade educacional entre raças pode ser explicada pela circunstância anterior. No entanto, estudos mostram que, mesmo após controlar pelo nível socioeconômico das famílias, as crianças negras ainda possuem um rendimento inferior em comparação com crianças brancas (Soares et al. 2005, Soares & Alves 2003). Obviamente, a raça da criança não é uma característica de escolha por ela, ou seja, é claramente um problema de oportunidades.

O gênero é um fator importante para a análise da desigualdade de renda. Estudar a desigualdade educacional pode contribuir para um maior entendimento de diferenciais de gênero no mercado de trabalho (Pinheiro et al. 2009). Está bem estabelecido na literatura que as meninas possuem melhor rendimento em português, enquanto que os meninos possuem melhor rendimento em matemática. Neste estudo, é possível avaliar como se dá a evolução desse *gap*, bem como a sua contribuição para a desigualdade educacional.

Por fim, o município também pode ser um fator de desigualdade educacional, pois representa, em síntese, a qualidade da rede escolar. Riani & Rios-Neto (2008) apontam que uma boa rede escolar municipal pode reduzir a influência do nível socioeconômico familiar. É importante ressaltar, que neste estudo, apenas quatro municípios compõem a amostra e, com isso, a análise desse fator fica prejudicada. Entretanto, mesmo diante de tal limitação, os resultados considerando essa circunstância serão apresentados, pois não existem estudos mostrando a evolução da contribuição dos municípios para a desigualdade educacional. Mas também serão apresentados resultados apenas com as outras três circunstâncias.

3 Base de Dados

A base de dados utilizada neste estudo é proveniente do Projeto Geres, que avalia alunos de uma amostra de escolas durante os quatro primeiros anos do Ensino Fundamental. Portanto, é um estudo longitudinal no qual é possível acompanhar a evolução do nível de habilidades das crianças por meio do nível de proficiência, ou seja, é possível medir a evolução do aprendizado da criança. Ao acompanhar uma amostra de alunos durante a primeira fase do Ensino Fundamental, a análise dos resultados do Geres é uma contribuição para as pesquisas de avaliação educacional no Brasil. Os objetivos do Geres são:

- identificar as características escolares que maximizam a aprendizagem dos alunos e que minimizam o impacto da origem social sobre o aprendizado;
- identificar os fatores escolares que diminuem a probabilidade de repetência dos alunos;
- identificar aquelas características da escola que reduzem a probabilidade do absentismo.

O Geres acompanhou alunos entre 2005 e 2008, realizando testes todos os anos, possibilitando estimar os níveis de proficiência em matemática e português. Em 2005, foram aplicados dois testes, um no início do ano, para ser utilizado como diagnóstico, e outro ao final do ano. Ao todo, foram cinco aplicações de testes, denominadas ondas. A onda 1 representa as informações de diagnóstico, a onda 2 as informações relativas à segunda aplicação dos testes, e assim por diante. Foram acompanhados alunos que em 2005 estavam matriculados na primeira série do Ensino Fundamental (ou no segundo ano, onde o Ensino Fundamental já era composto por nove anos). Além da aplicação dos testes, o projeto também aplicou questionários aos professores, diretores, pais e alunos com o objetivo de avaliar o efeito de fatores familiares e escolares no aprendizado. As escolas da amostra estão localizadas em cinco grandes cidades brasileiras — Belo Horizonte, Rio de Janeiro, Campo Grande, Salvador e Campinas, e seis universidades participaram do planejamento, coordenação e execução do projeto — UFMG, PUC-Rio, UEMS, UFBA UNICAMP e UFJF.

As avaliações em larga escala mais difundidas no país, SAEB e Prova Brasil, aplicam testes em crianças a partir da 4ª série/5º ano do ensino fundamental. O Geres inicia a aplicação dos testes para crianças na 1ª série/2º ano, ou seja, crianças mais novas. Esta característica exige alguns procedimentos diferenciados. Primeiro na forma de apresentação do teste para as crianças. Os itens

não foram apresentados de forma aleatória, como acontece no SAEB. Os itens foram apresentados de acordo com seu nível de dificuldade, do mais fácil para o mais difícil, método considerado mais adequado para as séries inferiores (Soares & Oliveira 2009). Além dessa ordenação, nas primeiras aplicações do teste, parte dos enunciados foi lida para as crianças.

O segundo procedimento diferenciado se refere à escala de proficiências. Optou-se por uma escala própria de proficiência e não pela escala do SAEB, cujo o grupo de referência é a 8ª série. De acordo com Soares & Oliveira (2009), utilizar a 8ª série como grupo de referência permite boa comparabilidade para as séries avaliadas no SAEB, mas não produz uma medida adequada para o Geres, dado que as crianças da 1ª série estão muito distantes dos alunos da 8ª série em termos de habilidades cognitivas. Entretanto, os itens aplicados na 4ª série do SAEB 2003 foram calibrados na escala Geres, permitindo que as proficiências dos alunos participantes do Geres possam ser comparadas com as proficiências do SAEB (Soares & Oliveira 2009).

A proficiência dos alunos foi estimada com a utilização da Teoria da Resposta ao Item (TRI), método amplamente utilizado nos programas de avaliação educacional (SAEB, Prova Brasil, SARESP, PISA). A TRI permite comparabilidade entre os resultados de testes aplicados em diferentes momentos³. A equação 1 apresenta a fórmula para estimação das proficiências.

$$P(Y_{ig} = 1 | a_i, b_i, c_i, \theta_j) = c_i + \frac{(1 - c_i)}{1 + e^{-Da_i(\theta_j - b_i)}}, \theta_j \sim N(\mu_g, \sigma_g), \quad (1)$$

onde, $P(Y_{ig} = 1 | \cdot)$ representa a probabilidade da resposta Y_{ig} do aluno j do grupo g estar correta dada a proficiência θ_j , a_i representa o parâmetro de discriminação, b_i o parâmetro de dificuldade e c_i o parâmetro de acerto casual (Soares & Oliveira 2009).

Após a estimação das proficiências pela fórmula apresentada na equação 1, é realizada uma padronização das notas. A padronização é uma transformação linear das notas originais para uma escala arbitrária, em que existe um movimento de rotação e de translação. Como será visto na seção metodológica, essa transformação impede o uso dos índices de desigualdade comumente utilizados na análise da desigualdade de renda, gini e theil, por exemplo, para a análise da desigualdade educacional. A padronização segue a fórmula apresentada na equação 2.

$$y_i = \widehat{\mu} + \frac{\widehat{\sigma}}{\sigma}(x_i - \mu), \quad (2)$$

onde, y_i é nota do aluno i após a padronização, $\widehat{\mu}$ e $\widehat{\sigma}$ representam a média e o desvio padrão arbitrários da distribuição padronizada, respectivamente, μ e σ representam a média e o desvio padrão da distribuição original, respectivamente, e x_i é a nota pré-padronização obtida pela aplicação da TRI, conforme equação 1. No caso do Geres, a média e o desvio padrão arbitrário foram definidos em 100 e 25, respectivamente (Oliveira 2008).

Neste trabalho, além dos níveis de proficiência dos alunos em português e matemática, são utilizadas a raça das crianças, o nível socioeconômico da família e o município de residência, variáveis retiradas dos questionários de pais e alunos para caracterizar a origem familiar.

³Para maiores detalhes da TRI ver, entre outros, (Baker & Kim 2004)

O nível socioeconômico é uma variável disponibilizada juntamente com o banco de dados do projeto Geres. É uma variável composta de diversas informações coletadas nos questionários aplicados aos pais e aos próprios alunos. São três dimensões compondo o índice: escolaridade da mãe, ocupação do pai e renda. Esta última é composta pela presença de bens de conforto no domicílio, como existência de computador, carro, aparelho de dvd, máquina de lavar, etc ⁴.

A tabela 1 apresenta a frequência e percentual das variáveis de raça, gênero e município ⁵. Tanto no banco de português quanto no banco de matemática, há uma concentração de crianças pardas, seguidas pelas brancas, negras, indígenas e amarelas, nesta ordem. O gênero está bem equilibrado, com um pouco mais de 50% de meninos nos dois casos. Em relação aos municípios, Belo Horizonte e Campinas possuem em torno de 30% cada, Rio de Janeiro 22% e Campo Grande 16%. O total de observações é de 5.374 crianças por onda em matemática e 5.281 em português.

Tabela 1: Quantidade e percentual de crianças por raça, gênero e município de residência

Circunstâncias	Matemática		Português	
	Quantidade	%	Quantidade	%
Raça				
Branco	1761	32,77	1731	32,78
Pardo	2791	51,94	2735	51,79
Negro	822	15,3	815	15,43
Gênero				
Meninos	2696	50,17	2662	50,41
Meninas	2678	49,83	2619	49,59
Município				
Belo Horizonte	1743	32,43	1694	32,08
Campinas	1822	33,9	1811	34,29
Campo Grande	943	17,55	913	17,29
Rio de Janeiro	866	16,11	863	16,34
Total	5374	100	5281	100

Fonte: GERES.

A tabela 2 mostra a média do nível socioeconômico dos alunos e da proficiência em cada onda de aplicação dos testes. Tanto considerando os alunos que fizeram todos os testes de português como aqueles que fizeram todos os testes de matemática, o nível socioeconômico dos alunos apresenta a mesma média, com uma pequena diferença no desvio padrão. Nota-se uma evolução das proficiências ao longo das ondas, que representa o aprendizado das crianças após o início do ensino fundamental.

⁴Para mais detalhes da construção desta variável, ver Soares & Oliveira (2009).

⁵Esses valores são os mesmos para todas as ondas, pois são características invariantes no tempo

Tabela 2: Média do NSE e Proficiência

	Matemática	Português
NSE	-0,13	-0,13
	Proficiência	
Onda1	101,37	100,60
Onda2	129,84	121,15
Onda3	147,64	136,99
Onda4	189,63	152,47
Onda5	233,23	164,87

Fonte: GERES.

4 Estratégia Empírica

Desde Roemer (1998), diversas metodologias vêm sendo aplicadas para mensurar a desigualdade de oportunidades. Os primeiros trabalhos empíricos buscavam avaliar a desigualdade de oportunidades em relação aos salários e rendimentos do indivíduo. Os trabalhos de Pignataro (2012) e Ramos (2012) trazem uma revisão dos principais métodos utilizados. As principais técnicas utilizadas se baseiam em decomposição de índices de desigualdades, principalmente do Índice de Theil.

No campo das desigualdades educacionais, as mesmas técnicas utilizadas para a desigualdade de renda foram aplicadas (Checchi & Peragine 2005, Gamboa & Waltenberg 2012). No entanto, Ferreira & Gignoux (2014) alertam para o problema de usar tais técnicas na análise da proficiência dos alunos quando estas são disponibilizadas em valores padronizados.

Ferreira & Gignoux (2014) apresentam as propriedades desejáveis que os índices de desigualdade devem atender.

- simetria: requer que o índice seja insensível a permutações no vetor de interesse;
- continuidade nos valores observados;
- princípio da transferência: requer que o índice seja sensível a uma transferência de alguém no topo da distribuição para alguém na parte inferior da distribuição e vice-versa;

Além dessas três propriedades, mais uma, entre duas outras, é desejável, porém nenhum índice de desigualdade atende as duas simultaneamente.

- invariância de escala: requer que o índice seja insensível à mudança de escala, $I(y) = I(\lambda y)$, $\lambda > 0$ e y o vetor de interesse, ou;
- invariância em relação à translação: requer que o índice seja insensível à uma translação do vetor de interesse, $I(y) = I(y + a)$, $a \neq 0$.

Ferreira & Gignoux (2014) destacam que nenhum índice de desigualdade produz uma medida cardinalmente igual para uma distribuição pré e pós padronização dos resultados de proficiência. Os autores ainda chamam a atenção que esse problema ocorre em função da padronização utilizada nos resultados dos testes, e não em função da aplicação da Teoria da Resposta ao Item. Portanto, esse problema ocorre no GERES da mesma maneira que ocorre em

outros casos de aplicações de teste, SAEB, Prova Brasil, PISA, entre outros, mesmo o GERES utilizando um número maior de parâmetros na estimação da proficiência.

Entretanto, não produzir uma medida cardinalmente invariante à padronização não é um problema muito grave. O que os autores chamam a atenção é que nenhum índice de desigualdade conhecido é ordinalmente equivalente quando aplicado a distribuições pré e pós padronizadas. Ou seja, ao se comparar a desigualdade entre duas ou mais distribuições, nada garante que o ranking entre essas distribuições se manterá quando o índice for aplicado à distribuição pré padronização e à distribuição pós padronização.

Apesar de não fazermos comparações entre países, como Gamboa & Waltenberg (2012) ou regiões, neste trabalho fazemos comparações entre distribuições ao longo do tempo. Aplicando os conhecidos índices de desigualdade nessas distribuições, já padronizadas, não poderemos ter confiança se a evolução observada nos índices reflete a verdadeira evolução na desigualdade de proficiência. Como solução, os autores sugerem a aplicação de uma medida simples, a variância. Gamboa & Waltenberg (2012) mostram que a variância aplicada à distribuição pós padronizada é apenas uma transformação monotônica da variância aplicada à distribuição pré padronização, o que permite uma equivalência ordinal entre as duas distribuições. Portanto, neste trabalho utilizamos a variância aplicada à distribuição da proficiência em matemática e português para as cinco ondas, permitindo acompanhar a evolução da desigualdade de proficiência nos primeiros anos do Ensino Fundamental para uma amostra de crianças.

Após calcular a evolução da desigualdade de proficiência, o nosso interesse recai sobre o quanto dessa desigualdade pode ser explicada por características do indivíduo fora de seu controle. As técnicas utilizadas se baseiam em decomposições da desigualdade. A desigualdade de oportunidades é medida como a relação entre a desigualdade entre os grupos definidos e a desigualdade geral. Seguimos o método usado por Ferreira & Gignoux (2014), que sugerem uma alternativa paramétrica baseada numa regressão estimada por Mínimos Quadrados Ordinário (MQO), com a proficiência sendo explicada pelas características não controladas pelo indivíduo. O índice de desigualdade de oportunidades será a razão entre a variância explicada pela regressão e pela variância total da proficiência, que é simplesmente o R quadrado da regressão. É importante destacar que, nesse tipo de construção da desigualdade de oportunidades, a regressão utilizada não pretende estimar relações causais, e o valor do R quadrado, irá fornecer uma medida subestimada da desigualdade de oportunidades.

Para um melhor entendimento do comportamento da desigualdade de oportunidades, os resultados são apresentados com a inclusão gradativa das circunstâncias, até o resultado final onde as quatro circunstâncias escolhidas são consideradas conjuntamente. A medida é aplicada para as cinco ondas do GERES permitindo uma avaliação da evolução da desigualdade de oportunidades durante os primeiros anos do período escolar.

Ferreira & Gignoux (2014) destacam algumas vantagens de se utilizar o R quadrado como medida de desigualdade de oportunidades. Entre elas está a possibilidade de decomposição da medida nos componentes individuais utilizados na regressão. Com isso é possível calcular a contribuição marginal de cada circunstância na desigualdade de oportunidades e conseqüentemente a evolução ao longo do tempo destas características. Os autores sugerem que o

R2 seja decomposto com o uso da decomposição de Shapley-Shorrocks.

5 Estatísticas Descritivas e Resultados

Esta seção apresenta primeiramente gráficos mostrando a evolução da proficiência para os grupos dentro de cada circunstância, ou seja, como se dá a evolução da proficiência comparativamente entre homens e mulheres, entre as diferentes raças dos alunos, entre os diferentes municípios e entre os quintis de nível socioeconômico. Em seguida, são apresentados os resultados para a evolução da desigualdade de proficiência, utilizando a variância como medida de desigualdade, e por fim os resultados dos exercícios econométricos, em que é possível visualizar a evolução da desigualdade de oportunidades e da contribuição de cada circunstância.

Antes de mostrarmos a evolução da proficiência para cada circunstância, destacamos a evolução da proficiência média de cada disciplina, mostrada no gráfico 1. Como se pode perceber, as duas proficiências começam com valores próximos, mas vão se distanciando ao longo das ondas. A proficiência em matemática cresce a taxas crescentes, enquanto que a proficiência em português cresce a taxas decrescentes. Esse fato também foi destacado em Brooke et al. (2014), que utilizaram os mesmos dados do que os utilizados neste artigo. Os autores ressaltam que a diferença entre os padrões das duas disciplinas é algo conhecido na literatura da educação. O aprendizado de português se dá antes mesmo da entrada da criança na escola, devido ao uso da língua na convivência social. Já o aprendizado de matemática é realizado fundamentalmente dentro do ambiente escolar, portanto, sua evolução é mais pronunciada após o início de período escolar. É possível notar que o crescimento do aprendizado em matemática se torna mais expressivo entre as 3^a e 4^a ondas, que representam as 2^a e 3^a séries. Esse aumento da inclinação a partir da 3^a série é compatível com o processo de ensino das duas disciplinas. Os dois primeiros anos do ensino fundamental (ou os 3 primeiros, quando considerado o ensino fundamental de 9 anos) é voltado mais para o processo de alfabetização, sendo que a partir da 3^a série aumenta a ênfase no ensino em matemática. Essa diferença entre o comportamento das duas disciplinas é importante para a análise da evolução das desigualdades de oportunidades e da contribuição de cada circunstância.

A figura 2 apresenta a evolução das proficiências em matemática e português, respectivamente, para cada raça conforme declarada nos questionários do Geres. Tanto para a proficiência em português quanto em matemática, as curvas para brancos e pardos evoluem de forma semelhante, mantendo-se a distância entre elas praticamente constante ao longo dos anos. Já a curva de proficiência para negros, além de já começar o período com uma distância maior em relação às outras duas curvas, aumenta esta distância ao longo do período analisado, indicando que a escola não está sendo capaz de reduzir a desigualdade entre as raças, e de forma mais preocupante, está contribuindo para um aumento destas desigualdades. O resultado reforça outras evidências já documentadas na literatura, como Soares et al. (2005); Soares & Alves (2003).

A figura 3 apresenta a evolução dos níveis de proficiência para cada quintil de nível socioeconômico. Da mesma forma como acontece com a raça, o formato das curvas é semelhante para cada quintil e a distância entre as curvas

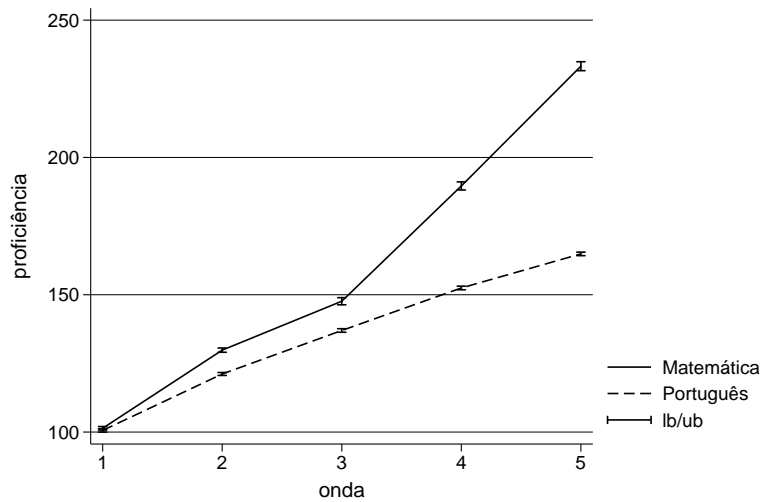


Figura 1: Evolução da proficiência por disciplina

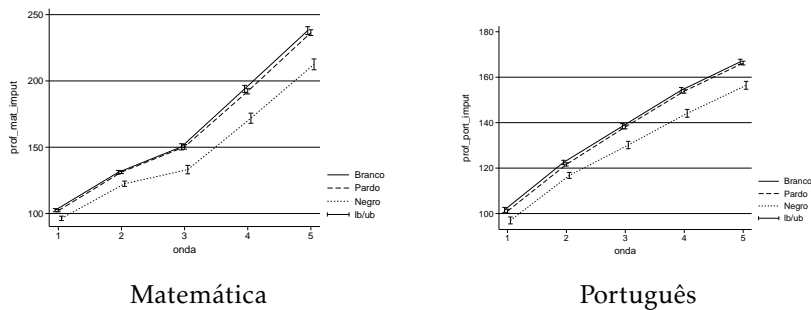


Figura 2: Evolução da proficiência por raça

na proficiência em matemática vai se ampliando, ou seja, a desigualdade entre os grupos socioeconômicos se amplia ao longo dos primeiros anos do Ensino Fundamental.

A figura 4 mostra a evolução da proficiência média por município em matemática e português, respectivamente. Nota-se uma evolução dos níveis de proficiência para todos os municípios e que as linhas não se cruzam, ou seja, os municípios mantêm suas posições relativas iniciais. Mas a distância entre as linhas se altera. Na proficiência em português, a distância entre os municípios se reduz enquanto em matemática a distância se altera menos. Destaque para os municípios de Belo Horizonte e Campo Grande, que iniciam praticamente com o mesmo nível de proficiência, no entanto, na última onda, esta distância é visualmente perceptível, ou seja, as escolas de Belo Horizonte apresentam um maior nível de aprendizado do que as escolas de Campo Grande, este efeito é mais acentuado na proficiência em matemática. Este mesmo efeito ocorre entre os municípios do Rio de Janeiro e de Campinas. Campinas inicia o período com média de proficiência inferior à média apresentada pelo

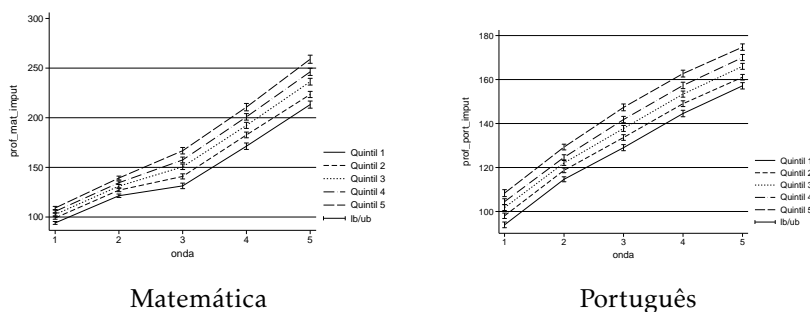


Figura 3: Evolução da proficiência por quintil de nível socioeconômico

município do Rio de Janeiro, mas ao final do período apresenta resultados estatisticamente iguais.

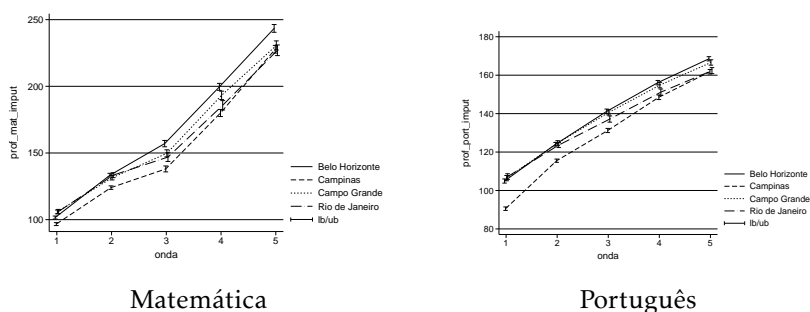


Figura 4: Evolução da proficiência por municípios

A figura 5 apresenta a evolução da proficiência para meninos e meninas separadamente. Para matemática, é possível perceber que no início do período escolar não havia diferenças significativas entre os dois grupos, permanecendo assim nas ondas dois e três. Já na quarta onda, as meninas apresentam uma proficiência média ligeiramente superior aos meninos, consolidando esta tendência na quinta onda. Vale ressaltar que esse resultado não retrata a realidade do Brasil, onde, de acordo com resultados da Prova Brasil, os meninos têm notas em matemática ligeiramente superior na quarta série. Para a proficiência em português, o resultado é semelhante ao observado no Brasil, ou seja, as meninas apresentam uma proficiência superior à dos meninos. No caso do GERES, essa diferença aumenta ao longo dos anos. Esse resultado é importante, pois, sugere que a desigualdade de renda observada no mercado de trabalho entre homens e mulheres não decorre das desigualdades educacionais.

A tabela 3 mostra o comportamento da desigualdade geral de proficiência nos quatro primeiros anos do Ensino Fundamental. O resultado para a onda um mostra que as crianças já iniciam o período escolar com uma desigualdade de conhecimentos, com valores próximos em ambas as proficiências. No entanto, a evolução da desigualdade é muito diferente para matemática e português. Enquanto que em português existe uma quase estabilidade da medida de desigualdade, com uma elevação de apenas 5%, para matemática,

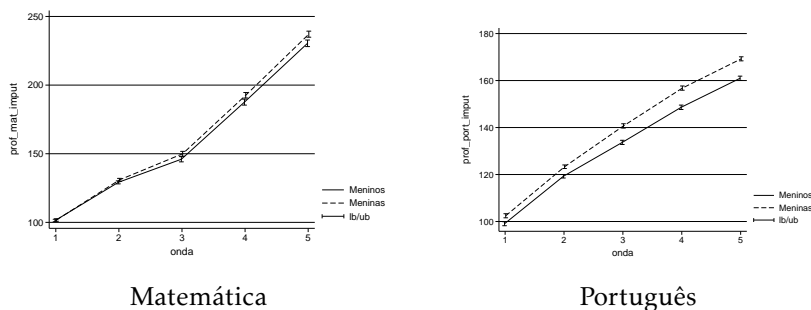


Figura 5: Evolução da proficiência por gênero

a elevação é de mais de 153%. Tal diferença pode ser explicada pela maior dificuldade em recuperar o atraso em matemática do que o atraso em português. Em função desse fato, uma parcela dos alunos, que entrou com uma habilidade em matemática mais baixa, pode ter uma dificuldade progressiva na aprendizagem, como bem esclarece Brooke et al. (2014).

Vale lembrar que a análise neste trabalho é apenas para os quatro primeiros anos do Ensino Fundamental. Se a tendência aqui encontrada continuar nos próximos anos do Ensino Fundamental e no Ensino Médio, a desigualdade em matemática irá atingir patamares ainda mais elevados ao final do Ensino Médio, o que é um alerta para a necessidade de investimentos nas crianças antes mesmo da entrada delas no Ensino Fundamental. Com isso, seria possível reduzir a desigualdade inicial e o atraso de parcela das crianças.

Tabela 3: Evolução da Desigualdade de Proficiência

Onda	Matemática	Português
1	24,12	22,70
2	29,14	19,76
3	48,85	24,01
4	55,41	24,59
5	61,14	23,83
Varição %	153,52	4,94

Fonte: GERES.

Os próximos resultados se referem à desigualdade de oportunidades. Os cálculos foram realizados utilizando a técnica descrita na seção metodológica. O primeiro exercício realizado estima uma regressão para cada circunstância separadamente, fornecendo a contribuição bruta de cada circunstância. A tabela 4 mostra os resultados para as quatro circunstâncias para as proficiências em matemática e português.

Em relação à proficiência de matemática, a primeira onda apresenta o nível socioeconômico com a maior contribuição para a desigualdade de habilidades (4,4%), seguida pelo município (2,5%), raça (0,8%) e gênero, sendo que esta última com praticamente nenhum efeito na desigualdade. A maior contribuição do nível socioeconômico está de acordo com a literatura da educação, que mostra que essa é a característica de maior impacto no aprendizado das crianças e também apresenta o maior impacto na desigualdade de notas (Diaz 2012, Ferreira & Gignoux 2014, Gamboa & Waltenberg 2012).

Ao analisar a evolução da contribuição de cada circunstância, nota-se um aumento no efeito da raça, nível socioeconômico e gênero e uma diminuição apenas no efeito do município. O maior aumento foi percebido pelo gênero da criança, com mais de 5.000% de elevação na contribuição. No entanto, continua sendo a circunstância com menor contribuição na quinta onda, apenas 0,3%. A contribuição da raça subiu praticamente 166%, passando de menos de 1% para mais de 2% e ocupando a segunda posição no ranking de maior contribuição. Já a contribuição do nível socioeconômico subiu em torno de 55%, passando a contribuir com quase 7% na desigualdade de matemática. Na última onda, o ranking de contribuição ficou com o nível socioeconômico em primeiro lugar (6,8%), seguido da raça (2,1%), município (1,4%) e gênero (0,3%).

Para a proficiência em português o ranking de contribuição na primeira onda aparece com o município em primeiro lugar (19,4%), seguido pelo nível socioeconômico (4,8%), o gênero (0,6%) e pela raça (0,05%). Apesar do município aparecer com a maior contribuição na primeira onda, essa contribuição diminui de forma significativa ao longo dos quatro anos, e, na quinta onda, tem a menor contribuição entre as circunstâncias analisadas (1,6%), representando uma queda de quase 85%. O nível socioeconômico tem um aumento de sua contribuição em mais de 40%, raça um aumento de mais de 377% e gênero de quase 450%. Ao final do período analisado, o ranking de contribuição para a desigualdade em português ficou com o nível socioeconômico em primeiro lugar (6,8%), seguido do gênero (3,1%), raça (2,3%) e por último o município (1,6%).

Tabela 4: Desigualdade de oportunidades - fatores individuais

Raça	Matemática			Português				
	Nse	Mun	Sexo	Raça	Nse	Mun	Sexo	
Onda 1	0,008	0,044	0,025	0,000	0,005	0,048	0,104	0,006
Onda 2	0,012	0,041	0,021	0,001	0,009	0,061	0,043	0,011
Onda 3	0,016	0,061	0,025	0,002	0,015	0,066	0,034	0,021
Onda 4	0,019	0,061	0,022	0,002	0,021	0,065	0,018	0,028
Onda 5	0,021	0,068	0,014	0,003	0,023	0,068	0,016	0,031
Varição %	165958	55175	-45401	5396515	377638	40990	-84676	449536

Fonte: GERES.

A tabela 5 apresenta os resultados da desigualdade de oportunidades considerando todas as circunstâncias conjuntamente, tanto para matemática quanto para português. São apresentadas as medidas para a desigualdade de oportunidades (R2) e para a contribuição marginal de cada circunstância, calculada pela decomposição de *shapley-shorroks*, para cada onda e a variação percentual entre a quinta e primeira onda.

Na primeira onda, a desigualdade de oportunidades em matemática era de 7,4%, enquanto que para português era de 16,8%. Uma análise da contribuição de cada circunstância, aponta que essa diferença se dá principalmente pelo efeito do município, que em português é de 10,8% e em matemática 2,5%. A evolução da desigualdade de oportunidades ao longo dos anos é distinta para as duas disciplinas. Enquanto que, para matemática, nota-se um crescimento de mais de 35%, em português, ocorre uma redução de mais de 20%. Mesmo com movimentos em sentidos opostos, ao final da quinta onda, a de-

sigualdade de oportunidades ainda é maior em português (13,3%) do que em matemática (10%). No entanto, na quinta onda, a diferença entre as duas disciplinas é explicada basicamente pelo gênero das crianças. Essa diferença entre os gêneros está de acordo com o que é apontado na literatura educacional e também com a tendência que já se delineava na figura 5. As meninas apresentam um aprendizado em português superior aos meninos, ao passo que, em matemática, essa diferença é menos significativa.⁶

É interessante comparar os resultados das contribuições de cada circunstância quando calculadas de forma bruta (tabela 4) de seus efeitos marginais (tabela 5). As contribuições do município de residência e do gênero não se alteram nas duas situações. Entretanto, as variáveis de raça e nível socioeconômico diminuem sua contribuição para a desigualdade, tanto na proficiência de matemática quanto na proficiência de português. Essa redução indica que existe uma correlação entre estas variáveis. Esse fato está de acordo com a literatura de desigualdade, que mostra que dentro do grupo de negros, o percentual de pessoas com baixo nível socioeconômico é maior do que entre o grupo de brancos. Apesar da redução do efeito da raça quando controlado pelo nível socioeconômico, ainda resta um efeito de quase 2% nas duas disciplinas. Este resultado é importante, pois chama a atenção para um problema de discriminação racial no sistema educacional brasileiro. Ainda mais preocupante é o fato que o efeito da raça na desigualdade cresce de forma abrupta nos primeiros anos escolares, em torno de 148% para matemática e 303% para português. Com os resultados obtidos neste artigo, não é possível determinar exatamente de que forma essa discriminação ocorre, se dentro da escola, se no acesso à melhores escolas ou se em características fora da escola. Soares & Alves (2003) e Soares et al. (2005) encontram resultados no mesmo sentido, mesmo controlando pelo nível socioeconômico, o desempenho dos negros é inferior ao desempenho dos brancos na escola.

É interessante examinar a desigualdade de oportunidades sem considerar o efeito do município, dado que a amostra deste estudo contém apenas quatro municípios e, portanto, pode enviesar os resultados. A tabela 6 apresenta os resultados para as estimações sem a circunstância município. As contribuições marginais de cada circunstância não sofrem alterações significativas em relação ao que foi apresentado na tabela 5. No entanto, a medida de desigualdade de oportunidades e, principalmente, sua evolução sofrem alterações interessantes. A desigualdade de oportunidades é inferior ao observado na tabela 5, em função de não contar mais com a contribuição do município. Já a evolução da desigualdade é crescente para as duas disciplinas, com um crescimento em torno de 80% para matemática e 108% para português.

6 Considerações Finais

O artigo avalia a evolução da desigualdade geral e de oportunidades no aprendizado de uma amostra de crianças durante os quatro primeiros anos do ensino fundamental. Ao utilizar uma base de dados longitudinal o artigo con-

⁶Na literatura educacional, os meninos aparecem com um rendimento em matemática superior ao das meninas, embora em menor magnitude que a diferença apresentada em português. Entretanto, no banco de dados utilizado neste estudo, as meninas aparecem com um resultado ligeiramente superior ao dos meninos. Apesar desta inversão entre os gêneros no resultado de matemática, o menor efeito do gênero na desigualdade de matemática está de acordo com o encontrado em outros estudos Gamboa & Waltenberg (2012)

Tabela 5: Desigualdade de oportunidade - nse, raça, sexo e município

Onda	r2	nse	mun	sexo	raça	nse %	mun %	sexo %	raça %
Matemática									
1	0,074	0,041	0,025	0,000	0,007	56295	34215	0,130	9360
2	0,072	0,039	0,022	0,001	0,010	54363	30311	1513	13814
3	0,102	0,060	0,027	0,002	0,014	58687	26077	1840	13396
4	0,099	0,058	0,023	0,002	0,016	58741	22914	2298	16047
5	0,100	0,065	0,014	0,003	0,017	65534	14027	3206	17233
Varição %	35180	57365	-44580	3244057	148881	16411	-59003	2373783	84111
Português									
1	0,168	0,049	0,108	0,006	0,005	29242	64447	3554	2757
2	0,126	0,061	0,046	0,011	0,008	48592	36350	8874	6185
3	0,135	0,066	0,035	0,022	0,012	48687	25818	16476	9019
4	0,129	0,064	0,019	0,029	0,018	49436	14457	22168	13939
5	0,133	0,067	0,016	0,031	0,019	50428	12065	23517	13990
Varição %	-20496	37106	-85116	426027	303418	72451	-81279	561637	407420

Fonte: GERES.

Tabela 6: Desigualdade de oportunidades - nse, raça e sexo

Onda	r2	Raça	nse	Sexo	Raça %	nse %	Sexo %
Matemática							
1	0,047	0,006	0,041	0,000	13549	86243	0,208
2	0,049	0,010	0,038	0,001	19710	77996	2294
3	0,074	0,014	0,058	0,002	18436	79065	2499
4	0,076	0,016	0,058	0,002	20714	76312	2974
5	0,085	0,017	0,064	0,003	20476	75716	3808
Varição %	80350	172556	58336	3200035	51127	-12206	1729797

Fonte: GERES.

tribui para a literatura da economia da educação avaliando o comportamento das desigualdades educacionais nos primeiros anos da escola, bem como as desigualdades de oportunidades. Com isso, é possível investigar o papel da escola no processo de geração/redução das desigualdades de habilidades entre os indivíduos. Essa análise foi realizada utilizando o arcabouço teórico da Igualdade de Oportunidades. Para mensurar as habilidades, foram utilizados os níveis de proficiência em matemática e português. Os resultados encontrados diferem entre as proficiências em matemática e português.

Os resultados devem ser analisados considerando três aspectos: desigualdade geral; desigualdade de oportunidades; e efeito marginal de cada circunstância. Em relação à desigualdade geral, destaca-se o aumento das desigualdades em matemática, enquanto que para português houve relativa estabilidade.

Os resultados para a desigualdade de oportunidades variam em função das circunstâncias consideradas. Quando consideradas as quatro circunstâncias conjuntamente, nível socioeconômico, gênero, raça e município de residência, a desigualdade de oportunidades em matemática cresce de aproximadamente 7% para 10%, enquanto que para português os valores se alteram de quase 17% para um pouco mais de 13%. Mas, apesar da desigualdade de oportunidades ser mais elevada para português, a taxa de crescimento desta desigualdade é positiva em matemática, cerca de 35%, enquanto que para português observa-se uma redução de pouco mais de 20%. No entanto, a queda na taxa

de desigualdade de oportunidades em português é impulsionada pelo efeito do município. Em função disso e da pouca representatividade do município na amostra utilizada, também foi testada uma especificação retirando o município da análise, mantendo as outras três circunstâncias. Nessa especificação, o valor da desigualdade de oportunidades é menor para as duas disciplinas, um pouco mais de 8% para matemática e de 11% para português. No entanto, as duas disciplinas apresentam um aumento nos valores ao longo dos quatro anos, sendo que português apresenta um crescimento superior ao de matemática, 109% e 80%, respectivamente.

Quando analisadas as circunstâncias isoladamente, o nível socioeconômico aparece como o fator de maior contribuição para a desigualdade educacional, reforçando evidências da literatura educacional que indicam que as características da família são os principais fatores determinantes do sucesso educacional. O município se mostra um fator importante no início do período escolar, no entanto, sua influência sobre a desigualdade educacional vai diminuindo ao longo do período escolar. Este resultado deve ser analisado com cautela, pois, nele estudo apenas quatro municípios fazem parte da amostra. A raça e o gênero da criança apresentam uma característica em comum, iniciam o período escolar influenciando pouco na desigualdade, tanto em matemática quanto em português, mas apresentam altas taxas de crescimento ao longo dos quatro anos. Vale lembrar que a análise é restrita aos quatro anos da amostra, ou seja, se a tendência de crescimento se mantiver no restante do período escolar, estes fatores irão explicar boa parte da desigualdade educacional.

Em uma perspectiva de políticas públicas, os resultados apresentados no artigo ressaltam a importância de investimentos na primeira infância para reduzir as desigualdades educacionais logo no início do período escolar. Pois, se a escola recebe crianças com uma diferença de habilidades muito acentuada, políticas educacionais voltadas apenas para melhorias de insumos escolares serão insuficientes para reduzir a desigualdade e principalmente a desigualdade de oportunidades.

Referências Bibliográficas

- Albernaz, A., Ferreira, F. & Franco, C. (2002), 'Qualidade e equidade no ensino fundamental brasileiro', *Pesquisa e Planejamento Econômico* 32(3).
- Arneson, R. (1989), 'Equality and equal opportunity for welfare', *Philosophical studies* 56(1), 77-93.
- Arneson, R. (1990), 'Liberalism, distributive subjectivism, and equal opportunity for welfare', *Philosophy & public affairs* pp. 158-194.
- Baker, F. & Kim, S. (2004), *Item response theory: Parameter estimation techniques*, CRC Press.
- Barro, R. (2000), 'Inequality and growth in a panel of countries', *Journal of economic growth* 5(1), 5-32.
- Barros, R. d., Foguel, M. & Ulyssea, G. (2006), Desigualdade de renda no Brasil: uma análise da queda recente, in 'Desigualdade de renda no Brasil: uma análise da queda recente', Ipea.

Barros, R. d. & Mendonça, R. (1995), 'Os determinantes da desigualdade no brasil'.

Betts, J. & Roemer, J. (2005), 'Equalizing opportunity for racial and socioeconomic groups in the united states through educational finance reform'.

Bonelli, R. & Sedlacek, G. (1988), 'Distribuição de renda: evolução no último quarto de século'.

Bourguignon, F. (1981), 'Pareto superiority of unegalitarian equilibria in stiglitz' model of wealth distribution with convex saving function', *Econometrica: Journal of the Econometric Society* pp. 1469–1475.

Brooke, N., da Silva Fernandes, N., de Miranda, I. & Soares, T. (2014), 'Modelagem do crescimento da aprendizagem nos anos iniciais com dados longitudinais da pesquisa geres', *Educ. Pesqui* 40(1), 77–94.

Checchi, D. & Peragine, V. (2005), Regional disparities and inequality of opportunity: the case of italy, Technical report, IZA Discussion Papers.

Checchi, D. & Peragine, V. (2010), 'Inequality of opportunity in italy', *The Journal of Economic Inequality* 8(4), 429–450.

Cohen, G. (1989), 'On the currency of egalitarian justice', *Ethics* pp. 906–944.

Deininger, K. & Squire, L. (1996), 'A new data set measuring income inequality', *The World Bank Economic Review* 10(3), 565–591.

Deininger, K. & Squire, L. (1998), 'New ways of looking at old issues: inequality and growth', *Journal of development economics* 57(2), 259–287.

Diaz, M. (2012), '(des)igualdades de oportunidades no ensino médio brasileiro: Escolas públicas e privadas', *Economia* 13(3a), 553–568.

Dworkin, R. (1981a), 'What is equality? part 1: Equality of welfare', *Philosophy & Public Affairs* pp. 185–246.

Dworkin, R. (1981b), 'What is equality? part 2: Equality of resources', *Philosophy & Public Affairs* pp. 283–345.

Ferreira, F. (2000), *Os determinantes da desigualdade de renda no Brasil: luta de classes ou heterogeneidade educacional?*, Pontifícia Universidade Católica de Rio de Janeiro, Departamento de Economia.

Ferreira, F. & Gignoux, J. (2014), 'The measurement of educational inequality: Achievement and opportunity', *The World Bank Economic Review* 28(2), 210–246.

Ferreira, F. & Litchfield, J. (1996), Growing apart: inequality and poverty trends in brazil in the 1980s, Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Ferreira, F. & Litchfield, J. (1999), 'Educacion o inflacion?: Explicando la desigualdad en brasil en la decada de los ochenta', *CÁRDENAS, M., LUSTIG, N.(eds.)* .

Ferreira, F., Walton, M. et al. (2005), *World development report 2006: equity and development*, Vol. 28, World Bank Publications.

Fishlow, A. (1972), 'Brazilian size distribution of income', *The American Economic Review* pp. 391–402.

Galor, O. & Moav, O. (2004), 'From physical to human capital accumulation: Inequality and the process of development', *The Review of Economic Studies* 71(4), 1001–1026.

Gamboa, L. & Waltenberg, F. (2012), 'Inequality of opportunity for educational achievement in latin america: Evidence from pisa 2006-2009', *Economics of Education Review* 31(5), 694–708.

Hoffman, R. (1989), 'Evolução da distribuição da renda no brasil, entre pessoas e entre famílias, 1979/86', *Mercado de trabalho e distribuição de renda: uma coletânea. Série Monográfica* 35.

Lam, D. & Levison, D. (1990), 'Idade, experiência, escolaridade e diferenciais de renda: Estados unidos e brasil', *Pesquisa e planejamento econômico* 20(2), 219–256.

Langoni, C. (2005), *Distribuição da renda e desenvolvimento econômico do Brasil*, FGV Editora.

Marrero, G. & Rodríguez, J. (2013), 'Inequality of opportunity and growth', *Journal of Development Economics* 104, 107–122.

Oliveira, L. (2008), Três investigações sobre escalas de proficiência e suas interpretações, PhD thesis, Tese (Doutorado em Educação)–Departamento de Educação, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Peragine, V. & Serlenga, L. (2007), Higher education and equality of opportunity in italy, IZA Discussion Papers 3163, Institute for the Study of Labor (IZA).

URL: <http://ideas.repec.org/p/iza/izadps/dp3163.html>

Pignataro, G. (2012), 'Equality of opportunity: Policy and measurement paradigms', *Journal of Economic Surveys* 26(5), 800–834.

Pinheiro, L., Oliveira Fontoura, N. & Querino, A. C. (2009), *Retrato das desigualdades de gênero e raça*, 3ra ed. Brasília: AECID; UNIFEM; IPEA.

Ramos, L. (1993), A distribuição de rendimentos no brasil: 1976/85, in 'IPEA', number 141, Ipea.

Ramos, X. (2012), 'Empirical approaches to inequality of opportunity: Principles, measures, and evidence'.

Rawls, J. (1971), 'A theory of justice'.

Rebelo, S. (1991), 'Long-run policy analysis and long-run growth', *Journal of Political Economy* 99(3), 500–521.

Riani, J. & Rios-Neto, E. (2008), 'Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros', *Revista Brasileira de Estudos de População* 25(2), 251–269.

Roemer, J. (1998), *Equality of opportunity*, Cambridge Univ Press.

Roemer, J. E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S. P., Lefranc, A., Marx, I., Page, M., Pommer, E., Ruiz-Castillo, J. et al. (2003), 'To what extent do fiscal regimes equalize opportunities for income acquisition among citizens?', *Journal of Public Economics* 87(3), 539–565.

Sen, A. (1979), 'Equality of what?'

Soares, J. & Alves, M. T. (2003), 'Desigualdades raciais no sistema brasileiro de educação básica', *Educação e Pesquisa* 29(1), 147–165.

Soares, J. & Oliveira, F. (2009), O cálculo de nível socioeconômico (nse) dos alunos do projeto geres, Technical report, Universidade Federal de Minas Gerais.

Soares, S., Beltrão, K., Barbosa, M. & Ferrão, M. E. (2005), Os mecanismos de discriminação racial nas escolas brasileiras, *in* 'Os Mecanismos de discriminação racial nas escolas brasileiras', Ipea.

Waltenberg, F. & Vandenberghe, V. (2007), 'What does it take to achieve equality of opportunity in education?: An empirical investigation based on brazilian data', *Economics of Education Review* 26(6), 709–723.

EFICIÊNCIA NOS MERCADOS FUTUROS AGROPECUÁRIOS BRASILEIROS

MARCOS AURELIO RODRIGUES *
JOÃO GOMES MARTINES FILHO †

Resumo

Objetivou-se testar a hipótese de passeio aleatório a contratos futuros agropecuários negociados na BM&FBOVESPA. Refutá-la, significa possível previsibilidade e, por conseguinte, os mercados não seriam fracamente eficientes. Correlações seriais e testes de razão de variância foram utilizados para verificá-las. Os resultados deram suporte à hipótese de passeio aleatório nos mercados futuros de café e da soja, eficientes na forma fraca, e evidências contrárias foram encontradas nos mercados do boi gordo, milho e etanol.

Palavras-chave: Razão de variância; eficiência de mercado; *commodities*.

Abstract

We aim to test the random walk hypothesis to agricultural future contracts traded at the Brazilian Board of Trade (BM&FBOVESPA). Refute this hypothesis means possible predictability, therefore these markets are not weakly efficient. We used tests of serial correlation and variance ratio to verify them. Our results do not reject the random walk hypothesis in coffee and soybeans markets but contrary evidences were found for live cattle, corn and ethanol markets.

Keywords: Variance ratio; market efficiency; *commodities*.

JEL classification: G14, Q14

DOI: <http://dx.doi.org/10.1590/1413-8050/ea91170>

* Tribunal de Justiça do Estado do Paraná. Curitiba, Paraná, Brasil. E-mail: rodriguesmarcosaurelio@gmail.com

† Escola Superior de Agricultura “Luiz de Queiroz” – USP. Piracicaba, São Paulo, BRasil. E-mail: martines@usp.br

1 Introdução

Inferências sobre predicabilidade dos retornos em finanças possuem importantes implicações práticas e teóricas. Os agentes do agronegócio podem tomar decisões errôneas com respeito ao conjunto de informações incompletas contidas nos preços passados, se os mercados forem não eficientes em sua forma fraca. Tomadas de decisões sujeitas à incompletude informacional implicam custos alocativos ineficientes de seus recursos produtivos.

Nos argumentos de Samuelson (1965) e Fama (1970), mudanças nos preços não podem ser previsíveis se incorporadas totalmente as informações e expectativas de todos os participantes dos mercados. Quanto mais eficiente for o mercado, mais aleatória será a sequência de preços gerados por ele, e o mais eficiente dos mercados será aquele cujas mudanças nos preços sejam completamente aleatórias e imprevisíveis.

Sujeitos a esse comportamento de mercado, investidores não devem esperar lucros que excedam o retorno total de mercado. Em um mercado eficiente, a arbitragem ou possibilidade de lucro livre de risco, com base na diferença de preços sob o mesmo ativo em diferentes mercados, não é possível ser obtida. Estratégias efetuadas levando-se em consideração que os retornos futuros podem ser preditos, somente com base em seu comportamento passado, partem da invalidade da hipótese de eficiência em sua forma fraca, e carteiras de investimento construídas sob modelos de risco retorno dependem do comportamento de passeio aleatório às séries, exemplificando a relevância em estudar o comportamento dos ativos agropecuários negociados na BM&FBOVESPA.

Embora a literatura internacional seja vasta em estudos que envolvam a análise de eficiência de seus mercados, no Brasil sua maturidade limita-se aos mercados de ativos financeiros. Os estudos sobre os mercados futuros agropecuários são incipientes e pouco explorados metodologicamente (Bitencourt 2007). Bressan & Leite (2001) fizeram-na com testes de correlação serial, Bitencourt (2007) e Righi & Ceretta (2011) testaram sob a abordagem de razão de variância. A hipótese conjunta de não enviesamento e eficiência a preços futuros, sob abordagem de cointegração, foi a metodologia mais empregada nos estudos de derivativos agropecuários brasileiros, realizada por Amado & Carmona (2004), Bitencourt (2007), Duarte et al. (2007), Alves et al. (2008), Moraes et al. (2009), Silva Neto et al. (2010) e Fraga & Silva Neto (2011).

A contribuição deste estudo está em reexaminar a forma fraca de eficiência no mercado futuro agropecuário brasileiro sob quatro aspectos. Primeiro, quantidade amostral maior que a empregada pela literatura aos contratos de boi gordo, café, etanol, milho e soja. Segundo, avaliar a forma fraca de eficiência nesses mercados com testes de razão de variância individuais propostos por Lo & MacKinlay (1988) e múltiplos de Chow & Denning (1993). Terceiro, examinar a hipótese de passeio aleatório nos mercados agropecuários com a recente contribuição de Kim (2006), que propôs o *wild bootstrap* dos testes individuais e múltiplos. Quarto, considerar a característica díspar de homocedasticidade e heterocedasticidade presente nas séries diárias dos contratos futuros agropecuários negociados na BM&FBOVESPA.

O estudo encontra-se estruturado em mais quatro seções além desta introdução: descrevem-se na segunda seção a discussão teórica sobre eficiência de mercado e a literatura empírica sobre eficiência nos mercados de *commodities*; na terceira seção a metodologia para o teste de hipótese à eficiência fraca nesses mercados; por conseguinte, na quarta seção os resultados empíricos

encontrados. Evidenciam-se em seguida as considerações finais.

2 Hipótese da eficiência de mercado

A hipótese de mercado eficiente remonta ao trabalho de Samuelson (1965), que é creditado por dar respeitabilidade acadêmica à hipótese de passeio aleatório (Lim 2009). Samuelson (1965) elucidou a ideia de mercados eficientes através do seu interesse na precificação temporal de *commodities* estocáveis. Demonstrou que em um mercado informacionalmente eficiente, as mudanças nos preços devem ser imprevisíveis se forem apropriadamente antecipadas, uma vez que as expectativas e informações de todos os participantes dos mercados são incorporadas completamente. Fama (1970) argumenta que antes de Samuelson (1965), nenhum outro estudo relacionou as teorias de mercados eficientes e passeio aleatório de forma rigorosa, sendo que foi realizado com base na análise dos contratos futuros em mercado de *commodities*.

O comportamento de caminho aleatório dos preços futuros resulta do perfeito funcionamento dos mercados futuros, sendo definido um mercado futuro perfeito, aquele em que o preço de mercado constitui, em todos os períodos, a melhor estimativa a ser feita, a partir da informação disponível corrente, do preço que será na data de vencimento dos contratos futuros (Working 1962). De acordo com Park & Irwin (2004), essa definição de mercado futuro perfeito é essencialmente idêntica à de mercado eficiente disposta em Fama (1970), a qual serve como definição padrão na literatura econômica financeira.

Um mercado no qual os preços sempre “refletem completamente” as informações disponíveis é chamado “eficiente” (Fama 1970). Neste estudo, revisou-se a literatura teórica e empírica à época e distinguiu-os entre três formas ao teste da hipótese de mercado eficiente, refletido o processo informacional em ordem crescente e sujeitos a subconjuntos de informações relevantes: fraca, semiforte e forte. Na primeira, o conjunto de informações são apenas os preços ou retornos históricos. Na segunda, incluem-se à anterior outras informações publicamente disponíveis. A terceira consiste em todas as informações públicas e, também, quaisquer informações privadas relevantes à formação dos preços que investidores ou grupos detêm.

Uma definição mais recente à eficiência de mercado foi proposta por Jarrow & Larsson (2012): um mercado (\mathbb{F}, S) é chamado de eficiente¹ em um horizonte de tempo finito $[0, T]$ com respeito a \mathbb{F} se existe um índice de preço de um bem de consumo ψ e uma economia $(\{P_k\}_{k=1}^K, \mathbb{F}, \{\epsilon_k\}_{k=1}^K, \{U_k\}_{k=1}^K)$ para cada (ψ, S) sendo um processo de equilíbrio de preço S em $[0, T]$. Se isso se mantém para todo $T < \infty$, o mercado é chamado de eficiente com respeito a \mathbb{F} . Segundo os autores, esse equilíbrio relaciona-se com a noção de equilíbrio de expectativas racionais totalmente revelada sob a forma forte de eficiência de mercado e parcialmente revelada sob a forma fraca de eficiência de mercado.

Segundo Charles & Darné (2009), a hipótese de passeio aleatório provê uma maneira de testar a forma fraca de eficiência de mercado proposta por Fama (1970) e, conseqüentemente, não previsibilidade em mercados financeiros. A forma fraca de eficiência de mercado é comumente examinada com a abordagem da hipótese de passeio aleatório (Lim 2009).

¹ \mathbb{F} := filtro de informação; S := ativo; P_k := finitas crenças dos investidores; ϵ := *endowment stream*; U := função de utilidade.

As versões de passeio aleatório foram exemplificadas em Campbell et al. (1997). Considere a especificação:

$$p_t = \omega + p_{t-1} + \epsilon_t, \quad (1)$$

em que p_t representa o logaritmo natural dos preços futuros, ω um possível *drift* e ϵ_t o erro. Os três passeios aleatórios de Campbell et al. (1997) diferem em relação às suposições a ϵ_t . No passeio aleatório 1 (PA1), o erro de (1) é independente e identicamente distribuído (iid),

$$\epsilon_t \sim iid(0, \sigma^2),$$

com média 0 e variância σ^2 . No passeio aleatório 2 (PA2), o erro de (1) é independente mas não identicamente distribuído (inid),

$$\epsilon_t \sim inid(0, \sigma^2).$$

No passeio aleatório 3 (PA3), o erro de (1) é não independente e não identicamente distribuído (niid),

$$\epsilon_t \sim niid(0, \sigma^2).$$

2.1 Evidências empíricas sobre eficiência nos mercados de *commodities*

Os apontamentos de Garcia et al. (1988) a fatores identificados na literatura que potencialmente influenciam a eficiência de mercados futuros agropecuários, dentre eles, as diferentes características das *commodities* e de seus mercados, o tipo de dados utilizados na análise e a natureza dos testes empregados servem ao delineamento a seção. Portanto, antes das evidências empíricas na literatura, com base nos testes realizados, destacaram-se os estudos segundo suas características, em conjunto com o intervalo amostral, e, em seguida, abordou-se a frequência dos dados utilizados nos estudos.

Devido às características de estocabilidades semelhantes, inerentes a cada *commodity*, podem ser divididas entre estocáveis e não estocáveis. Os estudos sobre eficiência de mercado no Brasil concentram-se nas mercadorias estocáveis. A análise dos contratos futuros de café foi feita por: Bressan & Leite (2001) entre 1992 e 1998, Amado & Carmona (2004) de 1995 a 2003, Bitencourt (2007) de 1996 a 2006 e Cruz Júnior & Silveira (2007) de 2002 a 2006; de açúcar por Amado & Carmona (2004) entre 1995 e 2003; de álcool anidro por Alves et al. (2008) entre 2000 e 2006; de milho por Amado & Carmona (2004) entre 1996 e 2003 e de soja por Duarte et al. (2007) no ano de 2005 e Fraga & Silva Neto (2011) de 2007 a 2008. A eficiência dos mercados à vista foi verificada somente por Righi & Ceretta (2011) entre 2006 e 2010, para algodão, café, milho e soja. Apenas dois trabalhos focaram na análise do contrato futuro de boi gordo, *commodity* não estocável, Moraes et al. (2009) de 2000 a 2004 e Silva Neto et al. (2010) entre 2007 e 2008.

Com respeito à frequência utilizada por estes, houve predominância do uso de séries diárias. Os únicos trabalhos que consideraram frequências diferentes foram Alves et al. (2008), que utilizaram dados semanais, e Bitencourt (2007), que utilizou dados diários, semanais e mensais. Sua justificativa foi relacioná-los a testes de curto, médio e longo prazo ao proporcionar robustez a amostra utilizada.

Conforme disposto em Lim & Brooks (2011), testes de correlação serial e análise espectral serviram como as primeiras ferramentas empregadas pela literatura para testar a hipótese de eficiência de mercado em sua forma fraca, de forma pioneira por Fama (1965) e Granger & Morgenstern (1963), respectivamente.

Bressan & Leite (2001) testaram se houve autocorrelação nas séries de café no mercado futuro da BM&F, verificado por correlogramas com até cinco defasagens e estatística Q de Ljung & Box (1978). Conduzidos esses testes para os 31 contratos analisados, no período entre março de 1992 e março de 1998, 42% deles apresentaram indícios de não eficiência. Sob uma abordagem que buscou verificar a existência de autocorrelação entre os erros de previsão em diferentes horizontes, cinco dias antes do vencimento até noventa dias, variando de cinco em cinco, não houve indícios de ineficiência, pois os coeficientes de autocorrelação foram estatisticamente iguais a zero. Concluíram que o contrato futuro de café foi eficiente no período analisado, sendo que os preços se comportaram como num processo de passeio aleatório.

De acordo com o *survey* de Lim & Brooks (2011), a presença de raiz unitária foi utilizada na literatura para testar a forma fraca da hipótese de eficiência de mercado. Os logaritmos dos preços em nível foram não estacionários em estudos revistos por esses autores, assim, os pesquisadores concluíram que os mercados sob estudo apresentaram evidência a favor da hipótese fraca de eficiência.

Apesar de a maioria dos estudos nos mercados agropecuários brasileiros não explicitarem, exceto em Cruz Júnior & Silveira (2007), a presença de raiz unitária poderia auxiliar na inferência sobre a hipótese de eficiência de mercado em sua forma fraca. Mas Campbell et al. (1997) ressaltam que devido à possibilidade de os ϵ 's serem processos estacionários com média zero arbitrários, tanto sob a hipótese nula em (1) como sob a hipótese alternativa, o foco desses testes de raiz unitária não é sob a previsibilidade de p_t . Mesmo sob a hipótese nula, os incrementos de p_t podem ser preditos. Conforme seus argumentos, uma vez que existem alternativas ao passeio aleatório sob a hipótese nula de raiz unitária, como, por exemplo, outras formas de dependência serial, esses testes claramente não foram concebidos para detectar a previsibilidade, mas, de fato, são insensíveis por sua construção.

Em resposta ao interesse sobre não estacionariedade nos preços, cointegração e modelos com correção de erro têm sido usados para testar o não enviesamento e eficiência nos mercados futuros (Garcia & Leuthold 2004). O conceito de não enviesamento é uma versão mais restritiva da forma fraca de eficiência de mercado sugerido por Fama (1970), ao implicar que o preço futuro corrente de uma *commodity* deve ser igual ao preço esperado no mercado à vista, na data de vencimento do contrato (McKenzie et al. 2002). Uma versão simplista a esse teste é verificar a hipótese nula $\alpha = 0$ e $\beta = 1$ em

$$S_t = \alpha + \beta F_{t-1} + \epsilon_t, \quad (2)$$

tal que S é o preço à vista, F o preço futuro e ϵ o erro com média 0 e variância constante. Variações mais complexas a esta foram realizadas, mas a intuição permanece a mesma.

Estudos que analisaram os mercados sob essa abordagem cointegrante ao teste conjunto de eficiência e não enviesamento foram Amado & Carmona (2004), Bitencourt (2007), Duarte et al. (2007), Alves et al. (2008), Moraes et al. (2009),

Silva Neto et al. (2010) e Fraga & Silva Neto (2011). A existência de cointegração às séries à vista e futuras, considerada como condição necessária, mas não suficiente, à eficiência de mercado, foi reportada em todos esses estudos².

Como requisito à hipótese de eficiência de mercado, a hipótese de restrição aos coeficientes α e β $[0, 1]$ sob teste de razão de verossimilhança, ao vetor cointegrante, atendeu à hipótese conjunta de eficiência de mercado e não enviesamento nos preços futuros de café, no estudo de Bitencourt (2007). Alves et al. (2008) concluíram que o mercado de álcool anidro não atendeu à hipótese de eficiência de mercado pois o teste sob $\beta = 1$ foi rejeitado. Duarte et al. (2007) concluíram que o mercado de soja futuro foi um preditor não enviesado e eficiente em contraposto ao estudo de Fraga & Silva Neto (2011), pois a hipótese de mercado eficiente para a soja não foi aceita. A restrição conjunta de α e β $[0, 1]$, imposta ao vetor de cointegração no estudo de Moraes et al. (2009), não foi rejeitada, conseqüentemente, deu suporte à eficiência e não viés para o mercado de boi gordo. Resultado semelhante ao de Silva Neto et al. (2010), ao sugerirem a não rejeição das hipóteses de mercado eficiente, entretanto evidenciaram a existência de prêmio de risco.

Desde o estudo de Lo & MacKinlay (1988), os testes de razão da variância emergiram como ferramenta primária para verificar se os retornos das séries acionárias são não correlacionados serialmente (Lim & Brooks 2011). Lo & MacKinlay (1988) examinaram os testes de razão de variância propostos por eles, o de raiz unitária sugerido por Dickey & Fuller (1979, 1981), além do teste para correlação serial de Box & Pierce (1970). Os autores encontraram que os de razão de variância proveem maior poder que os demais sob passeio aleatório heterocedástico. Os estudos de Bitencourt (2007) e Righi & Ceretta (2011) foram os únicos a utilizarem-nos na inferência sobre a eficiência fraca no mercado futuro de café e nos mercados à vista do algodão, café, milho e soja, respectivamente.

Bitencourt (2007) realizou o teste de razão de variância proposto por Lo & MacKinlay (1988) para retornos diários, semanais e mensais. Esses apresentaram evidências contrárias à hipótese de passeio aleatório para todas as frequências. A rejeição do passeio aleatório sugeriu que os retornos de curto prazo apresentaram persistência. Além do teste de Lo & MacKinlay (1988), realizado por Bitencourt (2007), Righi & Ceretta (2011) utilizaram os testes de Chow & Denning (1993), Wright (2000) e Chen & Deo (2006). Os autores concluíram que para todas as *commodities* se rejeitou a hipótese de passeio aleatório, tanto nos testes que consideram defasagens individualmente, quanto nas estatísticas conjuntas. Isso implicou a rejeição de eficiência de mercado na sua forma fraca aos mercados à vista analisados.

Conforme apontado por Garcia et al. (1988), há divergência nas evidências sobre eficiência nos mercados, devido à natureza dos testes e características dos mercados. Como pôde ser notado, a falta de consenso, tanto metodológico como nos resultados, reforça o objetivo do presente estudo para um maior entendimento com respeito ao comportamento das *commodities* agropecuárias negociadas na BM&FBOVESPA.

²Fraga & Silva Neto (2011) reportam que na praça de Dourados não houve cointegração com o mercado futuro de soja.

3 Metodologia

Estatísticas de portmanteau foram estabelecidas para verificar a hipótese de passeio aleatório 1 de Campbell et al. (1997). Uma vez que essa hipótese implica autocorrelações iguais a zero, as estatísticas Q de Box & Pierce (1970) e Ljung & Box (1978), aplicadas aos retornos, fornecem evidências sobre a eficiência dos cinco contratos agropecuários negociados na BM&FBOVESPA. A análise da hipótese de passeio aleatório 3 de Campbell et al. (1997) foi feita por meio de testes de razão de variância sugeridos por Lo & MacKinlay (1988) e Chow & Denning (1993), sob *wild bootstrap*, seguida a proposta de Kim (2006). O uso de suas versões robustas a heterocedasticidade foi escolhido após análise das estatísticas de portmanteau nos resíduos ao quadrado, assim como da estatística dos Multiplicadores de Lagrange, proposta por Engle (1982). Os testes são complementares e, aplicadas as três classes de testes, proveem robustez no estabelecimento das conclusões.

Apesar do desenvolvimento de inúmeros testes estatísticos para verificar a hipótese de passeio aleatório, a classe de testes com base na metodologia da razão de variância tem ganho popularidade nos anos recentes (Charles & Darné 2009). Segundo seu *survey*, a metodologia de razão de variância consiste em testar a hipótese de passeio aleatório contra a alternativa de estacionariedade, ao explorar o fato de que a variância dos incrementos de um passeio aleatório é linear em todos os intervalos amostrais, isto é, a variância amostral do retorno no período k , de uma série temporal y_t , é k vezes a variância amostral do retorno de um período.

Seja y_t o retorno de uma *commodity* no tempo t , tal que $t = 1, \dots, T$. A razão de variância da k -ésima diferença escalonada por k , com respeito à variância da primeira diferença, tende a ser igual a um, isto é,

$$VR(k) = \frac{\sigma^2(k)}{\sigma^2(1)}, \quad (3)$$

em que $\sigma^2(k)$ é $\frac{1}{k}$ a variância da k -ésima diferença e $\sigma^2(1)$ é a variância da primeira diferença. Sob a hipótese nula de passeio aleatório, $VR(k)$ deve aproximar-se à unidade. Se essa razão for menor do que 1 em longos horizontes, têm-se indícios de correlação serial negativa (reversão à média) e razões maiores do que 1 em horizontes longos indicam correlação serial positiva (aversão à média ou persistência) (Charles & Darné 2009).

Lo & MacKinlay (1988) propuseram dois testes estatísticos que exploram a propriedade de (3). Defina o estimador para a variância da k -ésima diferença, $\sigma^2(k)$, como

$$\sigma^2(k) = \frac{1}{Tk} \sum_{t=k}^{Tq} (y_t + \dots, y_{t-k+1} - k\hat{\mu})^2 \quad (4)$$

sendo $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$ e o estimador da variância da primeira diferença, $\sigma^2(1)$, como

$$\sigma^2(1) = \frac{1}{T} \sum_{t=k}^{Tq} (y_t - \hat{\mu})^2. \quad (5)$$

Os autores mostraram que, sob a suposição de homocedasticidade, então a hipótese nula que $V(k) = 1$, a estatística

$$M_1(k) = \frac{VR(k) - 1}{\varphi(k)^{\frac{1}{2}}} \quad (6)$$

é assintoticamente distribuída como uma $\mathcal{N}(0, 1)$, onde

$$\varphi_0(k) = \frac{2(2k-1)(k-1)}{3kT}. \quad (7)$$

Para acomodar a presença de heterocedasticidade em y_t , os autores propuseram uma segunda estatística, robusta sob heterocedasticidade e que segue uma distribuição normal assintótica, definida como:

$$M_2(k) = \frac{VR(k) - 1}{\varphi^*(k)^{\frac{1}{2}}} \quad (8)$$

tal que

$$\varphi^*(k) = \sum_{j=1}^{k-1} \left[\frac{2(k-j)}{k} \right]^2 \delta(j) \quad (9)$$

e

$$\delta(j) = \frac{\sum_{t=j+1}^T (y_t - \hat{\mu})^2 (y_{t-j} - \hat{\mu})^2}{\left[\sum_{t=1}^T (y_t - \hat{\mu})^2 \right]^2}. \quad (10)$$

A proposta de Lo & MacKinlay (1988) verifica a hipótese nula para um k valor individual. Mas a questão central é se as séries temporais revertem à média, requerendo a validade da hipótese nula para todos os valores de k . Assim, necessita-se um teste conjunto a múltipla comparação das razões de variância sob diferentes horizontes. Conduzi-los em separado, testando sequencialmente diversos valores de k pode levar a sobre rejeição da hipótese nula e superdimensionamento. Disso, a fraqueza do teste de Lo & MacKinlay (1988) é ignorar a natureza conjunta do teste à hipótese de passeio aleatório (Charles & Darné 2009).

Para contornar isso, Chow & Denning (1993) propuseram um teste múltiplo de razão de variância que requer somente o valor máximo absoluto de $VR(k)$ no conjunto de m testes estatísticos considerados. A estatística é definida por

$$CD_1 = \sqrt{T} \max_{1 \leq i \leq m} |M_1(k_i)| \quad (11)$$

e segue uma distribuição módulo máximo studentizada, com m e T graus de liberdade, $MMS(\alpha, m, T)$. A hipótese nula de passeio aleatório é rejeitada ao nível α de significância se a estatística MV_1 for maior do que $\left[1 - \frac{\alpha^*}{2}\right]$ -ésimo percentil da distribuição normal, tal que $\alpha^* = 1 - (1 - \alpha)^{\frac{1}{m}}$. Entretanto, essa estatística só é válida sob retornos homocedásticos. A versão robusta à heterocedasticidade sugerida por Chow & Denning (1993) pode ser escrita por

$$CD_2 = \sqrt{T} \max_{1 \leq i \leq m} |M_2(k_i)| \quad (12)$$

com os mesmos valores críticos de CD_1 .

Segundo Kim (2006), os testes de Lo & MacKinlay (1988) e Chow & Denning (1993) são assintóticos e podem resultar em deficiências sob pequenas amostras. Kim (2006) propôs como alternativa a esse possível problema desses testes o uso de *wild bootstrap*, descrito pelo autor como um método de reamostragem que aproxima a distribuição amostral de uma estatística e é aplicável sobre dados com formas desconhecidas de heterocedasticidade condicional e não condicional. Exemplificando o teste com *wild bootstrap*, embasado em CD_2 , conduzem-se os três seguintes procedimentos:

- formar uma amostra de *bootstrap* de T observações $y_t^* = \eta_t y_t$ ($t = 1, \dots, T$), tal que η_t é uma sequência aleatória com $E(\eta_t) = 0$ e $E(\eta_t^2) = 1$;
- calcular CD_2^* , a qual é a estatística da equação (12) das amostragens geradas por *bootstrap* no primeiro procedimento;
- repetir o primeiro e segundo procedimento n vezes para formar uma distribuição *bootstrap* $\{CD_2^*(j)\}_{j=1}^n$ para a estatística do teste.

A distribuição *bootstrap* $\{CD_2^*(j)\}_{j=1}^n$ é utilizada para aproximar a distribuição amostral da estatística CD_2 . O valor-p do teste é estimado como uma proporção de $\{CD_2^*(j)\}_{j=1}^n$ maior do que o valor amostral de CD_2 Kim (2006).

3.1 Base de dados

A base de dados foi composta pelos preços de fechamento dos contratos³ do boi gordo (BGI), milho (CCM), etanol (ETH), café (ICF) e soja (SOJ), mais próximos a vencer, negociados na BM&FBOVESPA e obtidos no sistema de recuperação de dados desta bolsa. Justifica-se o uso dos contratos mais próximos ao vencimento, pois foram os mais ativos em volume negocial no período em análise. A rolagem do contrato próximo a expirar, para o contrato subsequente, ocorreu no dia anterior ao encerramento de cada contrato.

Tabela 1: Base de dados

Ativos — Código	Período amostral		Número de observações
BM&FBOVESPA	Início	Final	
Boi gordo — BGI	25 de setembro de 2000	29 de dezembro de 2011	2790
Milho — CCM	19 de setembro de 2008	29 de dezembro de 2011	812
Etanol — ETH	17 de maio de 2010	29 de dezembro de 2011	406
Café — ICF	03 de janeiro de 2000	29 de dezembro de 2011	2972
Soja — SOJ	27 de agosto de 2004	09 de junho de 2011	1676

Fonte: Sistema de recuperação de dados da BM&FBOVESPA.

O período escolhido e disposto na Tabela 1 às séries deu-se à disponibilidade desses dados no sistema e também para que fossem mantidas as principais características contratuais do final amostral. O final amostral às séries foi 29 de dezembro de 2011, exceto para o contrato de soja, que se encerrou no dia 09 de junho de 2011.

³Utilizaram-se seus respectivos códigos na BM&FBOVESPA para denotá-los.

4 Resultados e discussões

Antes de reportar os resultados dos testes, com o intuito de verificar a hipótese de passeio aleatório e, conseqüentemente, a hipótese de eficiência fraca às *commodities* em análise, apresentam-se na Tabela 2 as estatísticas descritivas, ressaltando algumas características das séries diárias.

A estatística do teste Jarque & Bera (1980) indicou rejeição da normalidade às séries desses produtos. Suas distribuições apresentaram comportamento de não normalidade que se relacionam à assimetria e excesso de curtose. Tanto para as séries em nível como às de retorno, para o boi gordo, café e milho apresentaram assimetria à direita e para o etanol e soja assimetria à esquerda. Com relação à curtose, as séries em nível indicaram platocurtose e as de retorno leptocurtose.

Com base nas estatísticas dos testes de Ljung & Box (1978) e Box & Pierce (1970) detectou-se forte presença de autocorrelação, tanto para as séries em nível como de retorno, à significância estatística de 1%, exceto nos retornos de milho e soja, que foram estatisticamente significativos a 10% e 15%, respectivamente. Como o passeio aleatório 1 de Campbell et al. (1997) implica que todas as autocorrelações sejam iguais a zero, dessa autocorrelação nas séries em nível, têm-se as primeiras evidências sobre a hipótese de eficiência de mercado em sua forma fraca, rejeitando-a, conforme os primeiros estudos que a testaram. Entretanto, deve-se ter cautela nessa inferência, pois o estudo de Lo & MacKinlay (1988) demonstrou que esses testes têm baixo poder, comparados aos de razão de variância, sob passeio aleatório.

Entretanto, resultados obtidos por testes convencionais como estatística F, Box & Pierce (1970), Ljung & Box (1978) e Dickey & Fuller (1979, 1981) podem gerar resultados errôneos para verificar um passeio aleatório (Hakkio 1986). Em simulações de Monte Carlo, demonstrou que esses testes têm baixo poder relativo, pois não distinguem entre um passeio aleatório e algo próximo a um passeio aleatório.

Seguido o critério de Tsay (2005), para a seleção de defasagens ao teste de Ljung & Box (1978), $ln(T)$, as estatísticas de Ljung & Box (1978) e Box & Pierce (1970), às séries ao quadrado, assim como o teste dos multiplicadores de Lagrange de Engle (1982), sugerem existência de heterocedasticidade nos preços em nível e de retorno a três contratos: boi gordo, café e soja. Ressalta sua não presença nos retornos de milho e etanol, fato estilizado esperado em séries financeiras diárias.

A presença de autocorrelações indicadas na Tabela 3 sugere razões de variância diferentes à unidade, confirmadas na Figura 1. Nela dispõem-se as estimativas das razões de variância até 8 *lags* com intervalo de confiança a 95%.

Em um mercado eficiente, espera-se que as razões de variância sejam próximas à unidade. Entretanto, apenas para a série do café isso ocorre de forma clara. Com o aumento do horizonte temporal, houve indícios de reversão à média. Reversão à média também ocorreu no caso do etanol, mas com um horizonte de tempo maior, comparado ao contrato de café, logo, ambos serialmente correlacionados negativamente. Nas demais séries, o comportamento das razões de variância apresentou valores crescentes e superiores à unidade, característico de aversão à média, demonstrando crescimento da variância mais do que proporcional com o tempo. Isso sugere que os retornos dessas séries possuíram presença dominante de autocorrelações positivas.

Tabela 2: Estatísticas descritivas

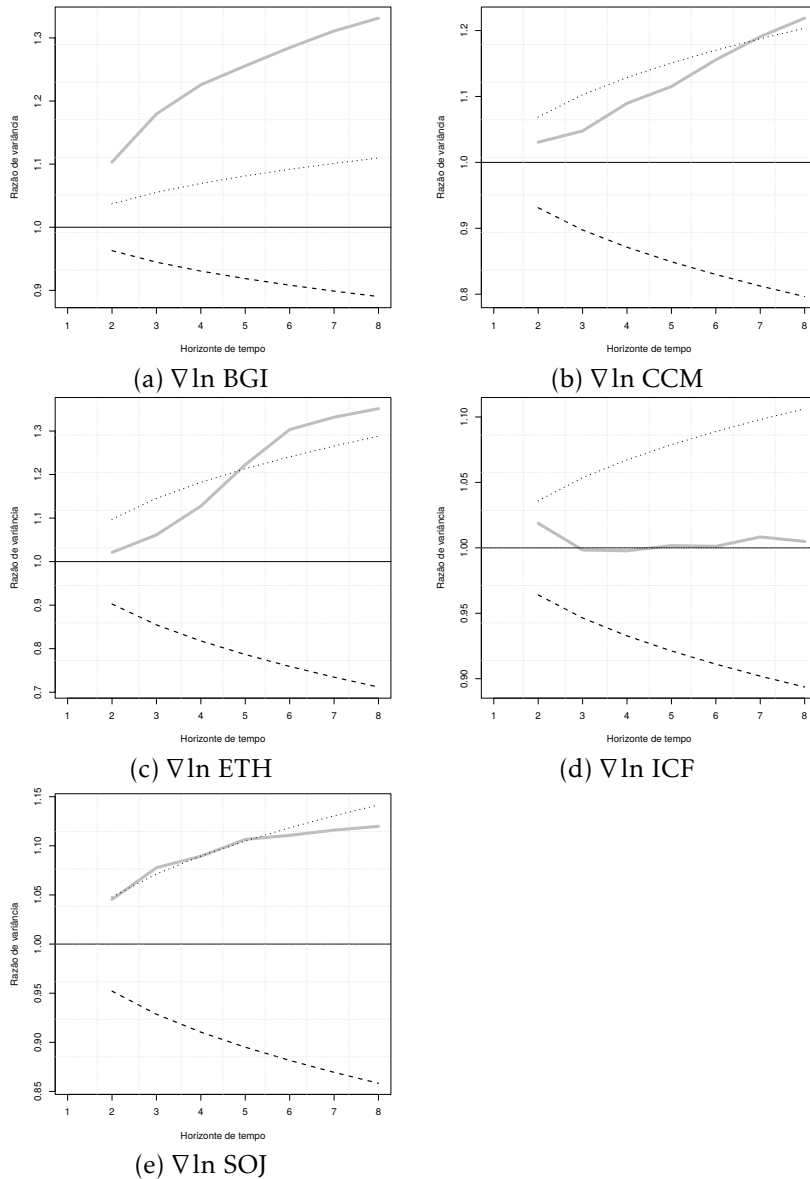
	BGI		CCM		ETH		ICF		SOJ	
	ln	∇ln	ln	∇ln	ln	∇ln	ln	∇ln	ln	∇ln
Observações	2790	2789	812	811	406	405	2972	2971	1676	1675
Mínimo	3,6243	-0,1049	2,8758	-0,0980	6,5827	-0,3352	3,7136	-0,1522	2,3933	-0,1616
Máximo	4,7532	0,0862	3,4819	0,0917	7,4413	0,0629	5,9839	0,1474	3,5779	0,0488
Média	4,1442	0,0003	3,1628	0,0002	7,0086	0,0013	4,7530	0,0003	2,9718	0,0005
Desvio Padrão	0,2798	0,0097	0,1794	0,0136	0,1847	0,0240	0,5243	0,0205	0,3213	0,0153
Assimetria	0,2534	0,0586	0,1517	0,0379	-0,5625	-9,7788	0,1840	0,3371	-0,0552	-1,3296
Curtose	-0,9585	13,3671	-1,4347	9,1537	-0,2118	127,1725	-0,4875	5,8234	-1,4006	11,4595
Jarque-Bera	136,36	20802,05	72,39	2850,18	22,23	282195,95	46,01	4262,98	137,47	9687,17
Valor-p	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Fonte: Dados da Pesquisa.

Tabela 3: Testes de autocorrelação e heterocedasticidade condicional

	BGI		CCM		ETH		ICF		SOJ	
	ln	∇ ln	ln	∇ ln	ln	∇ ln	ln	∇ ln	ln	∇ ln
Ljung-Box	22112,99	64,57	5536,93	11,77	2581,76	24,51	23705,38	31,72	13234,17	12,01
Valor-p	0,0000	0,0000	0,0000	0,1084	0,0000	0,0009	0,0000	0,0001	0,0000	0,1505
Ljung-Box ²	22125,46	33,31	5535,72	5,88	2577,10	0,25	23702,60	1880,38	13240,93	69,53
Valor-p	0,0000	0,0001	0,0000	0,5536	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000
Box-Pierce	22061,63	64,47	5496,27	11,67	2544,40	24,11	23653,66	31,66	13183,04	11,98
Valor-p	0,0000	0,0000	0,0000	0,1118	0,0000	0,0011	0,0000	0,0001	0,0000	0,1522
Box-Pierce ²	22074,08	33,22	5495,07	5,85	2539,81	0,25	23650,90	1876,71	13189,78	69,23
Valor-p	0,0000	0,0001	0,0000	0,5572	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000
ML-ARCH	2778,64	29,64	799,39	3,90	390,65	0,35	2959,71	574,95	1663,96	37,55
Valor-p	0,0000	0,0002	0,0000	0,8657	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000
ln(T)	8,0000		7,0000		7,0000		8,0000		8,0000	

Fonte: Dados da Pesquisa



Fonte: Dados da Pesquisa.

Figura 1: Razões de variância

Têm-se indícios que os mercados de boi gordo, milho e etanol, por estarem fora do intervalo de confiança às razões de variância empíricas, não sejam eficientes, e nos mercados de café e soja há evidência à eficiência. Com base nessas razões parte-se para o cálculo dos testes de razões de variância.

Tabela 4: Razões de variância com *wild bootstrap* proposto por Kim (2006) ao teste de Lo & MacKinlay (1988) sob os retornos

Teste	k	$\nabla \ln$ BGI	$\nabla \ln$ CCM	$\nabla \ln$ ETH	$\nabla \ln$ ICF	$\nabla \ln$ SOJ
M_1	2		0,8716	0,4257		
Valor-p			0,5260	0,8740		
M_1	4		1,3638	1,3712		
Valor-p			0,2730	0,2410		
M_1	6		2,1066	2,3908		
Valor-p			0,0600	0,0140		
M_1	8		2,1325	-0,3178		
Valor-p			0,0200	0,8540		
M_2	2	4,6618	0,6823	0,2956	0,5510	1,4743
Valor-p		0,0000	0,5260	0,8740	0,6170	0,1210
M_2	4	5,3052	1,1130	1,1524	-0,0355	1,5394
Valor-p		0,0000	0,2730	0,2410	0,9750	0,1060
M_2	6	4,8412	1,8024	2,1311	0,0501	1,2554
Valor-p		0,0000	0,0600	0,0140	0,9640	0,1740
M_2	8	3,0594	1,9081	-0,2650	-0,1409	0,8765
Valor-p		0,0010	0,0200	0,8540	0,8980	0,3510

Fonte: Dados da Pesquisa.

Conforme demonstrado na Tabela 3, os retornos do boi gordo, café e soja apresentaram heterocedasticidade, logo, apenas os testes robustos à heterocedasticidade foram reportados. De acordo com os testes individuais de Lo & MacKinlay (1988) sob *wild bootstrap* de Kim (2006) expostos na Tabela 4, a hipótese nula de um passeio aleatório foi rejeitada para os retornos do boi gordo. Isso evidencia poder de previsão para horizontes de curto prazo, entre 2 e 8 períodos, invalidando a hipótese fraca de eficiência de mercado, representada pelo modelo de passeio aleatório 3 de Campbell et al. (1997). Invalidades às hipóteses também ocorrem nos mercados de milho e etanol, mas em horizontes mais longos, 6 e 8 dias. Pode-se inferir que os mercados de café e soja foram eficientes na forma fraca, nos períodos analisados, devido à não rejeição da hipótese de caminho aleatório, aos níveis tradicionais de significância estatística.

Possibilidades de previsões discrepantes, em horizontes diferentes, impossibilitaram a unicidade das inferências sobre a hipótese de caminho aleatório. Para contornar esse problema, parte-se ao teste conjunto entre todos os horizontes, proposto por Chow & Denning (1993).

Com base na metodologia múltipla do teste de Chow & Denning (1993) sob *wild bootstrap* de Kim (2006), a hipótese nula à razão de variância igual a 1 foi rejeitada nos mercados em que o teste de Lo & MacKinlay (1988) foi inconclusivo. Assim, os resultados dos testes individuais reforçam o resultado do teste múltiplo e indicam eficiência de mercado na forma fraca apenas aos mercados do café e da soja.

As evidências desfavoráveis à eficiência possuem implicações nos estudos sobre *hedge* desses mercados. De acordo com Rodrigues & Alves (2010), nos

Tabela 5: Razões de variância com *wild bootstrap* proposto por Kim (2006) ao teste de Chow & Denning (1993) sob os retornos

Teste	$\nabla \ln$ BGI	$\nabla \ln$ CCM	$\nabla \ln$ ETH	$\nabla \ln$ ICF	$\nabla \ln$ SOJ
CD_1		2,1325	2,3908		
Valor-p		0,0570	0,0350		
CD_2	5,3052	1,9081	2,1311	0,5510	1,5394
Valor-p	0,0000	0,0570	0,0350	0,8870	0,2330

Fonte: Dados da Pesquisa.

estudos empíricos brasileiros sobre razão e efetividade de *hedge* dos mercados agropecuários há predominância no uso da função de utilidade média variância como arcabouço, assim como baixa efetividade do *hedge*. Segundo demonstrações⁴ de Kroner & Sultan (1993), esse arcabouço pressupõe que os retornos nos mercados futuros sigam um processo martingal — por conseguinte, eficientes na forma fraca — para que a razão de *hedge* seja ótima. Assim, pode-se atribuir a baixa efetividade de *hedge* nos estudos empíricos brasileiros sobre os mercados agropecuários à não eficiência dos mercados futuros, o que causa a não otimalidade da razão de *hedge*.

O funcionamento informacional nos mercados cuja hipótese de passeio aleatório não foi rejeitada, e, portanto, eficientes na forma fraca, auxiliam os agentes desses mercados na descoberta de preços. Assim, esses mercados futuros eficientes contribuem para as tomadas de decisões sobre a produção, comercialização e estocagem (Garcia et al. 1988). Entretanto, nos mercados em que a hipótese nula de passeio aleatório foi rejeitada, os agentes estão sujeitos às informações errôneas indicadas nos mercados futuros dos ativos. Isso ocasiona incompletude informacional e, segundo Garcia et al. (1988), possível redução do excedente econômico, devido à possível alocação equivocada dos recursos desses agentes.

Conforme argumentos de Garcia et al. (1988), as diferentes características das commodities e de seus mercados influenciam a eficiência dos mercados futuros agropecuários. Nesse sentido, particularidades desses mercados, dispostas na Tabela 6, como liquidez, custos de transação e o poder de mercado dos participantes podem auxiliar na explicação das rejeições e não rejeições à hipótese nula.

Observa-se que o ativo com a maior média diária de contratos negociados foi o café, sendo que neste ativo as evidências obtidas com os testes de razão de variância indicaram não rejeição à hipótese de eficiência de mercado. Note que Amado & Carmona (2004) encontraram evidências favoráveis a eficiência do mercado de café e indicaram a liquidez como fator para esse resultado.

A discrepância entre a evidência sugerida nos testes de razão de variância no mercado de soja e sua liquidez indica que o número de contratos negociados não corresponde como fator explicativo, haja vista sua ínfima liquidez comparada a outros ativos. Note que no *survey* de Garcia & Leuthold (2004), os estudos no mercado norte-americano da soja indicam eficiência deste mercado. Portanto, um possível fator que pode explicar a eficiência nesse mercado

⁴Para maiores informações, ver a primeira nota de rodapé de Kroner & Sultan (1993), assim como as passagens das equações (5-6) e (9-10), as quais são simplificadas com base no pressuposto de o mercado futuro seguir um processo martingal.

Tabela 6: Características dos mercados

	BGI	CCM	ETH	ICF	SOJ
Liquidez					
Média diária de contratos negociados	755	519	45	996	87
% de log retornos = 0	3,22	3,94	9,62	1,64	12,47
Tipos de participantes dos contratos em aberto					(%)
Pessoa Jurídica Financeira	5,9578	6,8968	2,2898	4,7176	2,0967
Investidor Institucional	19,5215	3,0077	0,7255	10,2125	2,2898
Investidores Não Residentes	7,2892	1,9100	3,3574	25,2093	1,4342
Pessoa Jurídica Não Financeira	34,8054	72,1241	84,3073	51,9801	85,6676
Pessoa Física	32,4271	16,0614	9,3200	7,8804	8,5116
Coeficiente de Gini	0,3313	0,6139	0,6968	0,4474	0,6995

Fonte: Sistema de recuperação de dados da BM&FBOVESPA.

Nota: O período compreende o disposto na Tabela 1. A média diária de contratos negociados e o percentual de log retornos iguais a zero se referem ao primeiro vencimento. Devido à disponibilidade no sistema de recuperação de dados da BM&FBOVESPA, os percentuais de contratos em aberto por tipo de participante referem-se a todos os vencimentos.

brasileiro é a formação de seus preços ser dada nos mercados internacionais, repercutindo no mercado brasileiro.

Os contratos de boi gordo e milho apresentaram médias de contratos negociados inferiores ao do café, enquanto que os testes indicaram rejeição à eficiência. Ademais, liquidez discrepante no etanol e soja revela que esses foram os menos líquidos no período de análise, sendo que houve evidência não favorável à eficiência no mercado da soja. Assim, apesar de Garcia et al. (1988), Amado & Carmona (2004), Gilson & Kraakman (2014), entre outros, indicarem a liquidez dos contratos para que esses sejam informacionalmente eficientes, possíveis explicações das não rejeições e rejeições à hipótese de eficiência de mercado não podem ser centradas apenas nesse indicativo.

Como consequência da não eficiência nos mercados futuros agropecuários brasileiros, os preços não respondem tão rapidamente às informações públicas contidas em seus preços passados, de forma que os investidores podem obter lucros por arbitragem ao negociar tal informação (Gilson & Kraakman 2014). Dois mecanismos podem explicar tal resposta às informações para os autores. Primeiro, pode-se considerar que os profissionais desses mercados não compreendem tão rapidamente as novas informações em um curto espaço de tempo, quando comparado aos mercados eficientes. Segundo, nos mercados eficientes, considera-se que praticamente todos os profissionais atuantes no mercado possuem aprendizado às novas informações quase simultaneamente. No entanto, a velocidade com que as informações refletem nos preços é função da liquidez do mercado. Os autores notam que esses mecanismos informacionais sob os preços dependem dos custos de obtenção das informações e dos custos de arbitragem – isto é – do custo de negociar tais informações.

Segundo Lesmond et al. (1999), um ativo com elevados custos de transação tem movimentos dos preços menos frequentes e maior frequência de retornos iguais a zero comparado a um ativo com baixos custos de transação. Utilizando-se o percentual de retornos iguais a zero, *proxy* sugerida pelos autores para estimar os custos de transação, observam-se elevados custos de transação nos mercados de etanol e de soja ao serem comparados aos demais. Assim como os custos de transação, a baixa liquidez dos contratos dificulta a

operacionalização das possíveis arbitragens existentes nos mercados não eficientes. Além disso, conforme McKenzie et al. (2002), o baixo nível de negócios dificulta a potencial descoberta de preços aos agentes do sistema de comercialização.

Assim como no mercado da soja, os mercados de etanol e milho apresentaram maior concentração dos contratos em aberto por tipo de participantes. Essa concentração, mensurada pelo coeficiente de Gini, possibilita inferir que há poder de mercado nas negociações desses ativos, em particular das pessoas jurídicas não financeiras. Ao contrário do que ocorre nos mercados de café e boi gordo, a não atração dos diversos tipos de participantes nesses mercados, dificulta a eficiência informacional nesses mercados. Ressalta-se que apesar de a menor concentração dos participantes no mercado de boi gordo ser favorável à eficiência desse mercado, Urso (2007) sugere existência de poder de mercado na aquisição de bois pelos frigoríficos no mercado à vista. Segunda a autora, a estrutura dessa indústria é de uma produção pecuária pulverizada, um número elevado de frigoríficos, mas que já apresenta sinais de concentração, e passa a se estruturar como um oligopólio. Os resultados encontrados em seu estudo corroboraram a visão de que os frigoríficos — em sua maioria associados a pessoas jurídicas não financeiras — têm mais informação, no mercado futuro, que os demais agentes.

Os resultados não condizentes com a eficiência dos mercados de milho e etanol também podem ser creditados às intervenções governamentais na estrutura de preços desses mercados. As intervenções governamentais, sugeridas como fator de influência ao bom funcionamento dos mercados no *survey* de Garcia et al. (1988), e utilizadas como justificativa no estudo sobre a eficiência do mercado de álcool anidro por Alves et al. (2008), têm influência adversa à eficiência informacional. Assim, devem-se considerar atenuações no grau dessas intervenções como forma de obter eficiência alocativa de recursos aos agentes dessa cadeia por meio do mercado futuro.

5 Considerações finais

Os resultados demonstraram que, no intuito de retornos especulativos, por mais que se tentem fazer previsões para essas séries, a evidência a favor de eficiência dos mercados de café e da soja indica que são imprevisíveis.

Os resultados da rejeição à hipótese de passeio aleatório aos retornos sugerem presença de predicabilidade nos mercados futuros agropecuários de boi gordo, milho e etanol. Refutá-la indica que os investimentos em ativos desses mercados podem proporcionar a agentes — investidores, arbitradores, especuladores e gestores financeiros — retornos superiores à média do mercado.

Desses resultados, as imperfeições existentes podem ocasionar a não proteção necessária a *hedgers* ao utilizarem contratos futuros que não refletem todas as informações existentes no mercado em seus preços passados. Isso dificulta a operacionalização de estratégias, com objetivo de redução de risco, embasadas na historicidade de suas cotações. A não eficiência fraca desses mercados pode explicar a baixa efetividade do *hedge* nesses ativos.

Do lado de eficiência alocativa necessária a agentes que fazem uso dos contratos nos mercados de *commodities* fracamente não eficientes, os mecanismos de apreçamento não asseguram alocação eficiente dos recursos aos agentes, com efeitos negativos para a cadeia dependente desse ativo. A ineficiência

evidenciada pode levar os membros que regulamentam e normatizam esses contratos negociados na BM&FBOVESPA a tomarem medidas necessárias à sua correção e que elevem o volume negociado desses ativos, indispensável para que ocorra maior aleatoriedade, com possível redesenho contratual e formadores de mercado.

A rejeição da hipótese de eficiência nesses mercados oferece evidência de que ainda são ineficientes, e não uma evidência que o arcabouço teórico, no qual a hipótese se baseia, pode estar equivocado.

Agradecimentos

Os autores agradecem aos comentários e sugestões do editor, dos pareceristas anônimos, assim como de: Alexandre Florindo Alves, José Cesar Cruz Júnior, Fabio Lanhoso de Mattos, Márcio Poletti Laurini e Rodrigo Lanna Franco da Silveira.

Referências Bibliográficas

- Alves, J. S., Duarte, G. S. & Lima, R. C. (2008), 'Teste da eficiência do mercado futuro do álcool anidro no Brasil: uma análise de co-integração', *Revista Econômica do Nordeste* 39(1), 173–184.
- Amado, C. F. P. & Carmona, C. U. M. (2004), 'Uma análise da eficiência dos mercados futuros agrícolas brasileiros', IV Encontro Brasileiro de Finanças, Rio de Janeiro.
- Bitencourt, W. A. (2007), 'Ensaio empírico sobre a eficiência do mercado futuro de café', Departamento de Administração e Economia, Universidade Federal de Lavras, Mestrado em administração.
- Box, G. E. P. & Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American Statistical Association* 65(332), 1509–1526.
- Bressan, A. A. & Leite, C. A. M. (2001), 'Eficiência do mercado futuro de café no Brasil', *Reuna* 6(1), 11–32.
- Campbell, J. Y., Lo, A. W. & MacKinlay, A. C. (1997), *The econometrics of financial markets*, Princeton University Press, Princeton.
- Charles, A. & Darné, O. (2009), 'Variance-ratio tests of random walk: an overview', *Journal of Economic Surveys* 23(3), 503–527.
- Chen, W. W. & Deo, R. S. (2006), 'The variance ratio statistic at large horizons', *Econometric Theory* 22(2), 206–234.
- Chow, K. & Denning, K. C. (1993), 'A simple multiple variance ratio test', *Journal of Econometrics* 58(3), 385–401.
- Cruz Júnior, J. C. & Silveira, R. L. F. (2007), 'Análise de eficiência, co-integração e exogeneidade nos mercados futuros de café na BM&F, NYB e LIFFE', XLV Congresso Brasileiro de Economia e Sociologia Rural.

- Dickey, D. A. & Fuller, W. A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**(366), 427–431.
- Dickey, D. A. & Fuller, W. A. (1981), 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica* **49**(4), 1057–1072.
- Duarte, G. B., Lima, R. C. & Alves, J. S. (2007), 'Co-integração e eficiência do mercado futuro da soja no Brasil', Congresso Brasileiro de Economia e Sociologia Rural, Londrina.
- Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**(4), 987–1007.
- Fama (1970), 'Efficient capital markets: a review of theory and empirical work', *The Journal of Finance* **25**(2), 383–417.
- Fama, E. F. (1965), 'The behavior of stock-market prices', *The Journal of Business* **38**(1), 34–105.
- Fraga, G. J. & Silva Neto, W. A. (2011), 'Eficiência no mercado futuro de commodity: evidências empíricas', *Revista Econômica do Nordeste* **42**(1), 125–137.
- Garcia, P., Hudson, M. A. & Waller, M. L. (1988), 'The pricing efficiency of agricultural futures markets: an analysis of previous research results', *Southern Journal of Agricultural Economics* **20**(1), 119–130.
- Garcia, P. & Leuthold, R. M. (2004), 'A selected review of agricultural commodity futures and options markets', *European Review of Agricultural Economics* **31**(3), 235–272.
- Gilson, R. J. & Kraakman, R. (2014), 'Market efficiency after the financial crisis: it's still a matter of information costs', *Virginia Law Review* **100**(2), 313–375.
- Granger, C. W. J. & Morgenstern, O. (1963), 'Spectral analysis of New York stock market prices', *Kyklos* **16**(1), 1–27.
- Hakkio, C. (1986), 'Does the exchange rate follow a random walk? a Monte Carlo study of four tests for a random walk', *Journal of International Money and Finance* **5**(2), 221–229.
- Jarque, C. M. & Bera, A. K. (1980), 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals', *Economics Letters* **6**(3), 255–259.
- Jarrow, R. A. & Larsson, M. (2012), 'The meaning of market efficiency', *Mathematical Finance* **22**(1), 1–30.
- Kim, J. H. (2006), 'Wild bootstrapping variance ratio tests', *Economics Letters* **92**(1), 38–43.
- Kroner, K. F. & Sultan, J. (1993), 'Time-varying distributions and dynamic hedging with foreign currency futures', *The Journal of Financial and Quantitative Analysis* **28**(4), 535–551.

Lesmond, D., Ogden, J. & Trzcinka, C. (1999), 'A new estimate of transaction costs', *Review of Financial Studies* 12(5), 1113–1141.

Lim, K. (2009), 'An empirical analysis of the weak-form efficiency of stock markets', Department of Econometrics and Business Statistics, Monash University, Doctor of Philosophy.

Lim, K. & Brooks, R. (2011), 'The evolution of stock market efficiency over time: a survey of the empirical literature', *Journal of Economic Surveys* 25(1), 69–10.

Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* 65(2), 297–303.

Lo, A. W. & MacKinlay, A. C. (1988), 'Stock market prices do not follow random walks: evidence from a simple specification test', *Review of Financial Studies* 1(1), 41–66.

McKenzie, A. M., Jiang, B., Djunaidi, H., Hoffman, L. A. & Wailes, E. J. (2002), 'Unbiasedness and market efficiency tests of the u.s. rice futures market', *Review of Agricultural Economics* 24(2), 474–493.

Moraes, A. S., Lima, R. C. & de Souza Melo, A. (2009), 'Análise da eficiência do mercado futuro brasileiro de boi gordo usando co-integração', *Revista de Economia e Sociologia Rural* 47(3), 601–614.

Park, C.-H. & Irwin, S. H. (2004), 'The profitability of technical analysis: a review', AgMAS Project Research Reports, n. 04.

Righi, M. B. & Ceretta, P. S. (2011), 'Previsibilidade e eficiência no mercado agrícola', *Ciência Rural* 41(10), 1844–1850.

Rodrigues, M. A. & Alves, A. F. (2010), 'Efetividade e razão ótima de hedge: um survey', XLVIII Congresso Brasileiro de Economia e Sociologia Rural, Campo Grande.

Samuelson, P. A. (1965), 'Proof that properly anticipated prices fluctuate randomly', *Industrial Management Review* 5(2), 41–49.

Silva Neto, W. A., Fraga, G. J. & Marques, P. V. (2010), 'Eficiência de mercado: evidências empíricas para os preços spot e futuro de boi gordo', *Revista de Economia* 36(3), 7–24.

Tsay, R. S. (2005), *Analysis of financial time series*, 2 edn, John Wiley & Sons.

Urso, F. S. P. (2007), 'A cadeia da carne bovina no Brasil: uma análise de poder de mercado e teoria da informação', Escola de Economia de São Paulo, Fundação Getúlio Vargas - Doutorado em Economia de Empresas.

Working, H. (1962), 'New concepts concerning futures markets and prices', *The American Economic Review* 52(3), 431–459.

Wright, J. H. (2000), 'Alternative variance-ratio tests using ranks and signs', *Journal of Business & Economic Statistics* 18(1), 1–9.

DESIGUALDADE DE OPORTUNIDADES NO ACESSO AO ENSINO SUPERIOR NO BRASIL: UMA COMPARAÇÃO ENTRE 2003 E 2013

MÁRCIA MARQUES DE CARVALHO *

FÁBIO D. WALTENBERG †

Resumo

Apesar dos avanços no ensino fundamental e médio nas últimas décadas, e da ampliação de vagas no ensino superior, apenas 12% dos jovens em idade universitária estão cursando o ensino superior no Brasil, contra 21% na Argentina, 65% nos EUA e 70% na Suécia. Neste estudo, a partir dos dados mais recentes da Pnad disponíveis (2013) e dos dados de uma década antes (2003, pouco antes do advento de políticas de ação afirmativa), inicialmente traçam-se perfis de grupos vulneráveis e não vulneráveis no acesso ao ensino superior. Depois, com base no índice de oportunidades humanas (Barros et al. 2009), e fazendo-se uso de regressões de variável dependente binária, mensuramos a desigualdade de oportunidades no acesso ao ensino superior para os dois anos. Os resultados indicam que houve realocação de oportunidades dos grupos não vulneráveis para os grupos vulneráveis, e que o acesso ao ensino superior também aumentou levemente no espaço de uma década. Contudo, o índice de oportunidades humanas ao acesso ao ensino superior no Brasil ($IOH = 0,281$ em 2013) está muito longe da situação ideal ($IOH=1$), que ocorre quando o acesso é universal e não há desigualdade de oportunidades no acesso. Por fim, aplicou-se aos indicadores obtidos uma decomposição por fator (de Shapley), cujos principais resultados são a importância da instrução do chefe e da renda domiciliar como circunstâncias limitadoras, e também uma contribuição moderada da cor/raça para a desigualdade de oportunidades no acesso ao ensino superior no Brasil.

Palavras-chave: Índice de Oportunidades Humanas; Regressão Logística; Decomposição de Shapley.

Abstract

* Professora do Departamento de Estatística da Universidade Federal Fluminense (UFF). Pesquisadora do centro de Estudos sobre Desigualdade e Desenvolvimento (CEDE). Niterói, Rio de Janeiro, Brasil. E-mail: mcarvalho@est.uff.br

† Professor do Departamento de Economia da Universidade Federal Fluminense (UFF). Pesquisador do centro de Estudos sobre Desigualdade e Desenvolvimento (CEDE). Niterói, Rio de Janeiro, Brasil. E-mail: waltenberg@economia.uff.br

In spite of overall improvements in basic and high school education levels in the last decades, and of an increase in the supply of higher education, only 12% of Brazilian youths in the expected age are attending higher education in Brazil, against 21% in Argentina, 65% in the USA and 70% in Sweden. In this study, based on the most recent Pnad data (2013) and on data from ten years earlier (2003, just before the implementation of affirmative action policies), we obtained profiles of vulnerable and non-vulnerable groups in terms of access to higher education in Brazil. Then through the index of human opportunities (Barros et al. 2009), and making use of binary dependent variable regressions, we were able to measure inequality of opportunity in terms of access to higher education for both years. Results indicate that a reallocation of opportunities from non-vulnerable to vulnerable groups has taken place, and that access to higher education has also risen slightly in the course of one decade. However, the index of human opportunity as applied to higher education in Brazil (HOI = 0.281 in 2013) is very far from the ideal situation (HOI=1), which occurs when access is universal and there is no inequality of opportunity in the access. Finally, we have applied a Shapley decomposition to the calculated indices, and the main results are the substantial contribution of household-head's education level and the household's income to inequality of opportunity, and also a moderate impact of race.

Keywords: Human Opportunity Index; Logistic Regression; Shapley Decomposition.

JEL classification: C43, D30, D63

DOI: <http://dx.doi.org/10.1590/1413-8050/ea124777>

1 Introdução

Estudar as dificuldades e os gargalos no acesso ao ensino superior no Brasil é relevante por vários motivos. O primeiro é a nossa defasagem de pessoas com esse nível de ensino, visto que, ao contrário de muitos países, o Brasil tem menos de 20% da sua população adulta com esta qualificação. Isso afeta a produtividade e a capacidade da economia brasileira em adquirir e desenvolver novas tecnologias. Pessoas mais instruídas ganham mais e estão menos expostas ao desemprego, elementos que contribuem para uma maior qualidade de vida para elas e sua família.

Como é amplamente sabido, a educação afeta diversas dimensões da vida social e econômica de um país, além de produzir benefícios que são privadamente apropriados pelos indivíduos, tais como habilidades, atitudes e valores (Barr 2012). Há diversas evidências que mostram que quanto mais se investe em educação, além dos efeitos diretos positivos na economia do país, maior é o retorno à sociedade em termos de bem-estar, redução das desigualdades e das taxas de fecundidade e mortalidade (Mendonça 2000). No caso da educação superior, os ganhos são mais elevados e os impactos no mercado de trabalho e na capacidade de absorção de inovação tecnológica e na produtividade são ainda mais expressivos.

No Brasil, a conclusão de um curso de graduação é acompanhada por uma baixa taxa de desemprego e por um retorno financeiro que, em média, é 2,6 vezes maior do que os obtidos por aqueles que pararam os estudos no ensino médio (Carvalho 2011). Em termos comparativos esse retorno do ensino superior é alto quando comparado com o observado nos países desenvolvidos: nos EUA é de 1,77, na Alemanha, 1,67; no Reino Unido, 1,54, na França, 1,50 e no Canadá, 1,42, segundo dados da OCDE (2011). Esse alto retorno no Brasil se deve ao fato de ser pequena a proporção da população adulta brasileira (25 a 64 anos) que completou o ensino superior, limitando, dessa forma, a oferta dessa mão-de-obra. Segundo dados da Pnad/2011, apenas 11% da população adulta brasileira tinham curso de graduação, enquanto no Canadá essa proporção chega a 49%, nos Estados Unidos, 41%, no Japão, 43%, no Reino Unido, 33% e na Alemanha, 25%.

O momento é oportuno para se estudar o acesso ao ensino superior por dois motivos. O primeiro é no sentido de se compreender melhor as políticas de ação afirmativa no acesso ao ensino superior, uma vez que muitas instituições públicas passaram a adotar políticas deste tipo ao longo da década passada e agora as instituições federais são obrigadas a reservarem uma proporção das suas vagas para egressos do ensino médio público com combinações de renda familiar e cor/raça (Lei nº 12.711 de 29 de agosto de 2012, doravante chamada de “Lei das Cotas”). O segundo é de aproveitar o momento do bônus demográfico brasileiro. Perdemos a primeira “onda” da população de 18 a 24 anos que ocorreu entre 2000 e 2010 e causou um aumento de dois milhões de pessoas deste grupo etário em apenas uma década. A próxima onda desse grupo etário irá ocorrer entre 2016 e 2026. Estamos neste momento (2015) passando pela onda do grupo etário de 25 a 34 anos que terá o seu ápice em 2016 e terminando em 2020.

Outra característica do ensino superior no Brasil é a pouca diversidade socioeconômica entre os estudantes. De fato, embora entre 2006 e 2008, 85% dos concluintes do ensino médio proviessem do ensino médio público, dos indivíduos que ingressaram nos cursos de graduação no Brasil nesse período,

essa proporção cai para somente 57% (MEC/Inep). Na mesma linha, em 2009, enquanto 45% das pessoas com ensino médio completo provinham de famílias relativamente pobres (com renda familiar de até 3 salários mínimos), entre os ingressantes do ensino superior, essa proporção caía para 39%. Considerando apenas as pessoas com ensino médio completo, 50,3% se declararam não brancas enquanto entre os ingressantes dos cursos de graduação a incidência desse grupo era de apenas 36,4%.

De acordo com a teoria de igualdade de oportunidades do economista Roemer (1998), muito em voga atualmente (Fleurbaey 2008, Ferreira & Gignoux 2014, Pignataro 2012, Ramos & Van De Gaer 2012), quando existe subrepresentação por parte de um grupo socioeconômico, definido pela sociedade como relevante e legítimo, no acesso a um serviço ou vantagem — como ocorre com o acesso de certos grupos ao ensino superior no Brasil — pode-se estar diante de um problema de desigualdade de oportunidades, se porventura a dificuldade de obter acesso ao serviço ou vantagem tiver sido causada, sobretudo, por circunstâncias desfavoráveis.

No caso do ensino superior, uma tentativa de mitigar o problema de acesso limitado de certos grupos consiste na aplicação de políticas de ação afirmativa. As ações afirmativas podem ser compreendidas como programas que buscam prover oportunidades ou outros benefícios para pessoas pertencentes a grupos específicos, alvo de discriminação ou com pouco acesso a recursos (IPEA 2008). Têm sido aplicadas em inúmeros países e em diferentes etapas da educação, bem como no mercado de trabalho. No Brasil, têm se concentrado no acesso aos cursos de graduação, por meio de diferentes instrumentos: cotas e bônus, ditas “raciais” ou “sociais”. As “raciais” utilizam como critério a cor/raça do aluno, de acordo com auto-declaração. Os critérios “sociais” baseiam-se na renda familiar que comprove carência ou no fato de o aluno ser oriundo do ensino médio público, ou seja, de escolas municipais, estaduais ou federais ou de cursos supletivos presenciais de educação de jovens e adultos. Há casos em que ambos os critérios são considerados simultaneamente, quando vagas são reservadas, por exemplo, a alunos negros pobres.

No dia 26 de abril de 2012 o Supremo Tribunal Federal (STF) considerou constitucional a política de reservas de vagas para negros e índios para garantir o acesso destes grupos nas universidades públicas. A “Lei das Cotas” regulamentou as cotas nas universidades federais em todo o país e reserva um percentual das vagas oferecidas em cada curso de graduação e turno para alunos que estudaram todo o ensino médio em escolas públicas. Esse percentual deveria ser de 12,5% das vagas em 2013; 25% em 2014, 37,5% em 2015 e 50% em 2016. Dessas vagas reservadas, a metade deve ser destinada a estudantes com renda familiar *per capita* de até 1,5 salário-mínimo e a outra metade a estudantes com renda superior a 1,5 salário-mínimo. Dentro de cada estrato de renda, será aplicada a porcentagem de cotas para pretos, pardos e indígenas observada em cada Estado e será definida pelo peso de cada uma dessas populações segundo o mais recente Censo do Instituto Brasileiro de Geografia e Estatística (IBGE). Cabe destacar que os percentuais definidos pela Lei não correspondem à representação proporcional dos grupos na população. Por exemplo, o percentual de vagas reservadas para egressos da escola pública em 2011 era de 50% enquanto o Censo da Educação Básica do Ministério da Educação e Cultura mostrou que 85% dos concluintes do ensino médio de todo o país eram de escolas públicas. Enquanto a Lei das Cotas reserva 50% das vagas para alunos egressos do ensino médio público com renda

domiciliar *per capita* inferior a 1,5 sm, a Pesquisa Nacional por Amostra de Domicílios (Pnad) do IBGE mostra que esse grupo representa 86,7% da demanda potencial ao ensino superior. Como existe uma relação forte entre a renda domiciliar *per capita* e a cor/raça autodeclarada, dentre os estudantes do ensino médio público com renda inferior a 1,5 sm, 60% eram não brancos; e entre os estudantes do ensino médio público com renda superior a 1,5 sm, apenas 40% se autodeclararam não brancos.¹

Uma política de ação afirmativa deve estar baseada na precisa identificação de segmentos sociais que, por desvantagens socioeconômicas, estão privados ou tem seu acesso limitado a um bem ou serviço. Em especial, quando se sabe que os retornos à educação superior são particularmente elevados, é de se questionar por que no Brasil, apesar desse enorme retorno em termos de aumento de salário e maior empregabilidade, muitos brasileiros não possuem este nível de ensino? Na década de 1980 e 1990 o gargalo poderia ser a insuficiência de vagas, mas entre 1999 e 2011, enquanto as vagas ofertadas nas instituições de ensino superior no país aumentaram 3,6 vezes, o número de ingressos no mesmo período aumentou 2,3 vezes (INEP 2011).

O presente trabalho traça um detalhado perfil dos matriculados dos cursos de graduação públicos e privados no Brasil a partir dos microdados da Pnad e identifica os grupos sociais mais vulneráveis ao acesso ao ensino superior e os menos vulneráveis. Em seguida, analisa quais são as características ou circunstâncias mais correlacionadas ao ingresso no ensino superior público e privado.

Uma vez identificados os fatores que restringem o acesso à educação superior e os segmentos sociais mais direta e intensamente afetados por essas restrições, ficarão evidentes para o leitor os elementos que poderiam justificar a adoção de políticas de ação afirmativa. Uma crítica que alguns pesquisadores têm feito com relação ao acesso à necessidade de se ampliar o acesso ao ensino superior é a de que nem todos os jovens que concluem o ensino médio gostariam de avançar os estudos e cursar a graduação. Uma pesquisa realizada em 2007 pelo Ipea em parceria com o Pnud e Cepal com jovens de 18 a 24 anos dos estados de São Paulo e Rio de Janeiro indagou que nível de ensino o jovem gostaria de alcançar. A resposta foi surpreendente: 77% dos jovens responderam que gostariam de alcançar o ensino superior, sendo que 48% deles ficariam satisfeitos com a graduação; 17% gostariam de cursar mestrado e 12% almejavam o doutoramento.

Para analisar a evolução da igualdade de oportunidades no acesso ao ensino superior no Brasil são utilizados os microdados da Pnad de 2013 (mais recente disponível) e também da Pnad de 2003 (antes da política de ações afirmativas no acesso ao ensino superior público). Com esses dados é possível mensurar a evolução, ao longo de um período de exatos dez anos, da taxa de cobertura do ensino superior e também a distribuição do acesso às vagas segundo as características dos alunos (circunstâncias) através do índice de dissimilaridade. Esses dois componentes multiplicados formam o Índice de Oportunidades Humanas (IOH), desenvolvido por Barros et al. (2009) para mensurar o acesso de crianças a um conjunto básico de bens e serviços, conforme proposto pelo Banco Mundial no seu Relatório de Desenvolvimento Mundial de 2006. Segundo o relatório, desigualdades de oportunidades cau-

¹O apêndice A apresenta um desenho esquemático com a distribuição percentual das vagas segundo a “Lei das Cotas” e a encontrada nos dados disponíveis.

sam privações e têm como consequência desperdício de potencial humano. O método de decomposição de Shapley será aplicado ao índice de dissimilaridade para analisar os fatores mais associados à desigualdade no acesso ao ensino superior.

Tendo em vista os objetivos detalhados acima, o presente trabalho está organizado em três seções, além desta introdução e da conclusão. A segunda seção é conceitual e metodológica e apresenta a teoria da igualdade de oportunidades de Roemer. Além disso, traz os índices que tentam medir essa desigualdade. A seção traz também a metodologia da decomposição de Shapley. A seção 3 é empírica e apresenta uma análise quantitativa da desigualdade de oportunidades no acesso ao ensino superior, calculando os índices descritos no capítulo metodológico. Os resultados apontam que houve um aumento na cobertura do ensino superior público e privado entre 2003 e 2013. Houve também um pequeno avanço em termos de distribuição destas vagas entre os grupos sociais: o índice de dissimilaridade caiu de 0,155 em 2003 para 0,108 em 2013, o que significa que a distribuição das vagas está ligeiramente mais homogênea entre os grupos.

2 O índice de desigualdade de oportunidades e a decomposição de Shapley

Segundo Peragine (2011), a teoria da igualdade de oportunidades foi formulada na filosofia política por Dworkin (1981), Arneson (1989) e Cohen (1989) e traduzida em modelos econômicos formais por Roemer (1993, 1998) e Fleurbaey (2008).

A teoria da igualdade de oportunidades baseia-se em três tipos de informação:

- Resultados: realizações educacionais, futuro rendimento, acesso a serviços básicos como água, luz e esgoto ou qualquer outra variável que expresse resultado;
- Circunstâncias: variáveis além da responsabilidade dos indivíduos (fora de seu controle) e que afetam a capacidade de atingir o resultado. Ex: gênero, etnia, local de nascimento, educação dos pais;
- Responsabilidades ou Esforço: variáveis pertencentes à esfera da responsabilidade dos indivíduos (esforço ou características de responsabilidade), em geral não observadas. Exemplos: esforça-se mais ou menos na escola, dedica-se mais ou menos tempo às tarefas escolares, trabalha mais ou menos horas etc.

A ideia básica da teoria é que as desigualdades determinadas pelo esforço são consideradas justas e somente as desigualdades resultantes das circunstâncias são consideradas eticamente inaceitáveis.² Para ilustrar o que seria uma desigualdade injusta, isto é, devido às circunstâncias, vejamos o exemplo a seguir.

²A contribuição das variáveis não observáveis em geral acaba sendo creditada à conta do “esforço”, não por assim determinar a teoria, mas sim por falta de alternativa metodológica melhor. Por essa razão, a repartição entre contribuição de “esforço” e de “circunstância” sempre merece ser interpretada com a ressalva de que as contribuições das circunstâncias estão subestimadas — subestimada estará, portanto, a desigualdade de oportunidades reportada.

Considere indivíduos de dois tipos no problema do acesso à educação superior no Brasil:³

- Tipo 1 consiste em indivíduos negros, morando na favela, com mãe chefe e vários irmãos, cujos pais não concluíram o ensino médio. Suponha que o nível de esforço do tipo 1 esteja entre 1 e 7, numa escala de 0 a 10, com mediana 2,5 (c_1).
- Tipo 2 consiste de indivíduos de classe média alta, morando com pai e mãe, com mais de 1 irmão e ambos os pais com ensino superior. Suponha que o nível de esforço do tipo 2 esteja entre 3 e 8 com mediana 5 (c_2).

Suponha que João, do tipo 1, exerceu esforço de nível 5, o que o coloca no 90º centil da distribuição de seu tipo (Figura 1, distribuição do tipo 1, $F(\frac{y}{c_1, e_{90}})$). Maria, do tipo 2, exerceu esforço de nível 7, que também corresponde ao 90º centil da distribuição do seu tipo (Figura 1, distribuição do tipo 2, $G(\frac{y}{c_2, e_{90}})$).

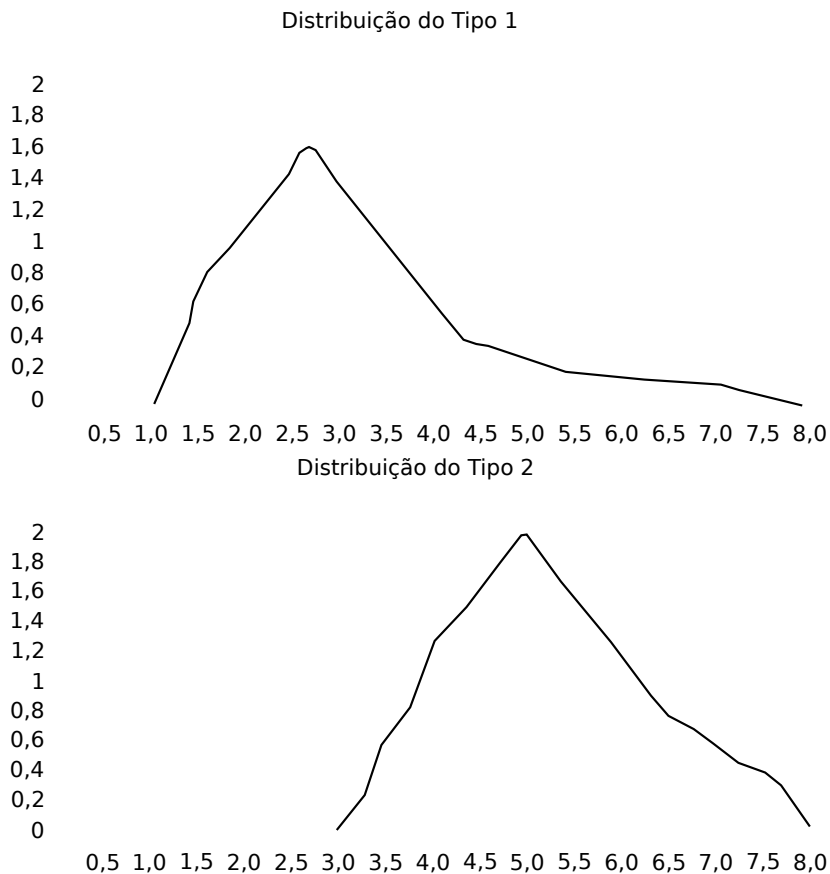


Figura 1: Distribuição dos Resultados das pessoas do tipo 1 e do tipo 2

Suponha também que a nota mínima para passar no vestibular e ingressar na universidade seja 6. Neste caso, João não ingressaria na universidade

³Estamos supondo uma relação não decrescente entre esforço e resultado.

porque obteve 5, e Maria ingressaria com sua nota 7. Entretanto, ambos exerceram o mesmo nível relativo de esforço e ficaram entre os 10% das melhores notas de seus tipos respectivos.

Roemer considera que, sob uma política de igualdade de oportunidades, os indivíduos que tenham exercido grau de esforço igual devem terminar com o mesmo resultado, ou seja:

$$F\left(\frac{y}{c1, e_{90}}\right) = G\left(\frac{y}{c2, e_{90}}\right) \quad (1)$$

A proposta de Roemer não implica que os resultados sejam igualados para todos. Ele propõe “nivelar o terreno”, o que significa garantir que aqueles que aplicaram igual esforço terminem com igual realização, independente de suas circunstâncias. No ensino superior, as notas que medem o desempenho no vestibular dependem não só do esforço do indivíduo, mas também de outros fatores tais como a capacidade cognitiva. Esta, por sua vez, depende de fatores hereditários, da situação da família, da criação recebida na infância, da qualidade do ensino recebido, entre outras circunstâncias. As cotas ou reservas de vagas consideram aptos para admissão no ensino superior os indivíduos que mais se esforçaram dadas suas circunstâncias. Dessa forma justiça seria realizada, porque seriam tratados igualmente pessoas com antecedentes e características semelhantes. Em suma, o que Roemer propõe é uma meritocracia condicional ou uma meritocracia entre iguais. Quando consideramos somente as notas (mérito), segundo a teoria de Roemer, não estamos sendo justos porque estamos tratando igualmente pessoas diferentes, com antecedentes e características distintas. Esse é o princípio da compensação, segundo o qual a desigualdade de resultados derivada por diferentes dotações iniciais deve ser eticamente inaceitável e compensada pela política pública.

Roemer discute o escopo e a extensão do conceito de igualdade de oportunidades. Ele propõe que o conceito deva ser aplicado quando a oportunidade em questão é a aquisição de um atributo necessário para concorrer a um cargo (emprego ou carreira). Nesse contexto a igualdade de oportunidades no acesso ao ensino superior cabe, porque o acesso ao ensino superior no Brasil significa acesso a uma ocupação (médico, advogado, engenheiro) e ter uma educação superior é uma vantagem necessária para competir por determinados cargos. Roemer lembra, entretanto, que uma oportunidade é um acesso a uma vantagem (frase de Cohen 1989) mas o indivíduo é responsável em transformar esse acesso em vantagem pela aplicação do esforço.

Uma vez apresentado o modelo teórico de Roemer, o desafio agora é traduzir a teoria em medida. Pesquisas recentes têm procurado quantificar a desigualdade de oportunidades de um resultado, seja ele expresso em categorias, como os níveis educacionais e o acesso ou não a um bem ou serviço; ou expresso por uma variável contínua, como renda ou desempenho acadêmico.

Com relação à mensuração da desigualdade de oportunidades de variáveis categóricas, depois de Barros et al. (2009), outras propostas foram desenvolvidas. Podemos destacar as contribuições dos seguintes autores:

- Pignataro (2012) destaca que, no caso do acesso a um serviço ou a um nível educacional específico, deve-se utilizar o índice de oportunidades humanas (IOH) desenvolvido por Barros et al. (2009), uma vez que as categorias não envolvem um ordenamento;

- Hoyos & Narayan (2012) também utilizam o IOH para mensurar a desigualdade de oportunidades no acesso à educação e à saúde (vacinação) em 47 países. Além de mensurar a desigualdade de oportunidades, eles examinaram a contribuição de uma circunstância, a saber o gênero da criança, na desigualdade no acesso a esses serviços. Eles concluíram que, para a maioria dos países, a contribuição do sexo da criança para a frequência escolar e vacinação tende a ser baixa ou muito abaixo da contribuição de outras circunstâncias como status socioeconômico e a localização do domicílio (área urbana ou rural).

No Brasil, além do pioneiro trabalho de Barros et al. (2009), podemos destacar os trabalhos de:

- Dill & Gonçalves (2012) analisam a igualdade de oportunidade no acesso a serviços como energia elétrica, água limpa, saneamento e escola, na linha de Barros et al. (2009), e sua evolução no Brasil, nos anos de 1999 e 2009, utilizando a Pnad. As características utilizadas como circunstâncias foram: gênero, raça, área de residência, gênero da pessoa de referência do domicílio, presença da mãe no domicílio, educação da pessoa de referência do domicílio, renda mensal domiciliar *per capita* e número de pessoas que vivem no domicílio.
- Diaz (2012) mensura a desigualdade de oportunidades de desempenho no ensino médio no Brasil a partir dos microdados do Saeb para o período 1995 a 2005. Utilizando a metodologia de Ferreira & Gignoux (2008), que compara a desigualdade existente na distribuição alisada com a desigualdade na distribuição original, captando a desigualdade entre os grupos, os resultados indicam que a desigualdade de oportunidades foram menores em Língua Portuguesa (16% da desigualdade total, média do período) do que em Matemática (24% da desigualdade total, média do período).
- Carvalho et al. (2012) mensuram a igualdade de oportunidades educacionais utilizando os dados do Pisa, além de desenvolverem uma técnica alternativa e multidimensional que considera acesso, dissimilaridade na distribuição do acesso e desempenho dos estudantes.

2.1 O Índice de Oportunidades Humanas (IOH)

Após essa breve resenha da literatura sobre a mensuração da igualdade de oportunidades, cabe ressaltar que essa metodologia foi desenvolvida muito recentemente (desde 2005), principalmente a que trata de resultados categóricos (desde 2009).

Como o objetivo desse trabalho é mensurar o grau de desigualdade de oportunidades no acesso ao ensino superior (ou seja, uma variável binária) será calculado o índice de oportunidades humanas (IOH), desenvolvido por Barros et al. (2009), que traduz a teoria de Roemer a resultados binários, mas que não foram utilizados no Brasil para acesso ao ensino superior. O IOH combina dois elementos:

- O nível de cobertura do acesso ao bem ou serviço em estudo na sociedade (\hat{p});

- O grau em que o acesso a esse bem ou serviço depende de variáveis que não são de responsabilidade do indivíduo (circunstâncias), tais como gênero, etnia, local de nascimento e educação dos pais. Essa mensuração é feita via índice de dissimilaridade (D), comumente utilizado em sociologia, e que mede a proporção do total de recursos (ou pessoas) que deveria ser realocada para restabelecer, no caso em tela, a igualdade de oportunidades. O índice D é calculado da seguinte forma:

$$D = \frac{1}{2\hat{p}} \sum_k w_k (p_k - \hat{p}) \quad (2)$$

onde:

k denota o grupo não-vulnerável, isto é, com acesso ao ensino superior ($p > \hat{p}$);

p_k é a probabilidade de acesso do grupo k e \hat{p} a taxa de cobertura média da população;

w_k é a participação do grupo não vulnerável k no total da população.

O IOH combina acesso a um bem ou serviço (\hat{p}) e a forma de alocação (D) com base no princípio da igualdade de oportunidade da seguinte forma:

$$IOH = \hat{p}(1 - D) \quad (3)$$

De maneira intuitiva, o IOH capta a cobertura do bem ou serviço que é adequadamente alocada entre os diversos grupos de circunstâncias. Se o acesso dos indivíduos é independente de suas circunstâncias ($p_k = \hat{p}$ logo $D = 0$) então o IOH será igual à taxa de cobertura do bem e serviço. O IOH é igual a 1 quando o acesso é universal ($\hat{p} = 1$) e não há desigualdade no acesso ($p_j = \hat{p}$).

Observe que qualquer realocação de oportunidade dos grupos “não vulneráveis” (aqueles com $p_k > \hat{p}$) para os grupos “vulneráveis” (aqueles com $p_k < \hat{p}$) reduzirá D e aumentará o IOH, que é o efeito esperado das ações afirmativas no Brasil.

Seguindo Hoyos & Narayan (2012), o IOH possui algumas propriedades importantes, tais como:

- a) sensibilidade à escala: se a cobertura para todos os grupos muda aditivamente ou de forma multiplicativa pelo mesmo fator k , o IOH também muda (de forma aditiva ou multiplicativa) pelo mesmo fator:

$$\hat{p}^* = k\hat{p} \Rightarrow IOH^* = (k\hat{p})(1 - D) = kIOH \quad (4)$$

- b) sensibilidade para melhoria de Pareto: se a cobertura de um grupo aumenta sem diminuir a taxa de cobertura dos demais grupos o IOH também aumenta;
- c) sensibilidade à redistribuição: se a taxa de cobertura de um grupo vulnerável aumenta mantendo a taxa de cobertura total constante, o IOH também aumenta.

As mudanças do IOH ao longo do tempo (decomposição dinâmica) podem ser decompostas em dois componentes, segundo Barros et al. (2009):

- a) efeito distribuição (ΔD): se refere à contribuição nas mudanças na distribuição das circunstâncias para a mudança do IOH;
- b) efeito escala ($\Delta \hat{p}$): se refere à contribuição de uma variação proporcional na taxa de cobertura de todos os grupos.

Neste artigo, em que se calcula o IOH para 2003 (antes das ações afirmativas e primeiro ano em que a Pnad coleta a dependência administrativa da instituição de ensino) e para 2013 (dez anos após a implementação, ano mais recente disponível), a mudança temporal no IOH pode ser decomposta da seguinte forma:

$$\Delta_{2013,2003} = IOH_{2013} - IOH_{2003} = \Delta \hat{p} + \Delta D \quad (5)$$

onde o efeito escala e o efeito distribuição são definidos como segue:

$$\Delta \hat{p} = \hat{p}_{2013} (1 - D_{2013}) - \hat{p}_{2003} (1 - D_{2003}) \quad (6)$$

$$\Delta D = \hat{p}_{2013} (1 - D_{2013}) - \hat{p}_{2003} (1 - D_{2003}) \quad (7)$$

Outra propriedade do índice D, e por consequência do IOH, é que ele é uma função do conjunto de circunstâncias consideradas, mas, claramente, do número de grupos de circunstâncias consideradas. Mais formalmente, considere o índice de dissimilaridade D calculado a partir do vetor de circunstâncias x : $D = D(x)$. Se mais circunstâncias são adicionadas, por exemplo, vetor z , o índice de dissimilaridade para uma oportunidade irá aumentar, ou seja:

$$D(x) \leq D(x, z) \quad (8)$$

O resultado acima significa que o índice D não pode ser menor, mesmo que circunstâncias em falta são adicionadas ao conjunto inicial de circunstâncias, isto é, o índice D será sempre um limite inferior do nível de dissimilaridade que se poderia obter no caso da perfeita observação das circunstâncias. A consequência dessa propriedade é que o índice IOH, assim como qualquer outra medida de igualdade de oportunidades, subestima o grau de desigualdade de oportunidade e que a presente injustiça no acesso ao bem e serviço em questão pode ser maior do que a detectada pelo índice IOH.

Vega et al. (2010) fazem uma ressalva ao IOH: não se trata de uma medida direta da desigualdade de oportunidade, mas uma medida que revela como as oportunidades em questão estão disponíveis e são alocadas numa sociedade com base no princípio da igualdade. Peragine (2011) também apresenta algumas reticências sobre a consistência do índice IOH com a teoria da igualdade de oportunidades, mas elogia o índice por resumir em uma única medida a extensão das oportunidades agregadas para determinado país como avaliação de justiça em termos de como as oportunidades são distribuídas.

2.2 A contribuição de cada circunstância na desigualdade: a decomposição de Shapley

Políticas de ação afirmativa (cotas ou bônus; “raciais” ou “sociais”) têm sido implementadas no Brasil nos últimos dez anos com o objetivo de reduzir a desigualdade de oportunidades, por meio do aumento da probabilidade de acesso de grupos desfavorecidos ao ensino superior. Em abril de 2012, o STF

declarou constitucional a reserva de vagas no ensino superior para afrodescendentes. Nesta seção, vamos apresentar a metodologia que mensura a contribuição da cor/raça e outras características circunstanciais na desigualdade de oportunidades no acesso ao ensino superior.

Segundo Litchfield (1999), a técnica de decomposição de índices de desigualdade foi iniciada por Bourguignon (1979), sendo seguido por Cowell (1980) e Shorrocks (1980).

Existem três técnicas de decomposição:

- a) Por subgrupos da população: o objetivo desta decomposição é separar a desigualdade total na distribuição entre grupos (I_b) e intra-grupos (I_w). Essa técnica de decomposição foi aplicada por Dill & Gonçalves (2012) ao índice de Dissimilaridade D e ao IOH no acesso a serviços como energia elétrica, água limpa, saneamento e escola.
- b) Decomposição dinâmica: considera mudanças no nível de desigualdade ao longo do tempo e as classifica em efeito distribuição e efeito escala. Foi essa a utilizada por Barros et al. (2009).
- c) Decomposição por fator: considera as fontes e/ou fatores que determinam uma variável e mensura a contribuição de cada uma delas. Foi a escolhida por Hoyos & Narayan (2012) na decomposição do IOH entre crianças para examinar o quanto o gênero contribui para a desigualdade no acesso aos serviços essenciais de frequência escolar e vacinação.

Neste artigo, optou-se pela terceira técnica de decomposição porque, segundo Shorrocks (1999), esse procedimento pode ser utilizado sempre que se deseje avaliar a importância relativa das variáveis circunstanciais no índice de desigualdade, que é o que se deseja neste estudo. Cabe ressaltar que, seguindo a literatura, decompõe-se somente o componente de desigualdade (D) do índice composto (IOH), e não o índice em si. Isto nos parece fazer sentido, pois o foco central dessa etapa de nosso estudo é avaliar a contribuição das diferentes circunstâncias para a parcela do IOH referente à desigualdade de oportunidade (intergrupos) na probabilidade de ingresso no ensino superior e não à carência de oportunidades em geral (válida para todos os grupos e expressa por \hat{p}).

Essa decomposição por fator teve originalmente como inspiração a composição da renda total de um indivíduo, geralmente composta por mais de uma fonte: rendimentos do trabalho, rendimentos de capital, transferências públicas e privadas etc. Seguindo esse exemplo, vamos definir o vetor total de rendas y de uma população de tamanho n como a soma das rendas y_1, y_2, \dots, y_n onde n é o tamanho da população. Seja $I(y)$ uma estimativa da desigualdade na distribuição da renda. Uma regra natural de decomposição permite descrever o índice de desigualdade como uma soma ponderada da renda, de acordo com Shorrocks (1982):

$$I(y) = \sum_{i=1}^n a_i(y) y_i \quad (9)$$

onde a_i é o peso, y_i é a renda, i é indivíduo.

Sob essa estrutura, a contribuição proporcional ou relativa da fonte ou fator k (s^k) na desigualdade total será calculada por:

$$s^k = \frac{s^k}{I(y)} = \frac{\sum_{i=1}^n a_i(y) y_i^k}{\sum_{i=1}^n a_i(y) y_i} \quad (10)$$

onde S^k é a contribuição absoluta.

Um resultado imediato da razão (10) é que $\sum_k s^k = 1$. Para exemplificar, no caso do índice de Gini (I_{Gini}) teremos a contribuição relativa do fator k medida por (s_{Gini}^k) e calculada por:

$$s_{Gini}^k = \frac{s_{Gini}^k}{I_{Gini}} = \frac{\sum_{i=1}^n \left(i - \frac{n+1}{2}\right) y_i^k}{\sum_{i=1}^n \left(i - \frac{n+1}{2}\right) y_i} \quad (11)$$

De acordo com Sastre & Trannoy (2002), a contribuição absoluta⁴ da fonte k de acordo com a decomposição de Shapley por fator é dada por:

$$s^k = \sum \frac{(s-1)!(j-s)!}{j!} [I(y(S)) - I(y(S - \{k\}))] \quad (12)$$

Traduzindo a contribuição absoluta da um fator k de um índice de desigualdade $I(y)$ para a circunstância A de um índice de Dissimilaridade $D(S)$, de acordo com Hoyos & Narayan (2012):

$$S^A = \sum_{S \subset \{A\}} \frac{(s-1)!(c-s)!}{c!} [D(S\{A\}) - D(S)] \quad (13)$$

onde:

C é o conjunto de todas as circunstâncias, que contém c circunstâncias;

S é o subconjunto de C que contém s circunstâncias;

$D(SUA)$ é o índice de dissimilaridade calculado no conjunto de circunstância S e A ;

$D(S)$ é o índice de dissimilaridade calculado com o conjunto de circunstâncias S .

Para ilustrar melhor a equação (13) suponha que o índice de dissimilaridade é calculado usando três circunstâncias: C_1, C_2 e C_3 . Dado que as circunstâncias são correlacionadas entre si, o impacto marginal dependerá que qual conjunto pré-existente de circunstâncias C_1 é adicionado: se C_1 é adicionado ao conjunto $\{C_2, C_3\}; \{C_2\}$ ou $\{C_3\}$. Portanto, o impacto da adição da circunstância C_1 será dado por:

$$S^1 = \frac{2}{6} [D(C_1, C_2, C_3) - D(C_2, C_3)] + \frac{1}{6} [D(C_1, C_2) - D(C_2)] + \frac{1}{6} [D(C_1, C_3) - D(C_3)] + \frac{2}{6} [D(C_1) - 0] \quad (14)$$

A contribuição relativa da circunstância C_1 de acordo com a decomposição de Shapley por fator será:

$$s_{D-indice}^1 = \frac{S^1}{D} \quad (15)$$

onde D é o índice de dissimilaridade.

⁴A contribuição absoluta da fonte k também pode ser escrita em termos de variâncias conforme discussão apresentada em Ferreira & Gignoux (2014).

3 Dados

A base de dados selecionada para a mensuração da desigualdade de oportunidades no acesso ao ensino superior foi a Pesquisa Nacional por Amostra de Domicílios (Pnad) do IBGE. Dois motivos justificam essa escolha: primeiro, por ser uma pesquisa domiciliar de abrangência nacional e de periodicidade anual; segundo por investigar simultaneamente diversas características da população tais como educação, trabalho, rendimento e habitação, além de migração, fecundidade e características gerais dos domicílios.

O ano de 2013 foi selecionado por ser o ano mais recente disponível da Pnad. Escolhemos o ano de 2003 para definirmos um período exato de uma década entre os dois pontos de análise, por já haver informação do tipo de escola ou instituição de ensino que o aluno estava cursando (se pública ou privada) e também por coincidir temporalmente com o advento das políticas de ação afirmativa no ingresso ao ensino superior. No Brasil, as cotas foram instituídas pela primeira vez em 2001, na Universidade Estadual do Rio de Janeiro (UERJ) mas foram implementadas em 2003.

Como o nosso objeto de estudo são fatores relacionados ao acesso ao ensino superior, foram utilizados os seguintes filtros:

- Pessoas que moram em regiões urbanas;
- Que possuem entre 11 e 14 anos de estudo, isto é, que possuem o ensino médio completo como nível de instrução mais alto alcançado. Em 2013 tínhamos 44,9 milhões nesta situação escolar (26,3% do total da população urbana).
- Entre 17 e 29 anos de idade, o que representa 73,4% dos matriculados no ensino superior em 2013. Segundo o Mec/Inep, a faixa etária adequada/correta para se cursar o ensino superior no Brasil é entre 18 e 24 anos. Decidimos expandir essa faixa etária por causa do atraso escolar no ensino fundamental e médio e do grande estoque de pessoas sem ensino superior no Brasil. Outro motivo para a expansão da faixa etária é para a inclusão da idade média dos estudantes do ensino superior: 26,4 anos em 2013.
- Que são filhos na unidade domiciliar. Tivemos que usar esse filtro para poder capturar a instrução do chefe do domicílio na base de dados da Pnad. Outra justificativa é que, sendo filho na unidade domiciliar, o indivíduo não tem a obrigação de sustentar uma família e por isso tem mais opção entre escolher cursar uma universidade ou entrar no mercado de trabalho. Como o nosso objetivo neste capítulo é analisar o acesso ao ensino superior da demanda potencial, achamos que os filhos na unidade domiciliar pertencem a esse grupo.⁵ São filhos 54,0% das pessoas urbanas entre 17 e 29 anos de idade que possuem ensino médio completo.

As pessoas com características diferentes das citadas acima foram eliminadas da base de dados. A base de dados (ponderada) em 2003 totalizou 8,1

⁵Os dados da Pnad de 2013 comprovam esta hipótese. Entre as pessoas urbanas, de 17 a 29 anos de idade que estão matriculadas no ensino superior, 68,4% são filhos na unidade domiciliar; 11,2% são as pessoas de referência, 12,9% são outro parente e 7,5% são cônjuges.

milhões de pessoas e 9,87 milhões de pessoas em 2013. Com a aplicação dos filtros acima, acreditamos que estamos tratando da demanda potencial ao acesso ao ensino superior no Brasil.

Segundo Hoyos & Narayan (2012), para calcular o IOH para oportunidades de educação e saúde, além de uma pesquisa domiciliar (que é o caso da Pnad), é apropriado ter um conjunto mínimo de informações a nível individual, como gênero, localização (urbano/rural ou regionais), características dos pais, alguma medida do status socioeconômico do domicílio (rendimento, consumo ou riqueza). As características das pessoas consideradas como circunstâncias para analisar a probabilidade de acesso ao ensino superior foram:

- Gênero (duas categorias), medida por uma variável binária: 1 se mulher e 0 se homem;
- Etnia (duas categorias), medida por uma variável binária: 1 se branco ou amarelo e 0 se negro (preto, pardo ou mulato) ou indígena;
- Local de residência (cinco categorias), medida por quatro variáveis binárias regionais, com a região Sudeste como categoria base para analisar o efeito da oferta do ensino superior na concorrência de ingresso nos cursos;
- Nível de educação do chefe do domicílio (três categorias), medida por duas variáveis binárias, tendo ensino fundamental como categoria base como proxy para a origem familiar e *background* cultural da família.
- Grupo de riqueza da familiar que o indivíduo pertence, medido pelo rendimento domiciliar *per capita* (duas categorias): até 1,5 salários-mínimos ou mais de 1,5 salários-mínimos. Essa variável indica o status social ou a riqueza do domicílio. Foi medida com 1,5 salários mínimos porque esse é o critério utilizado para os beneficiados das ações afirmativas nas universidades federais e estaduais e também pelo programa Universidade para Todos (ProUni) para as instituições privadas do ensino superior.

Essas características pessoais consideradas como circunstâncias não são as únicas possíveis. Mas são as que estão disponíveis na base de dados utilizada. Por isso, no cálculo do índice de oportunidades humanas, temos a consciência de que se trata de um “lower bound”, ou seja, de uma cota inferior da medida dado o número limitado de circunstâncias utilizadas.

A Tabela 1 apresenta um breve perfil do grupo aqui denominado “demanda potencial ao ensino superior”, segundo suas circunstâncias, nos últimos 10 anos. A demanda potencial ao ensino superior aumentou 21,8% no período e o número de matriculados entre 2003 e 2013 aumentou 33,9%. Apenas 28,9% da demanda potencial ao ensino superior estavam matriculados em 2003 e este número aumentou em 2013 para 31,7% da demanda. Com relação à cor/raça do indivíduo, a demanda potencial ao ensino superior dos negros aumentou 60% entre 2003 e 2013 e a matrícula no ensino superior deste grupo duplicou, passando de 512 mil pessoas em 2003 para 1 milhão de pessoas em 2013.

Para fins comparativos, nos Estados Unidos, que possuem ações afirmativas para minorias há 30 anos, apenas 5,75% dos candidatos a uma vaga nas faculdades e universidades mais seletivas eram negros em 1989 (Bowen & Bok

2004). O Brasil que possui uma proporção da população negra muito maior, representando 47% da demanda potencial em 2013 (4,6 milhões sobre 9,8 milhões), instituiu reserva de vagas para os negros em todas as universidades federais apenas em 2012 com a “Leis das Cotas”.

Com relação à origem familiar, o volume da demanda potencial com chefe do domicílio com ensino médio como nível mais alto de educação aumentou 70% entre 2003 e 2013 e a matrícula deste grupo no ensino superior aumentou 40%, passando de 817 mil pessoas em 2003 para um milhão em 2011. Entretanto, em termos relativos, a maior incidência de matrícula no ensino superior é observada em filhos com chefe com ensino superior: 69,6% da demanda potencial estava matriculada no ensino superior em 2003 e 59,1% desse grupo em 2013 em comparação aos filhos com chefe com ensino médio: 42,8% em 2003 e 36,5% em 2013. Para captar o efeito dos recursos do domicílio disponíveis para o acesso ao ensino superior, apesar do aumento expressivo da demanda potencial com até 1,5 salário mínimo de renda domiciliar *per capita* no período, apenas 23,3% deste grupo garantiu sua matrícula no ensino superior em 2013.

Em suma, as categorias de circunstâncias que mais evoluíram no período entre 2003 e 2013 em termos de demanda e número de matrículas no ensino superior foram os pretos/pardos/mulatos, os residentes da região Norte, as pessoas com pai com ensino médio completo e também as pessoas com renda domiciliar *per capita* de até 1,5 sm. Mas o ingresso no ensino superior é maior, em termos relativos, para os indivíduos com renda domiciliar *per capita* superior a 1,5 sm e com pai com ensino superior, indicando a forte relação entre matrícula no ensino superior com pai muito instruído e com alto nível social.

A Tabela 2 apresenta um perfil dos ingressantes do ensino superior público e privado entre a demanda potencial. Nos 10 últimos anos de ação afirmativa (cotas, bônus, ProUni), a taxa de ingresso da demanda potencial nas instituições de ensino superior (IES) públicas aumentou de 9,4% em 2003 para 10,9% em 2013. A taxa de ingresso nas IES privadas aumentou de 23,2% em 2003 para 25,5% em 2013. As categorias com maior taxa de ingresso são as mulheres, os brancos, pai com educação superior e domicílios com renda alta. Entretanto, a taxa de ingresso dos negros aumentou no período analisado, tanto nas IES públicas quanto nas privadas. A taxa de ingresso dos filhos cujo pai possui pelo menos o ensino médio reduziu entre 2003 e 2013 nas instituições públicas e privadas.

4 Resultados

O primeiro passo para o cálculo do IOH é obter as probabilidades condicionais de acesso à oportunidade em questão, que neste trabalho consiste em ingressar numa instituição de educação superior (IES) com base nas suas circunstâncias. Como o nível de concorrência entre as IES públicas é muito maior do que nas IES privadas, gostaríamos de estimar duas regressões logísticas separadas: uma para as IES públicas e outra para as privadas. Os microdados da Pnad de 2003 e 2013 trazem essa informação da dependência administrativa da escola/IES que a pessoa está cursando.⁶

⁶A Pnad é uma pesquisa que utiliza um plano amostral complexo e isso implica a necessidade de alguns procedimentos na estimação de modelos estatísticos. Por isso, a estimação dos modelos econométricos foi feita por meio de comando de pacote estatístico que leva em conta a estrutura

Tabela 1: Perfil da demanda potencial ao ensino superior — Brasil (2003 e 2013)

Circunstâncias	Categoria	Demanda Potencial ao Ensino Superior, em 1.000 pessoas					
		2003			2013		
		Total	Matriculados IES		Total	Matriculados IES	
	Nº	%	Nº	%			
Total		8117	2342	28,9	9890	3135	31,7
Sexo	Homem	3421	1062	31,0	5007	1438	28,7
	Mulher	3817	1281	33,6	4873	1697	34,8
Etnia	Branco e Amarelo	5196	1830	35,2	5224	1971	37,7
	Preto, Pardo e Mulato	2920	512	17,5	4656	1164	25,0
Região de Residência	Norte	401	92	22,9	663	190	28,7
	Nordeste	1506	375	24,9	2189	673	30,8
	Sudeste	4540	1249	27,5	4990	1440	28,8
	Sul	1157	442	38,2	1308	516	39,5
	Centro-Oeste	513	183	35,7	730	315	43,2
Escolaridade	Até Fundamental	5275	882	16,7	5165	1066	20,6
Do chefe	Médio	1909	817	42,8	3156	1151	36,5
	Superior	922	642	69,6	1548	915	59,1
Renda Dom.	Até 1,5 sm	3835	480	12,5	6008	1400	23,3
Per capita	Mais de 1,5 sm	4032	1758	43,6	3140	1486	47,3

Fonte: IBGE, Microdados das Pnads 2003 e 2013.

¹ Pessoas que vivem em áreas urbanas, entre 11 e 14 anos de estudo (ensino médio completo como nível de instrução mais alto alcançado), entre 17 e 29 anos de idade e filho como posição no domicílio.

² O salário mínimo nominal vigente na época da pesquisa (setembro) era: 240 reais em 2003 e 678 reais em 2013. Os pontos de corte (1,5 sm) foram: 360 reais em 2003 e 1017 reais em 2013.

Precisamos estimar um modelo em que Y é a variável dependente binária, que assume o valor 1 se o indivíduo teve acesso à oportunidade que é ingressar no ensino superior e 0 se não teve. Os dois modelos mais comuns para estimar variável dependente binária são o probit binário e o logit binário. O modelo logit usa a função logística acumulada e o modelo probit usa a função de distribuição acumulada da normal. Graficamente, a função logística tem as caudas ligeiramente mais achatadas, ou seja, a função acumulada da normal se aproxima do eixo mais rapidamente do que a curva logística. Não há razão técnica para se preferir um a outro. Neste artigo, optou-se pelo modelo logit. A equação 16 apresenta o modelo logit na forma linear nos parâmetros. A matriz X é formada pelos vetores x_j que representam as circunstâncias disponíveis na Pnad e utilizada na pesquisa e β_j representa os $k + 1$ parâmetros que serão estimados pelo modelo. O $\pi(x_i)$ é uma simplificação da notação, que representa a média condicional de Y dado X quando a distribuição logística é utilizada.

Tabela 2: Perfil da dos ingressantes do ensino superior entre a demanda potencial — Brasil (2003 e 2013)

Variáveis	Categorias	IES PÚBLICAS				IES PRIVADAS			
		2003		2013		2003		2013	
		Não Ingr.	Ingr.	Não Ingr.	Ingr.	Não Ingr.	Ingr.	Não Ingr.	Ingr.
Total		90,6	9,4	89,1	10,9	76,8	23,2	74,5	25,5
Sexo	Masculino	91,0	9,0	90,0	10,0	78,5	21,5	77,4	22,6
	Feminino	90,1	9,9	88,1	11,9	75,2	24,8	71,5	28,5
Etnia	branco / amarela	88,9	11,1	86,5	13,5	70,5	29,5	69,0	31,0
	preta / parda / indígena	93,0	7,0	91,7	8,3	87,9	12,1	80,4	19,6
Região	Norte	88,6	11,4	87,3	12,7	85,6	14,4	79,6	20,4
	Nordeste	86,5	13,5	85,9	14,1	85,0	15,0	78,1	21,9
	Sudeste	93,3	6,7	92,3	7,7	76,4	23,6	75,6	24,4
	Sul	87,3	12,7	86,3	13,7	67,9	32,1	66,9	33,1
	Centro-oeste	88,0	12,0	83,0	17,0	70,5	29,5	64,3	35,7
Escol_chefe	Até EFund.	95,3	4,7	94,4	5,6	86,8	13,2	83,3	16,7
	EMédio	83,6	16,4	86,8	13,2	64,4	35,6	70,3	29,7
	ESup	62,3	37,7	69,6	30,4	37,3	62,7	49,8	50,2
Renda Dom.	Até 1,5 SM	94,6	5,4	91,8	8,2	92,1	7,9	82,3	17,7
	Acima de 1,5 SM	85,6	14,4	82,4	17,6	62,3	37,7	59,3	40,7

Fonte: IBGE, Pnads 2003 e 2013. Tabulações dos autores.

¹ Pessoas de 17 a 29 anos de idade, residentes em áreas urbanas, com ensino médio completo como nível de instrução mais alto alcançado e filho na posição do domicílio.

$$\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \ln\left(\frac{P(Y_i=1|X=(x_1, x_2, \dots, x_k))}{P(Y_i=0|X=(x_1, x_2, \dots, x_k))}\right) = \beta_0 + \sum_{j=1}^k \beta_j x_{i,j} \quad (16)$$

onde $\pi(x_i) = E(Y|X) = \frac{e^{\beta_0 + \sum \beta_j x_{i,j}}}{1 + e^{\beta_0 + \sum \beta_j x_{i,j}}}$

Essas estimações estão sujeitas a viés de simultaneidade — por exemplo, entre ingresso ou não no ensino superior e renda —, o que pode tornar questionável o processo de inferência e os resultados (Cruz & Moreira 2005). Diante da impossibilidade de se usarem variáveis instrumentais, ou outra técnica, para resolver o problema, optou-se por se apresentar os resultados de qualquer forma, fazendo-se a ressalva de que não há, na estimação dessas equações, nenhuma pretensão de estabelecimento de causalidade. Isto significa que, em havendo disponibilidade de dados que permitam corrigir o problema, é desejável que resultados de estudos futuros sejam comparados aos deste trabalho.

A Tabela 3 apresenta os coeficientes estimados (β), o erro padrão $Ep(\beta)$ e a razão de chances $\exp(\beta)$. Como o modelo de regressão é o logit, podemos interpretar seus coeficientes por meio da razão de chances, $\exp(\beta)$. Observe-se que a circunstância que obteve coeficiente com maior valor para o acesso aos cursos de graduação no Brasil em todos os anos foi a chefe e superior, que significa que domicílio com chefe com ensino superior possui uma chance de ingressar é 6,3 vezes maior que a de não ingressar em 2003 e 3,82 vezes em 2013. Separando o ingresso entre as instituições públicas e privadas em 2003, a variável que mais contribuiu para o ingresso na instituição pública foi o fato de o chefe do domicílio ter instrução superior, apesar de estar reduzindo sua participação: a chance de um filho(a) cujo pai possui instrução superior ingressar era de 8,67 em 2003 e reduziu para 5,54 em 2013. Entre as instituições privadas, além do nível de instrução do pai, a renda domiciliar também possui uma alta razão de chance para o ingresso: 4,17 em 2003 e 2,34 em 2013.

A Tabela 3 também apresenta algumas estatísticas dos testes de validação do modelo. A rejeição da hipótese nula de que todos os coeficientes (exceto a constante) são conjuntamente iguais a zero (teste de Qui-Quadrado) possui uma interpretação análoga ao teste global F do modelo de regressão linear múltiplo: que pelo menos um dos coeficientes é diferente de zero. A proporção de Y corretamente previsto é superior a 70% em todos os modelos estimados. O R^2 de McFadden, que é um índice da razão de verossimilhança, denotado por $R^2 = 1 - \frac{\ln L(\text{Modelo completo})}{\ln L(\text{Modelo Intercepto})}$, também foi calculado para todos os modelos. Com todas essas estatísticas, podemos concluir que os modelos estimados passaram pelo teste de validação.

Com os coeficientes estimados, pode-se obter para cada indivíduo da amostra sua probabilidade prevista de acesso ao ensino superior \hat{p}_i :

$$\hat{p}_i = \frac{\exp\left(\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{i,j}\right)}{1 + \exp\left(\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{i,j}\right)} \quad (17)$$

Usando as probabilidades previstas na equação (17) e os pesos amostrais w , calculamos a taxa média de acesso ao ensino superior \bar{p} , também conhecida como a taxa global de cobertura prevista:

Tabela 3: Resultados da estimação do Modelo Logit, por categoria administrativa da IES — Brasil (2003 e 2013)

Variáveis / Circunstâncias	IES Total						IES Públicas						IES Privadas						
	2003			2013			2003			2013			2003			2013			
	β	EP(β)	Exp(β)	β	EP(β)	Exp(β)	β	EP(β)	Exp(β)	β	EP(β)	Exp(β)	β	EP(β)	Exp(β)	β	EP(β)	Exp(β)	
Mulher	0,38	0,02	1,46	0,45	0,02	1,57	0,23	0,03	1,26	0,29	0,03	1,34	0,44	0,02	1,55	0,50	0,02	1,65	
Branco / Amarelo	0,53	0,02	1,69	0,34	0,02	1,41	0,42	0,04	1,52	0,41	0,03	1,51	0,58	0,03	1,78	0,32	0,02	1,38	
Região Norte	0,25	0,05	1,29	0,20	0,03	1,22	1,02	0,06	2,78	0,81	0,05	2,25	-0,11	0,06	0,90	-0,03	0,04	0,97	
Região Nordeste	0,36	0,03	1,44	0,34	0,02	1,41	1,18	0,04	3,24	0,98	0,03	2,65	-0,05	0,03	0,95	0,12	0,02	1,12	
Região Sul	0,46	0,03	1,59	0,30	0,02	1,35	0,65	0,04	1,92	0,50	0,04	1,66	0,40	0,03	1,50	0,24	0,03	1,27	
Região C-Oeste	0,47	0,04	1,59	0,63	0,03	1,87	0,72	0,06	2,06	0,89	0,05	2,43	0,40	0,04	1,49	0,55	0,03	1,73	
Chefe_EMédio	0,99	0,02	2,70	0,66	0,02	1,93	1,14	0,03	3,11	0,83	0,03	2,29	0,94	0,02	2,56	0,60	0,02	1,83	
Chefe_ESuperior	1,84	0,03	6,30	1,34	0,02	3,82	2,16	0,04	8,67	1,71	0,04	5,54	1,74	0,03	5,70	1,21	0,02	3,37	
Renda_ > 1,5 SM	1,21	0,02	3,34	0,75	0,02	2,13	0,67	0,03	1,95	0,48	0,03	1,62	1,43	0,02	4,17	0,85	0,02	2,34	
Constante	-2,92	0,03	0,05	-2,12	0,02	0,12	-4,14	0,05	0,02	-3,73	0,04	0,02	-3,34	0,03	0,04	-2,36	0,02	0,09	
% Y estimado cor- retamente	76,92			72,12			90,79			89,20			79,88			75,82			
Estatística de Qui- Quadrado (9 gl)	17477,91			11638,66			5762,61			5188,46			15718,46			9403,86			
p-valor	< 0,0001			< 0,0001			< 0,0001			< 0,0001			< 0,0001			< 0,0001			
McFadden's Square	R	0,1862		0,1023		0,1503		0,1078		0,2010		0,0993							

Fonte: Fonte: Elaboração própria a partir dos microdados da Pnad.

Nota: β = parâmetro estimado; EP(β) = erro-padrão, Exp(β) = razão de chances.

$$\bar{p} = \sum_{i=1}^n w_i \hat{p}_i \quad (18)$$

Com o número de categorias em cada circunstância (sexo=2, etnia=2, região=5, educação do chefe=3, grupos de quintos de rendimento=2) podemos combinar todas essas categorias em 120 grupos (2x2x5x3x2) e assim estimar a probabilidade média prevista de ingresso dos indivíduos ao ensino superior em cada grupo k .

4.1 Grupos sociais “vulneráveis” no tocante ao ensino superior

Os grupos cuja probabilidade média de ingresso (p_k) é inferior à probabilidade média de acesso (\bar{p}) são chamados de grupos de “pessoas vulneráveis no que se refere ao acesso ao ensino superior”. Ordenando as probabilidades de acesso de cada grupo do menor para o maior, obtemos as Figuras A.1 e A.2 (Apêndice).

O grupo mais vulnerável no acesso ao ensino superior público, isto é, com menor probabilidade de acesso, é formado por homens não brancos, residentes da região Sudeste, com pai com até ensino fundamental como nível de instrução e renda domiciliar alta. A probabilidade média de acesso desse grupo em 2003 era de 1,57% e em 2013, 2,33%. O grupo com maior probabilidade de acesso nas instituições públicas (menos vulnerável) era formado por mulheres brancas, residentes na região nordeste, com pai com nível de instrução superior e renda baixa. As probabilidades de acesso desse grupo em 2003 e 2013 são respectivamente 62,5% e 53,6%.

Com relação às instituições privadas, o grupo mais vulnerável no acesso (menor probabilidade) é composto por homens, não brancos, residentes da região Norte, com pai com até ensino fundamental e renda alta. O grupo menos vulnerável é formado por mulheres brancas, residentes na região sudeste, com pai com educação superior e renda baixa. As probabilidades de acesso são respectivamente 77,7% em 2003 e 74,5% em 2013.

4.2 Índice de igualdade de oportunidades ao acesso ao ensino superior

Agora podemos calcular o índice de dissimilaridade (D) relativo ao acesso ao ensino superior, que mede qual proporção do total de oportunidades deveria ser realocada para restabelecer a igualdade de oportunidades. O cálculo do índice D é baseado nos grupos com oportunidade de acesso ao ensino superior, isto é, os “não vulneráveis” ($p_k > \bar{p}$):

$$D = \frac{1}{2\bar{p}} \sum_{k=1}^K w_k (\bar{p} - p_k) \quad (19)$$

$$IOH = \bar{p}(1 - D) \quad (20)$$

Os componentes D e \bar{p} foram combinados segundo a equação (20) para gerar o IOH (Tabela 4). O índice de dissimilaridade (D) em 2003 indica que, para alcançar a igualdade de oportunidades no acesso ao ensino superior, 15,5% das pessoas deveriam ser realocadas — isto é, deveriam ceder seus lugares na universidade ou ocupar esses lugares cedidos. O índice de dissimilaridade de 2013 caiu para 10,8%, o que indica que houve realocação de oportunidade dos

grupos “não vulneráveis” para os grupos “vulneráveis”. A disponibilidade do ensino superior para a sociedade (taxa de cobertura) aumentou de 28,5% em 2003 para 31,6% em 2013, e a desigualdade no acesso ao ensino superior reduziu-se, o que resulta em um IOH maior em 2013.

Tabela 4: IOH no acesso ao ensino superior e seus componentes — Brasil (2003 e 2013)

Indicadores	Total		Públicas		Privadas	
	2003	2013	2003	2013	2003	2013
Mínima	0,051	0,107	0,015	0,023	0,031	0,084
Máxima	0,817	0,801	0,625	0,536	0,777	0,745
Média	0,285	0,316	0,113	0,121	0,237	0,261
D	0,155	0,108	0,196	0,158	0,177	0,120
IOH	0,240	0,281	0,091	0,102	0,195	0,230

Fonte: Elaboração própria a partir dos microdados da Pnad.

Entretanto, o IOH para o acesso ao ensino superior está muito longe da situação ideal ($IOH = 1$) que ocorre quando o acesso é universal ($\bar{p} = 1$) e não há desigualdade no acesso ($\bar{p} = p_k \Rightarrow D = 0$). O IOH calculado evidencia quão distante está o Brasil de um acesso amplo e igualitário ao ensino superior, indicando como as características pessoais impactam na probabilidade de ter acesso a este nível de ensino. Para se ter uma ideia dessa reduzida igualdade de oportunidades no acesso ao ensino superior comparado com outros níveis educacionais e serviços, vamos comparar o resultado encontrado aqui com o de Dill & Gonçalves (2012), que utilizaram a Pnad e a mesma metodologia de Barros et al. (2009). Com relação ao componente educação, eles consideraram a probabilidade de se completar a sexta série na idade adequada. Observe que houve aumento de cobertura deste componente de 62,12% em 1999 para 73,16% em 2009 (Tabela 5). Também observaram uma redução na desigualdade no acesso a este componente educacional, o que resultou em um IOH maior em 2009, alcançando 0,6984. Entre os componentes analisados, o que alcançou maior IOH foi acesso à eletricidade no domicílio, que com uma cobertura de 98,22% dos domicílios em 2009 e uma desigualdade no acesso a este serviço muito baixa (1,18%), alcançou um IOH de 97,06%, muito próximo de 100%.

Tabela 5: IOH no acesso à educação básica, saneamento, água e energia e seus componentes — Brasil (1999 e 2009)

Ano	Indicadores	Educação	Saneamento	Água	Energia
1999	Cobertura	0,6212	0,4407	0,8042	0,9332
	Dissimilaridade	0,0895	0,2419	0,1118	0,0472
	IOH	0,5656	0,3341	0,7143	0,8892
2009	Cobertura	0,7316	0,4926	0,8920	0,9822
	Dissimilaridade	0,0452	0,2021	0,0624	0,0118
	IOH	0,6984	0,3930	0,8364	0,9706

Fonte: Dill & Gonçalves (2012). Microdados da Pnad.

A mudança do IOH ao longo do tempo, isto é, de 2003 a 2013, também medido por $\Delta_{2013,2003} = IOH_{2013} - IOH_{2003} = 0,041$, pode ser decomposta em

dois componentes:

- Efeito escala, que se refere à contribuição de uma variação na taxa de cobertura e calculada por: $\Delta\bar{p} = \bar{p}_{2013}(1 - D_{2013}) - \bar{p}_{2003}(1 - D_{2003}) = 0,026$.
- Efeito distribuição, que se refere à contribuição nas mudanças da distribuição das circunstâncias para mudanças no IOH: $\Delta D = \bar{p}_{2013}(1 - D_{2013}) - \bar{p}_{2003}(1 - D_{2003}) = 0,015$.

Observe-se que, no período de 2003 a 2013, o efeito escala foi maior que o efeito distribuição, indicando que as ações afirmativas (cotas, bônus, ProUni) tiveram maior contribuição na taxa de cobertura do ensino superior do que na distribuição das vagas segundo as circunstâncias consideradas.

Cabe ressaltar que os cálculos utilizados nesta seção possuem uma limitação: levam em consideração os filtros adotados, isto é, trata-se do acesso e da dissimilaridade da demanda potencial ao ensino superior, que foi definida como as pessoas urbanas de 17 a 29 anos de idade, que são filhos na unidade domiciliar. O avanço do ensino superior com relação às pessoas de 30 anos ou mais de idade não foi considerado nesse cálculo.

4.3 A contribuição de cada circunstância: a decomposição por fatores de Shapley

Conforme a Tabela 6, a variável cor/raça não é a circunstância com maior contribuição na desigualdade de oportunidades no acesso ao ensino superior, medido pelo índice de dissimilaridade D , mas tem um efeito razoável que, se acumulado com as demais circunstâncias, pode ser decisivo para impedir o acesso à universidade.⁷ Em 2003, a variável cor/raça explicava 13,8% do índice de D . Em 2013, esta circunstância aumentou sua participação para 14,7% na desigualdade. As circunstâncias que mais contribuem para o acesso ao ensino superior no Brasil são as relacionadas ao *background* familiar, mensurada pela instrução do chefe do domicílio e também pela renda domiciliar *per capita*: em 2003 e em 2013 contribuíram com mais de 70% do total de desigualdade. Note-se que a instrução do chefe do domicílio é a circunstância com maior contribuição ao acesso dos cursos das IES públicas e a renda domiciliar *per capita* é a circunstância com maior contribuição dentre as instituições privadas. Note-se também na Tabela 6 que houve uma redução da contribuição da instrução do chefe no acesso aos cursos de instituições públicas e um aumento na contribuição da cor/raça, talvez pelo fato da primeira não fazer parte do critério de concessão de bolsas do ProUni e das cotas e da segunda circunstância fazer parte.

Litchfield (1999) adverte que os resultados de qualquer decomposição devem ser discutidos e analisados como meramente descritivos porque a desigualdade é determinada por muitos fatores e em equilíbrio geral. Acrescenta-se que não há preocupação, nessa literatura, de se estabelecerem relações causais. Trata-se de correlações condicionais entre um conjunto de variáveis dependentes e uma variável de resposta.

⁷Vale notar que outras definições de grupos (ex. uma definição mais fina de cor/raça que a adotada aqui) poderiam conduzir a resultados diferentes dos apresentados nesta seção. Por isso, adota-se cautela nas interpretações.

Tabela 6: Contribuição Relativa das Circunstâncias ao Acesso ao ensino superior, utilizando a decomposição por fator de Shapley — Brasil (2003, 2013)

Circunstância	Todas as IES		IES Públicas		IES Privadas	
	2003	2013	2003	2013	2003	2013
Gênero	1,7	4,0	1,0	2,5	2,4	5,2
Cor/Raça	13,8	14,7	6,2	10,8	14,8	13,1
Região	3,7	8,2	11,3	17,3	5,5	6,9
Instrução Chefe	40,7	39,0	59,8	48,1	36,8	36,1
Renda Domiciliar	40,1	34,1	21,7	21,3	40,5	38,7
Total	100,0	100,0	100,0	100,0	100,0	100,0

Fonte: IBGE, Pnad.

5 Conclusões

Apesar dos avanços no ensino fundamental e médio nas últimas décadas, e da ampliação de vagas no ensino superior, apenas 12% dos jovens em idade universitária estão cursando o ensino superior no Brasil. Na Argentina essa cobertura chega a 21%, nos Estados Unidos a 65% e na Suécia a 70%.

Neste estudo, a partir dos microdados mais recentes da Pnad disponíveis (2013) e dos dados de uma década anterior, antes das ações afirmativas no acesso ao ensino superior serem adotadas no Brasil (2003), traça-se um perfil dos grupos “vulneráveis” e “não vulneráveis” ao acesso ao ensino superior, considerando a demanda potencial ao acesso ao ensino superior.

Mensuraram-se também as desigualdades de oportunidades no acesso ao ensino superior no Brasil. O índice de dissimilaridade (D) em 2003 indica que 15,5% das oportunidades deveriam ser realocadas para restabelecer a igualdade. Já o mesmo índice em 2013 reduziu para 10,8%, o que indica que houve realocação de oportunidade dos grupos “não vulneráveis” para os grupos “vulneráveis”. Assim, a disponibilidade do ensino superior para a sociedade aumentou (de 28,5% para 31,6%) e a desigualdade no acesso ao ensino superior caiu. Entretanto, o índice de oportunidades humanas ao acesso ao ensino superior no Brasil ($IOH=0,281$ em 2013) está muito longe da situação ideal ($IOH = 1$) que ocorre quando o acesso é universal ($\bar{p} = 1$) e não há desigualdade no acesso ($D=0$).

Utilizando a decomposição de Shapley por fatores, as circunstâncias que mais contribuem para a desigualdade de oportunidades no acesso ao ensino superior são: a instrução do chefe do domicílio e a renda domiciliar *per capita*, ou seja, o *background* cultural da família e o *status* econômico. Esses dois fatores conjuntamente explicavam 80,8% do índice de dissimilaridade (D) do acesso ao ensino superior em 2003; e 73,2% do índice D em 2013. Nas instituições públicas, destaca-se a instrução do pai como a principal circunstância, com uma contribuição de 48,1% em 2013 e nas instituições privadas destaca-se a renda domiciliar *per capita*, com uma contribuição de 40,5% em 2013.

Agradecimentos

Gostaríamos de agradecer a Maria Dolores Montoya Diaz, Rosana Rodrigues Heringer, Hustana Vargas e Danielle Carusi Machado por valiosos comentários e sugestões feitos na defesa de tese de M. Carvalho, tese que continua o

estudo do qual se originou este artigo. Também somos gratos aos membros do Núcleo de Estudos em Educação, do Centro de Estudos em Desigualdade e Desenvolvimento da UFF (CEDE), que colaboraram em todas as etapas do desenvolvimento desta pesquisa. Também devemos agradecimentos a pareceristas anônimos deste periódico, que nos permitiram aprimorar a qualidade do trabalho. Os autores são os únicos responsáveis por esta versão final.

Referências Bibliográficas

- Arneson, R. (1989), 'Equality of opportunity of welfare', *Philos. Stud.* **56**, 77–93.
- Barr, N. (2012), *Economics of the Welfare State*, Oxford University Press.
- Barros, R. P., Ferreira, F. H. G., Vega, J. R. M. & Chanduvi, J. S. (2009), Measuring inequality of opportunities in latin américa and caribbean, Technical report, The World Bank. Washington D.C.
- Bourguignon, F. (1979), 'Decomposable income inequality measures', *Econometrica* **47**, 901–920.
- Bowen, G. & Bok, D. (2004), *O curso do Rio: um estudo sobre a ação afirmativa no acesso à universidade*, Garamond, Rio de Janeiro.
- Carvalho, M. (2011), 'A educação superior no brasil: o retorno privado e as restrições ao ingresso', *Revista Sinais Sociais* **5**(15), 82–109.
- Carvalho, M., Gamboa, L. & Waltenberg, F. (2012), Equality of education opportunity employing pisa data: taking both achievement and access into account, Working Paper Series 277, ECINEQ.
- Cohen, G. (1989), 'On the currency of egalitarian justice', *Ethics* **99**, 906–944.
- Cowell, J. (1980), 'On the structure of additive inequality measures', *Review of Economic Studies* **47**, 521–531.
- Cruz, L. & Moreira, M. (2005), 'On the validity of econometric techniques with weak instruments: Inference and returns to education using compulsory school attendance laws', *Journal of Human Resources* **40**(2), 393–410.
- Diaz, M. (2012), '(des)igualdades de oportunidades no ensino médio brasileiro: escolas públicas e privadas', *Revista Economia* **13**(3), 553–568.
- Dill, H. & Gonçalves, F. (2012), 'Índice de oportunidade no brasil entre 1999 e 2009: estimação e decomposição através do valor de shapley', *Pesquisa e Planejamento Econômico* **42**(2), 187–210.
- Dworkin, R. (1981), 'What is equality? part 2: equality of resources', *Philos Aff.* **10**, 283–345.
- Ferreira, F. & Gignoux, J. (2008), 'The measurement of inequality of opportunity: Theory and an application to latin america', *Review of Income and Wealth* **57**(4), 622–657.
- Ferreira, F. & Gignoux, J. (2014), 'The measurement of educational inequality: Achievement and opportunity', *World Bank Econ.* **28**(2), 210–246.

Fleurbaey, M. (2008), *Fairness, Responsibility and Welfare*, Oxford University Press.

Hoyos, A. & Narayan, A. (2012), Inequality of opportunities among children: how much does gender matter?, in 'The World Development Report 2012: Gender Equality and Development'.

INEP (2011), Sinopse estatística da educação básica: Censo escolar, Technical report, INEP, Brasília: O Instituto.

IPEA (2008), Políticas sociais — acompanhamento e análise, Technical Report 15, IPEA.

Litchfield, J. (1999), Inequality: Methods and tools, Technical report.

URL: Disponível em: <<http://www.worldbank.org/poverty/inequal/index.htm>>.

Mendonça, R. (2000), A oportunidade imperdível: expansão educacional e desenvolvimento humano no Brasil, PhD thesis, Instituto de Economia da Universidade Federal do Rio de Janeiro (UFRJ).

OCDE (2011), Education at a glance, Technical report, OCDE, Paris, Unesco.

Peragine, V. (2011), 'Review of "measuring inequality of opportunities in latin america and the caribbean"', *Journal of Economic Inequality* 9, 137–143.

Pignataro, G. (2012), 'Equality of opportunity: Policy and measurement paradigms', *Journal of Economic Surveys* .

Ramos, X. & Van De Gaer, D. (2012), Empirical approaches to inequality of opportunity: Principles, measures, and evidence, Working Papers 259, ECINEQ Society for the Study of Economic Inequality.

Roemer, J. (1993), 'A pragmatic theory of responsibility for the egalitarian planner', *Philos, Public Fff* 22, 146–166.

Roemer, J. (1998), *Equality of Opportunity*, Harvard University Press.

Sastre, M. & Trannoy, A. (2002), 'Shapley inequality decomposition by factor components: some methodological issues', *Journal of Economics Supplement* 9, 51–90.

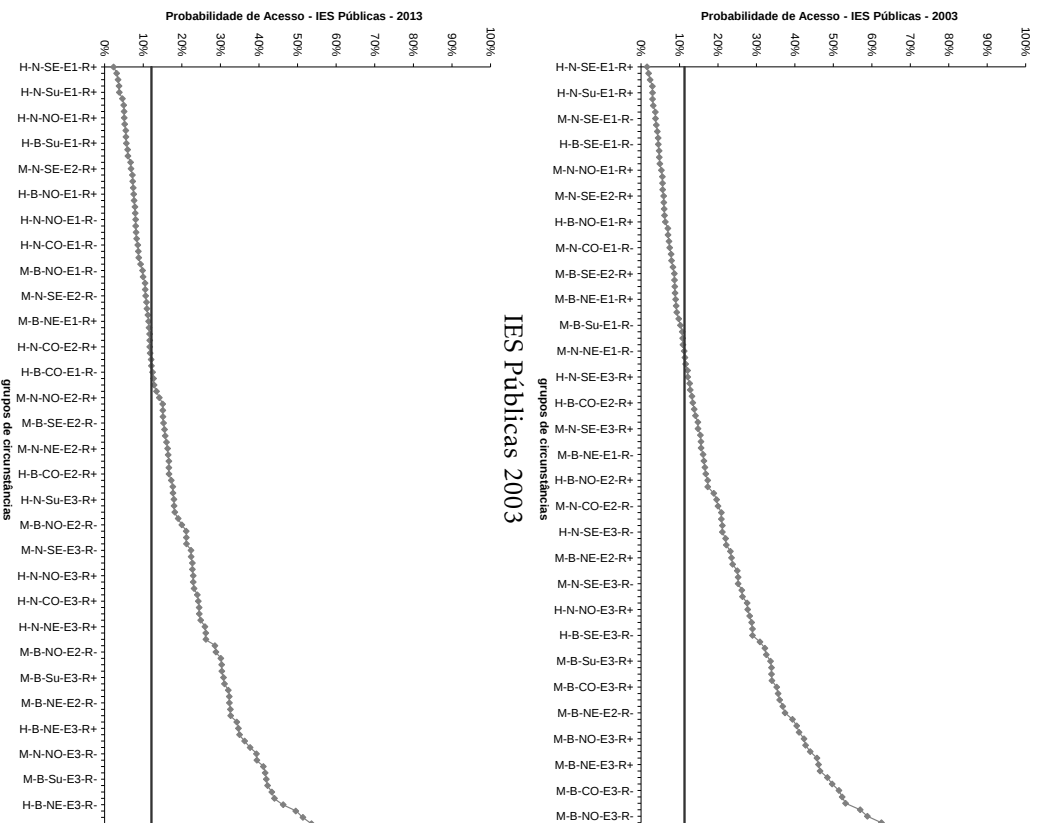
Shorrocks, A. (1980), 'The class of additively decomposable inequality measures', *Econometrica* 48, 613–625.

Shorrocks, A. (1982), 'Inequality decomposition by factor components', *Econometrica* 50, 193–211.

Shorrocks, A. (1999), 'Decomposition procedures for distributional analysis: a unified framework based on shapely value', Department of Economics, University of Essex.

Vega, J. R. M., Barros, R. P., Saavedra, J. & Giugale, M. (2010), Do our children have a chance?, in 'The 2010 Human Opportunity Report for Latin America and the Caribbean', Word Bank, Washington DC, p. 176 p.

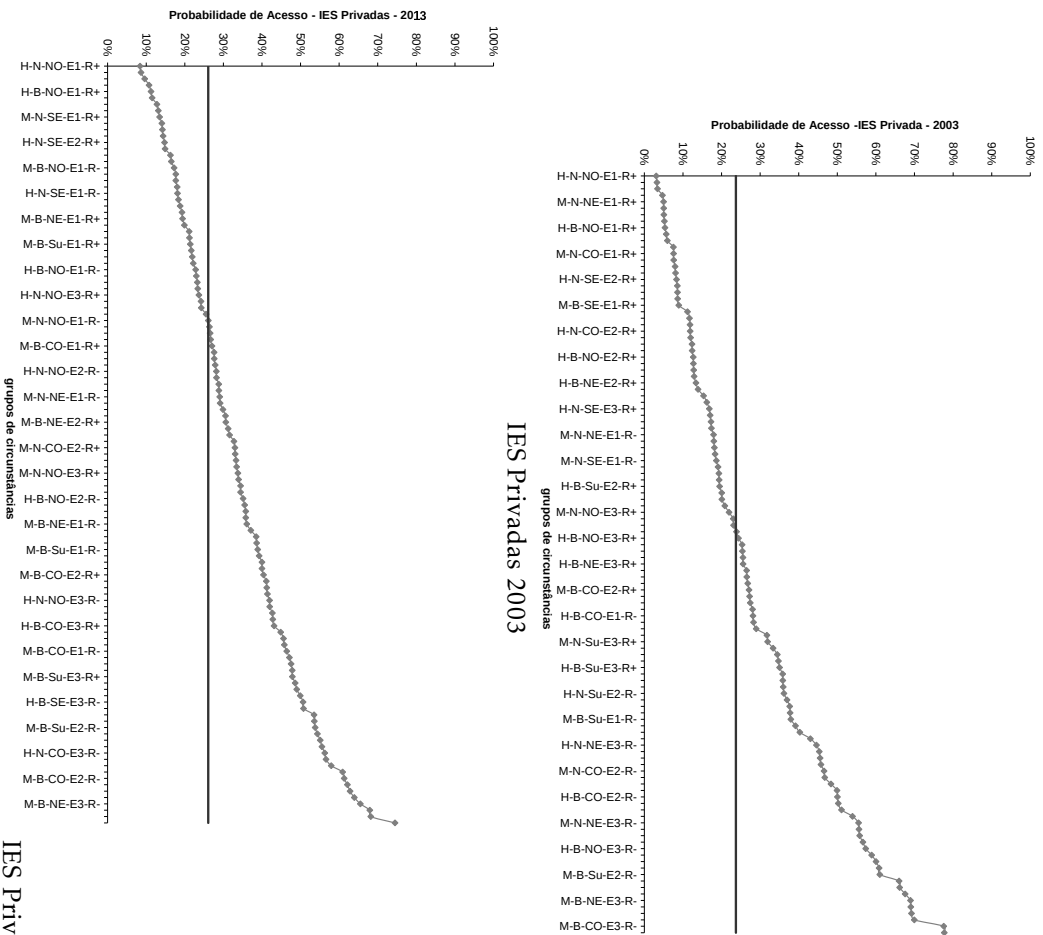
Apêndice A



Fonte: IBGE, Pnads 2003 e 2013. Tabulação dos autores.

Legenda: H=Homem, M=Mulher, N=não branco, B=Branco, NO=Norte, NE=Nordeste, SE=Sudeste, Su=Sul, CO=Centro-Oeste, E1=Até E,Fundamental, E2=Ensino Médio, E3=Ensino Superior; R=renda domiciliar per capita inferior a 1,5 SM; R+=renda domiciliar per capita superior a 1,5 SM.

Figura A.1: Probabilidade de acesso ao ensino superior segundo os grupos



Fonte: IBGE, Prads 2003 e 2013. Tabulação dos autores.

Legenda: H=Homem, M=Mulher, N=não branco, B=Branco, NO=Norte, NE=Nordeste, SE=Sudeste, Su=Sul, CO=Centro-Oeste, EI=Até E,Fundamental, E2=Ensino Médio, E3=Ensino Superior, R=renda domiciliar per capita inferior a 1,5 SM; R+=renda domiciliar per capita superior a 1,5 SM.

Figura A.2: Probabilidade de acesso ao ensino superior segundo os grupos

Informações aos autores

Objetivo e escopo

A ECONOMIA APLICADA é uma publicação trimestral do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, com uma linha editorial voltada exclusivamente para artigos científicos em economia aplicada. Tem por foco a publicação de trabalhos de análise econômica aplicada a problemas específicos, tanto de interesse do setor público como privado, e que contenham, particularmente, estudos quantitativos, cujos resultados possam aproximar a teoria da realidade. A interação entre trabalho empírico e teoria, análise e política econômica é, portanto, um importante diferencial da revista. A revista ECONOMIA APLICADA coloca-se como um espaço aberto para toda a comunidade acadêmica, nacional e internacional.

Política editorial

Os artigos submetidos são avaliados inicialmente pelos Editores quanto a sua adequação à linha editorial da Revista. Uma vez aceita a submissão, o artigo é encaminhado para avaliação por pareceristas designados pelos Editores, no sistema double-blind.

Apresentação de Artigos

Os artigos submetidos para avaliação devem obedecer às seguintes normas:

1. Apresentar ineditismos no país e no exterior
2. Na primeira página devem constar o título do artigo, nome dos autores com endereço para contato e afiliação institucional, resumo com até 100 palavras em português e inglês, palavras chave, em português e inglês e classificação JEL.
3. Na versão para “avaliação”, basta um arquivo no formato pdf com todas as equações, tabelas e figuras no corpo do texto, e referências no final.
4. Caso o artigo venha a ser aceito para publicação, a equipe da revista contatará os editores para as atividades de editoração. Nessa ocasião serão requeridos os seguintes arquivos:
 - Arquivo fonte no formato \LaTeX .
 - Arquivo de referências bibliográficas no padrão Bib \TeX .
 - Arquivo com as figuras em formato TIFF ou EPS.
 - Planilha eletrônica com todas as tabelas.

Responsabilidade e Direitos Autorais

Os autores que submeterem ou tiverem o artigo aceito para publicação na Revista ECONOMIA APLICADA declaram:

- Ter participado da concepção do artigo para tornarem públicas suas responsabilidades pelos seus conteúdos, que não omitiram quaisquer ligações ou acordos de financiamento entre autores e companhias que possam ter interesse na publicação deste artigo.
- Que o manuscrito é inédito e que o artigo, em parte ou na íntegra, ou qualquer outro artigo com conteúdo substancialmente similar, de sua autoria, não foi enviado à outra revista e não o será enquanto sua publicação estiver sendo considerada pela revista ECONOMIA APLICADA, que seja no formato impresso ou no eletrônico, exceto o descrito em anexo.
- Que a revista ECONOMIA APLICADA passa a ter os direitos autorais para a veiculação dos artigos, tanto em formato impresso como eletrônico a eles referentes, os quais se tornarão propriedade exclusiva da ECONOMIA APLICADA. É permitida a reprodução total ou parcial dos artigos desta revista, desde que seja explicitamente citada a fonte completa.

ISSN 1413-8050



The Brazilian Journal of Applied Economics is a quarterly publication of the Departments of Economics of the FEA-RP of the University of São Paulo. The published papers are sole responsibility of the author(s). As the author(s) send a manuscript for submission to the Journal they tacitly agree that, if accepted for publication, the copyright of the material will be transferred to the ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS). The Journal may then publish the material in paper and electronic formats that will be exclusive property of the ECONOMIA APLICADA (BRAZILIAN JOURNAL OF APPLIED ECONOMICS). The publication of the whole papers or parts of it is permitted as long as the source is explicitly mentioned. The Brazilian Journal of Applied Economics is published with the financial support of the CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico. This journal is indexed by Journal of Economic Literature, electronic on line and CD-ROM, and Scielo Brasil.

Secretary, Sales and Delivery: Juliana de Freitas Vitória Cover: Christof Gunkel
Linguistic revision (Portuguese and English): Iulo Feliciano Afonso (Solution Translations Eireli — Epp)
Diagramming: Bruna Carolina Garcia

Mailing Address

Revista Economia Aplicada, Departamento de Economia, FEA-PR/USP
Avenida do Bandeirantes, 3900, Ribeirão Preto, SP.
CEP 14040-905
Telefone: (0xx16) 3602-3910, fax: (0xx16) 3633-4488
email:revecap@usp.br
site: <http://www.usp.br/revecap>

Subscriptions for Delivery Outside Brazil

Individuals \$ 90.00
Institutions \$110.00

* Air mail included

† Individual issues can
also be purchased at
the above address.

Editor

Reynaldo Fernandes

Associate Editors

Luiz Guilherme Scorzafave

Angelo Costa Gurgel

Rodrigo De Losso

Cláudio Ribeiro de Lucinda & Walter Belluzzo Junior

Sergio Naruhiko Sakurai

Eliezer Martins Diniz

Irineu E. de Carvalho Filho

Editorial Board

Affonso Celso Pastore

USP

Cássio F. Camargo Rolim

UFPR

Clélio Campolina Diniz

CEDEPLAR

Eleutério F. S. Prado

USP

Francisco Anuatti Neto

USP/RP

Geraldo Sant'ana de Camargo Barros

ESALQ/USP

José Marcelino da Costa

NAEA/PA

Marcelo Portugal

UFRGS

Márcio Gomes Pinto Garcia

PUC/ RJ

Paulo César Coutinho

UnB

Pierre Perron

Boston University

Ricardo R. Araújo Lima

UnB

Roberto Smith

UFCE

Rogério Studart

UFRJ

Sérgio Werlang

FGV/RJ

Victor Bulmer-Thomas

University of London

Wilson Suzigan

Unicamp

Antônio Barros de Castro

UFRJ

Cláudio Monteiro Considera

UFF

Denisard C. de Oliveira Alves

USP

Fernando de Holanda Barbosa

FGV-UFF

Geoffrey J. D. Hewings

University of Illinois

Gustavo Maia Gomes

IPEA

José A. Scheinkman

Princeton University

Maria José Willumsen

Florida International University

Mário Luiz Possas

UFRJ

Paulo Nogueira Batista Júnior

FGV/SP

Pedro Cesar Dutra Fonseca

UFRGS

Robert E. Evenson

Yale University

Rodolfo Hoffmann

ESALQ/USP

Russell E. Smith

Washburn University

Tomás Málaga

FGV/SP

Werner Baer

University of Illinois